

Active discovery of catalysts for sustainable energy storage

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy
in the
Chemical Engineering Department

Kevin Tran

BSChE, University of Delaware

Carnegie Mellon University
Pittsburgh, PA

January, 2021

Active discovery of catalysts for sustainable energy storage

by

Kevin Tran

Submitted to the Chemical Engineering Department
on 27 January 2021, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Consumption of fossil fuels has caused climate to change at unmanageable rates, making adaptation difficult both environmentally and economically. This is shown by an orders-of-magnitude increase in global extinction rates, signaling environmental destabilization. It is also shown by an increasing rate and severity of extreme weather events such as droughts, floods, or wildfires, straining our economies and endangering our food & water supplies. We may be able to slow these changes by transitioning from fossil fuels to solar energy, but solar energy’s inconsistent availability makes implementation difficult. This could be addressed by storing the energy in solar fuels, which are fuels created from solar energy, CO_2 , and H_2O . Unfortunately, solar fuels are hindered by a lack of commercial viability. We could solve this issue by finding catalysts that produce solar fuels more quickly, selectively, and efficiently.

This thesis comprises several projects aimed at discovering catalysts for solar fuel production. We first show how recent advances in computation and data science can accelerate the catalyst discovery process. We illustrate this point by creating a software framework that performs high-throughput density functional theory (DFT) calculations, which can be used to predict catalyst performance. Then we combine our framework with a heuristic method for “active discovery”. Active discovery is the automated process of: using a dataset to choose next experiments; adding the experiment results to the dataset, and repeating this process iteratively. We used active discovery to identify several catalysts for CO_2 reduction and H_2 evolution. We then work with collaborators to find experimental evidence showing that one of the candidates, CuAl , can reduce CO_2 to ethylene with selectivity of up to 80%. Next, we improve our active discovery process by creating a multiscale model that predicts macro-scale catalyst performance from atomic-scale DFT functional theory and machine-learned predictions. Lastly, we combine this model with a multiscale sampling strategy for selecting calculations, and we show how this strategy can be used to discover catalysts even more efficiently.

Thesis Supervisor: Zachary W. Ulissi
Title: Assistant Professor

Acknowledgments

I acknowledge the various institutions that have given me financial support during my studies, including: the Department of Chemical Engineering at Carnegie Mellon University; the Graduate Student Association at Carnegie Mellon University; the Neil and Jo Bushnell fellowship; the North American Catalysis Society; and Lawrence Livermore National Lab. I also acknowledge my doctoral committee, including Lord Professor Andrew Gellman, Associate Professor Chrysanthos Gounaris, and Assistant Professor Aaditya Ramdas. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

I also thank all of my collaborators, peers, and colleagues. Special thanks are given to: My team members Pari Palizhati and Muhammed Shuaibi for helping me support the group and for commiserating with me; to Seoin Back for teaching me about catalysis and DFT; and to Willie Neiswanger for teaching me about iterative design of experiments. Additionally, I wish to highlight my deep appreciation for the mentors in my life that have helped me grow into the scientist and person I am today, including Babatunde Ogunnaike, Lauren (Lupinski) DeHart, Jeremy Trethewey, Varghese Maniyatte, and Mark Edmundson. I note especially how thankful I am to add Zack Ulissi to this esteemed list of mentors, who has been quite possibly the best advisor that I have ever heard of. Thank you for your patience, your empathy, your selflessness, and your respect. I could never have imagined a better advisor.

I cannot express enough how much I love my entire family, including my parents Duong & Mai, my brothers Bill & Allan, and my wife Zariel. I would never have been able to do any of this without you. Most importantly, I dedicate this thesis to my father-in-law, Apperson, and to my nieces: Liana, Kadence, & Elyse. May Apperson's legacy live on through Zariel and me, and may this legacy leave a better world for Liana, Kadence, and Elyse to live in.

Contents

1	Introduction	43
1.1	Motivation: climate change & sustainability	43
1.2	The role of catalysis in sustainable energy	45
1.3	Catalysis informatics	46
1.4	Research objective	49
2	Dynamic Workflows for Routine Materials Discovery in Surface Science	51
2.1	Abstract	51
2.2	Introduction	52
2.3	Review of informatics tools for inorganics and surfaces	53
2.3.1	Databases (DBs)	53
2.3.2	Workflow Managers (WFMs)	54
2.3.3	Surrogate Models (SMs)	55
2.3.4	Combining DBs, WFMs, and SMs	56
2.4	Generalized Adsorption Simulations in Python: GASpy	58
2.4.1	Dynamic Workflows via Database Queries	58
2.4.2	Combining workflow managers, databases, and surrogate modeling	62
2.4.3	Application of GASpy	63
2.5	Perspective on informatics tools	65
2.5.1	Considerations when using informatics tools	65
2.5.2	Considerations when building informatics tools	69

2.6	Conclusions	70
3	Active learning across intermetallics to guide discovery of electro-catalysts for CO₂ reduction and H₂ evolution	73
3.1	Abstract	73
3.2	Introduction	74
3.3	Results	76
3.3.1	Framework Construction	76
3.3.2	Enumeration of search space	77
3.3.3	Active learning optimization of catalyst descriptors	78
3.3.4	Performance of active learning optimization	80
3.3.5	Discovering potential intermetallic catalysts for CO ₂ reduction	83
3.3.6	Discovering trends in CO ₂ reduction	83
3.3.7	Discovering potential catalysts for H ₂ evolution	87
3.4	Conclusions	88
3.5	Methods	89
3.5.1	Enumerating search space	89
3.5.2	Addressing models with a variable number of features	90
3.5.3	Calculating optimal adsorption energies	90
3.5.4	Calculating adsorption energies with DFT	90
3.5.5	Automating DFT calculations	91
4	Accelerated discovery of CO₂ electrocatalysts using active machine learning	93
4.1	Abstract	93
4.2	Introduction	94
4.3	Methods	97
4.3.1	Experimental methods	97
4.3.2	Computational methods	100
4.4	Results	100
4.4.1	Initial catalyst exploration	100

4.4.2	System optimization	104
4.4.3	System characterization	104
4.5	Conclusions	106
5	Methods for comparing uncertainty quantifications for material prop- erty predictions	107
5.1	Abstract	107
5.2	Introduction	108
5.3	Methods	109
5.3.1	Dataset information	109
5.3.2	Regression methods	112
5.3.3	Performance metrics	116
5.4	Results	119
5.4.1	Illustrative examples	119
5.4.2	Summary results	121
5.5	Conclusions	128
6	Computational catalyst discovery: Active classification through my- opic multiscale sampling	131
6.1	Abstract	131
6.2	Introduction	132
6.3	Methods	135
6.3.1	Multiscale Modeling	135
6.3.2	Sampling Strategy	139
6.3.3	Active Learning Stopping Criteria	143
6.3.4	Management of Data Queries	144
6.3.5	Estimating Performance through Simulation	145
6.4	Results	148
6.4.1	Performance on Synthesized Data	148
6.4.2	Performance on Density Functional Theory (DFT) Data . . .	150
6.4.3	Recommended diagnostics	152

6.5	Conclusions	154
7	Conclusions	157
A	Supplementary Information for Chapter 3	179
A.1	Supplementary Figures	179
A.2	Supplementary Notes	181
A.2.1	Supplementary Note 1	181
A.2.2	Supplementary Note 2	182
A.2.3	Supplementary Note 3	186
A.3	Supplementary Methods	187
A.3.1	Calculation of free energies	187
A.3.2	Previous surrogate modeling methods	190
A.3.3	t-SNE methods	192
A.4	Supplementary Tables	193
B	Supplementary Information for Chapter 4	223
B.1	Computational Methods	223
B.1.1	DFT/ML screening methods	223
B.1.2	DFT settings for screening	225
B.1.3	Calculating the optimal ΔE_{CO} value	225
B.1.4	Method for creating 2-dimensional activity & selectivity volcanoes	227
B.1.5	Method for creating t-SNE diagrams	228
B.1.6	DFT calculations	229
B.2	Experimental Methods	230
B.2.1	Preparation of evaporated Cu on gas diffusion layers (GDLs) .	230
B.2.2	Preparation of ion-implanted Al-on-Cu on GDLs	231
B.2.3	Preparation of evaporated-etched Al-on-Cu on GDLs	232
B.2.4	Preparation of de-alloyed Cu-Al and nanoporous Cu on GDLs	232
B.2.5	Preparation of de-alloyed Cu-Al on polytetrafluoroethylene (PTFE) substrates	233

B.2.6	Characterizations	233
B.2.7	Electrochemical reduction of CO ₂	235
B.2.8	The stability of Al for CO ₂ electroreduction under our testing condition	249
B.2.9	In situ and ex situ X-ray absorption studies	272
C	Supplementary Information for Chapter 5	285

List of Figures

1-1	Example of a electrochemical cell that uses solar energy to convert CO ₂ into hydrocarbon fuel. (Credit: Clarissa Towle/Berkeley Lab)	44
1-2	Illustration of how DFT can be used to calculate adsorption energies of reaction intermediates. These adsorption energies are useful for performing microkinetic modeling, which provides vital reaction information. Figure obtained directly from Peterson <i>et al.</i> [143].	46
1-3	A set of activity and selectivity volcanoes for CO ₂ reduction as a function of CO adsorption energy (G_{CO^*}). These are used to predict catalyst performance from relatively simple descriptors (e.g., G_{CO^*}) instead of performing entire microkinetic modeling analyses. Figures obtained directly from Liu <i>et al.</i> [105].	47
1-4	Example of an “active” method whereby a dataset is used to decide which datum to query next, after which the datum is obtained and added to the dataset. This cyclical process is continued iteratively until some stopping criteria are met that indicate some goal has been achieved (e.g., a surrogate model has been built, a function has been optimized, or some data were discovered). Reproduced from Settles.[163]	48
2-1	Literature approaches of various combinations of databases, workflow managers, and surrogate models.	56

2-2	GASpy workflow in the form of a directed acyclic graph. Each box is a task, and all tasks are managed by Luigi. The colors of the tasks indicate which package that Luigi uses to execute the task. Triggering any single task will cause Luigi to trigger all necessary upstream tasks. Thus querying the database for the adsorption energy of a specific site will automatically trigger any previously incomplete tasks.	60
2-3	Active learning/sample-point refinement process for GASpy. Green tasks are performed with Mongo and gray tasks are performed with Python. The final task here, “Query GASpy”, is the same final task shown in Figure 2-2. This active learning/sample-point refinement process is repeated daily.	64
2-4	Summary of CO ₂ reduction results from GASpy. a , The number of intermetallic surfaces with near optimal ΔE_{CO} values for CO ₂ RR identified as a function of time. b , Latent space visualization using t-SNE [113] of all the adsorption sites simulated with DFT. Proximity in this reduced space indicates similarity in the structures of the adsorption sites. Both figures were reproduced with permission from Tran & Ulissi.[171]	66
2-5	Estimated difficulty for automating a type of task vs. the field’s current scientific understanding. In general, it is more difficult to create informatics tools for applications where we do not yet have systematic and generic methods developed.	67
2-6	Various informatics and data science tools and how they interact with each other. All of these tools are supported by a foundation of good coding practices.	68

3-1	Workflow for automating theoretical materials discovery. a , The experimental workflow for finding catalysts is accelerated by b , a DFT workflow for screening the catalysts <i>ab initio</i> . c , Conventional workflows (blue) require scientific intuition to select candidates for DFT screenings, while d , our workflow (red) uses machine learning to select candidates systematically and automatically. Red text outlines the framework that this study used to perform active machine learning/surrogate-based optimization.	77
3-2	Fingerprint of coordination site. Adsorption sites are reduced to numerical representations, or fingerprints, and these fingerprints are used as model features by TPOT[135] to predict ΔE_{CO} . Z is the atomic number of an element; χ is the Pauling electronegativity; CN is the number of atoms of an element within a shell; and $\left(\widetilde{\Delta E}\right)$ is the median adsorption energy between the adsorbate and the pure element.	79
3-3	Identification of surfaces with near-optimal ΔE_{CO} values for CO₂RR. a , Both the number of near-optimal surfaces identified and violin plots of the absolute error in predicting ΔE_{CO} as a function of time. The outer shells of the violins bound all data; narrow vertical lines bound 95% of the data; thick vertical lines bound 50% of the data; and white dots represent medians. Apparently missing months are not shown because no data were collected during those months. b , The normalized distribution of the low-coverage, DFT-calculated CO adsorption energies of all of the DFT-analyzed surfaces in this study. The sub-distribution for copper is also illustrated in orange. Dashed lines indicate the ± 0.1 eV range around the optimal ΔE_{CO} value of -0.67 eV. c , Surfaces whose low-coverage CO adsorption energies have been calculated and verified with DFT. d , Surfaces whose low-coverage CO adsorption energies have been calculated only by the machine learning models.	82

3-4	CO₂ reduction activity map for bimetallics. Visualization of two component intermetallics whose surfaces have low-coverage CO adsorption energy (ΔE_{CO}) values inside the range of $[-0.77, -0.57]$ eV. White shading indicates an absence of any enumerated surfaces; grey shading indicates that all ΔE_{CO} values are outside the range of $[-0.77, -0.57]$ eV; and colored shading indicates possible activity. The ΔE_{CO} values used to create the upper half of this figure were calculated by DFT, and the values used to create the bottom half were calculated by the surrogate ML model.	84
3-5	Active site motif analysis. Latent space visualization using t-SNE [113] of all the adsorption sites simulated with DFT. Proximity in this reduced space indicates similarity in the structures of the adsorption sites. Black points represent sites that bind too strongly; dark purple points represent sites with optimal binding; and purple points indicate sites that bind too weakly. Stronger binding sites are overlaid on top of weaker binding sites to indicate that the stronger sites have a greater influence on activity than weaker sites due to their greater thermodynamic stability. We labeled dark purple clusters/materials, because we expect them to be better candidates for further investigation and experimentation.	86
3-6	Analysis of results for HER performance. a , t-SNE [113] visualization of all the adsorption sites simulated with DFT. Similar to Figure 3-5, stronger binding sites are overlaid on top of weaker binding sites, and dark purple clusters/materials are labeled because we expect them to be better candidates for further investigation and experimentation. b , Normalized distribution of low-coverage ΔE_{H} values calculated by our DFT workflow. Dashed lines indicate the ± 0.1 eV range around the optimal ΔE_{H} value of -0.27 eV.	87

4-1 Screening of Cu and Cu based compounds using computational methods. a, A 2-dimensional activity volcano plot for CO₂ reduction. b, A 2-dimensional selectivity volcano plot for CO₂ reduction. CO and H adsorption energies in Figs. a and b were calculated with DFT. Points in yellow are average adsorption energies of monometallics; points in green are average adsorption energies of copper alloys; and points in magenta are average, low-coverage adsorption energies of Cu-Al surfaces. c, t-SNE[113] representation of *ca.* 4,000 adsorption sites that we performed DFT calculations on with Cu-containing alloys. The Cu-Al clusters are labeled numerically. d, Representative coordination sites for each of the clusters labeled in the t-SNE diagram. Each site archetype is labeled by the stoichiometric balance of the surface, i.e., Al heavy, Cu heavy, or balanced, and the binding site of the surface. . 96

4-2 Schematic and characterization of de-alloyed Cu-Al catalyst. a, Schematic of a de-alloyed nanoporous Cu-Al catalyst on a gas diffusion layer for CO₂ electroreduction. b, Scanning electron microscopy (SEM) and high angle angular dark field-scanning transmission electron microscopy (HAADF-STEM) images of de-alloyed Cu-Al catalyst before (left) and after (right) 5 hours CO₂ electroreduction at an applied current density of 600 mA cm⁻² in flow cells. The scale bars for SEM images are 500 nm (top-left) and 200 nm (top-right). The scale bars for TEM images are 200 nm (bottom-left) and 100 nm (bottom-right). c, HAADF-STEM image, and electron energy loss spectroscopy (EELS) spectra of the de-alloyed Cu-Al catalyst. #1, 2, 3 curves in the EELS spectra represent the EELS results measured at #1, 2, 3 areas in the corresponding HAADF-STEM image. Al, Al₂O₃, CuO, Cu₂O and Cu EELS results are plotted as references. The scale bar is 5 nm. 99

4-3	CO ₂ electroreduction performance on de-alloyed Cu-Al, porous Cu and deposited Cu catalysts on C-GDL substrates in a 1 M KOH electrolytes. a, C ₂ H ₄ production partial current density vs. potential with de-alloyed Cu-Al, nanoporous Cu and evaporated Cu catalysts. b, Faradaic efficiencies for gaseous products with de-alloyed Cu-Al catalysts at different applied current densities and with nanoporous Cu and evaporated Cu catalysts at a constant current density of 600 mA cm ⁻² obtained using chronopotentiometry. c, Faradaic efficiencies for all products at an applied current density of 600 mA cm ⁻² with 17 de-alloyed Cu-Al samples measured. d, Half-cell power conversion efficiency of C ₂ H ₄ with de-alloyed Cu-Al catalysts at different applied current densities and with nanoporous Cu and evaporated Cu catalysts at a constant current density of 600 mA cm ⁻² obtained using chronopotentiometry.	102
4-4	CO ₂ electroreduction performance on de-alloyed Cu-Al catalysts on PTFE substrates in alkaline electrolytes at different pH values. a, C ₂ H ₄ production partial current density vs. potential with de-alloyed Cu-Al in 0.3, 1, 3 and 10 M KOH electrolytes. b, Faradaic efficiencies for gaseous products with its corresponding C ₂ H ₄ power conversion efficiencies of the de-alloyed Cu-Al catalysts in the different electrolytes and at different applied current densities. c, The CO ₂ electroreduction stability of the carbon NPs/de-alloyed Cu-Al/PTFE electrode in a 1 M KOH electrolyte at an applied current density of 400 mA cm ⁻² . Left axis: potential (V vs. RHE) vs. time (s), right axis: C ₂ H ₄ Faradaic efficiency (%) vs. time (s). d, The CO ₂ electroreduction stability of the carbon NPs/de-alloyed Cu-Al/PTFE electrode in a 3 M KOH electrolyte at an applied current density of 150 mA cm ⁻² . Left axis: potential (V vs. RHE) vs. time (s), right axis: C ₂ H ₄ Faradaic efficiency (%) vs. time (s).	103

5-1	Overview of proposed procedure for judging the quality of models with uncertainty estimates. First and foremost, the models should be accurate. Second, the models should be “calibrated”, which means that their uncertainty estimates should be comparable with their residuals. Third, the models should be “sharp”, which means that their uncertainty estimates should be low. Lastly, the models should be “disperse”, which means that the distribution of the uncertainty estimates should be wide. This study demonstrates how to visualize and quantify these characteristics so that different methods of uncertainty quantification (UQ) can be compared objectively.	110
5-2	Overview of the various UQ methods we investigated in this study. ΔE represents DFT-calculated adsorption energies; $\Delta \hat{E}$ represents Machine Learning (ML)-predicted adsorption energies; UQ represents ML-predicted uncertainty quantifications; μ represents the mean of a sample of points; σ represents the standard deviation of a sample of points; ϵ represents the residuals between DFT and ML; and $\hat{\epsilon}$ represents the residuals between ML-predicted ϵ and the actual ϵ	113
5-3	Results of the crystal graph convolutional Neural Network (NN) ensemble. Each figure here was created with the test set of 8,289 points.	120
5-4	Calibration curves and parity plots of an overconfident NN ensemble, an underconfident Gaussian Process (GP), and better-calibrated Convolution-Fed Gaussian Process (CFGP). The vertical uncertainty bands in the parity plots indicate ± 2 standard deviations in the uncertainty predictions of each model. For clarity, we sampled only 20 points of the 8,289 test points to put in the parity plots. It follows that relatively overconfident models would have more points with uncertainty bands that do not cross the diagonal parity line; relatively underconfident models would have more points that cross the diagonal parity line; and a well-calibrated model would have <i>ca.</i> 19 out of 20 points cross the parity line.	122

5-5	Parity plots for all UQ methods used in this study. Shading plots were used in lieu of scatter plots because the large number of test points (8,289) obfuscated patterns. Darker shading indicates a higher density of points. Logarithmically scaled shading was used to accentuate outliers. The dashed, diagonal lines indicate parity.	123
5-6	Calibration curves for all UQ methods used in this study. Dashed, blue lines indicate perfect calibration while solid orange lines indicate the experimental calibration of the test set. The blue, shaded area between these lines is defined as the miscalibration area.	124
5-7	Distribution plots of the ML-predicted standard deviations for each method. Sharpness values are indicated by vertical lines.	125
6-1	Illustration of Myopic Multiscale Sampling (MMS). Given a database of DFT-calculated adsorption energies (a), we train a ML model to predict adsorption energies (b). Then we use those adsorption energies to estimate activities of catalyst surfaces (c), which we then use to estimate the activities of the bulk catalysts (d). Then we choose which catalyst to sample next (e); then we choose which surface on the catalyst to sample (f); then we choose which site on the surface to sample (g); then we perform DFT of that site to add to the database (h). This procedure is repeated continuously with the goal of classifying all catalysts as either “relatively active” or “relatively inactive”.	136
6-2	Multiscale modeling strategy for estimating the activity of a catalyst. For each adsorption site, we obtain a machine-learned estimate of its adsorption energy along with uncertainty. Then we aggregate the energy distributions for all sites within each surface through a minimum operator across sites. Next we transform the energy distributions for all surfaces into activities using a Sabatier relationship. Finally we average all the surface activities to obtain an estimate of overall catalyst activity.	138

6-3	Myopic Multiscale Sampling (MMS) overview. At the highest level, we choose a catalyst to query using level-set estimation—to be specific, we use the probability of incorrect classification as our acquisition function. At the medial level, we choose a surface of the catalyst using uncertainty sampling. At the lowest level, we choose a site on the surface using Bayesian optimization to find the lowest energy site. . .	141
-----	---	-----

6-4	Performance and convergence results for the simulations on the synthesized dataset. a. $F1$ score of the multiscale model during simulation of the synthesized data. For clarity of visualization, we plotted the rolling average of the $F1$ score using a window of 20 batches. b. predicted change in $F1$ score ($\Delta\hat{F}$) of the multiscale model during simulation of the synthesized data. For clarity of visualization, we plotted the rolling average of $\Delta\hat{F}$ using a window of 40 batches (excluding the Myopic Multiscale Sampling (MMS) null line, where no averaging was done). RS represents “random search” while MMS represents Myopic Multiscale Sampling.	149
-----	---	-----

6-5	Performance and convergence results for the simulations on the the Generalized Adsorption Simulator for Python (GASpy) dataset. a. $F1$ score of the multiscale model during simulation of the GASpy dataset. b. $\Delta\hat{F}$ of the multiscale model during simulation of the synthesized data. RS represents “random search” while MMS represents Myopic Multiscale Sampling.	151
-----	--	-----

6-6	Example of diagnostic plots that we recommend monitoring during an active discovery campaign: a. predicted change in $F1$ score ($\Delta\hat{F}$); b. residuals between the real data and the surrogate model's predictions; c. expected calibration error[170] of the surrogate model; d. the predicted uncertainties of surrogate model in the form of the predicted standard deviation (σ); and e. the negative-log-likelihood of the surrogate model.[170] These results were simulated by using the Myopic Multiscale Sampling (MMS) method with the Convolution-Fed Gaussian Process (CFGP) model on the GASpy dataset. For clarity of visualization, we plotted rolling averages of all values in this figure using a window of 100 queries (excluding the $\Delta\hat{F}$ values, where no averaging was done)	153
-----	--	-----

A-1	Elements considered in this screening. Shading of an element indicates its inclusion in this screening study.	179
-----	--	-----

A-2	Identification of surfaces with near-optimal ΔE_H values for HER. a, Distribution of the number of near-optimal surfaces identified. b, The normalized distribution of the low-coverage H adsorption energies of all of the surfaces enumerated by this study. Dashed lines indicate the ± 0.1 eV range around the optimal ΔE_H value of -0.27 eV. c, Surfaces whose low-coverage H adsorption energies have been calculated and verified with DFT. d, Surfaces whose low-coverage H adsorption energies have been calculated only by the machine learning models.	180
-----	--	-----

A-3	H₂ evolution efficiency map for bimetallics. Visualization of two component intermetallics whose surfaces have low-coverage H adsorption energy (ΔE_H) values inside the range of $[-0.37, -0.17]$ eV. White shading indicates an absence of any enumerated surfaces; grey shading indicates that all ΔE_H values are outside the range of $[-0.37, -0.17]$ eV; and colored shading indicates possible efficiency. The ΔE_{CO} values used to create the upper half of this figure were calculated by DFT, and the values used to create the bottom half were calculated by the surrogate model.	181
A-4	Surrogate modeling performance of final model. Performance metrics for 16 models created from 90/10 train/test splits on all DFT data for predicting (a) ΔE_{CO} (n=19,644) and (b) ΔE_H (n=23,141). Black points represent results for each of the 16 models, bar heights indicate average values, and red lines indicate average values ± 1 standard deviation. Parity plots when predicting catalyst performance for (c) CO ₂ RR and (d) HER. Parity plots when predicting (e) ΔE_{CO} and (f) ΔE_H . Darker shading indicates a higher density of points; dashed lines indicate parity; and dotted lines indicate 95% prediction intervals—i.e., parity $\pm 2 \times$ MAE. The parity plots in (c-f) were generated using a model that was trained using a 100/0 train/test split. . .	184
A-5	Learning curves. Error metrics vs. size of training set when calculating adsorption energies for (a) CO and (b) H. RMSE is the root mean squared error; MAE is the mean absolute error; and MAD is the median absolute deviation.	186
A-6	Benchmarking of DFT results. The adsorption energies (ΔE) of various adsorbate/surface combinations calculated by our framework vs. the energies calculated by various literature sources.[7, 164] The diagonal line indicates parity.	187

A-7	Alternative method for modeling adsorption energy. a, Alternative fingerprinting method. Each item in the “coordination” vector represents the coordination number for a particular element, e.g., Pt or Ni. Each item in the “neighbors’ coordination” array represents this same coordination vector, but for each of the adsorbate’s neighbors. Note that this example is for illustrative purposes only. The vectors and arrays actually used contained enough items to represent 31 different elements. b, Old method for performing regression. Coordination was used by TPOT to estimate adsorption energy, then the residuals of that model were combined with the neighbors’ coordinations and used by a Gaussian process regressor to estimate residuals of the first model. Summation of the results of the first and second models yielded the final estimate.	191
-----	--	-----

B-1	An example of the numerical encoding of an adsorption site. Vectors are created for each element present within the first and second neighbouring shells of CO. Each vector contains the atomic number of the element (Z), the Pauling electronegativity of the element (χ) the number of atoms of that element within each respective shell (CN), and the median monometallic adsorption energy of CO on that element ($\Delta\tilde{E}$). Color codes for elements: orange is Cu, light gray is Al, dark gray is C, and red is O.	236
-----	--	-----

B-2	A simplified example of a numerical representation of a coordination site which was used for t-SNE analysis. Each item in the “coordination” vector represents the coordination number for a particular element, e.g., Pt or Ni. Each item in the “neighbours” coordination” array represents this same coordination vector, but for each of the adsorbate’s neighbours. Note that this example is for illustrative purposes only. The vectors and arrays actually used contained enough items to represent 31 different elements, totalling in 31×32 features. Color codes for elements: green is Ni, light gray is Pt, black is C, and red is O. . .	236
B-3	Schematic for the synthesis of different catalysts on gas diffusion layers. a, evaporated Cu. b, ion-implanted Al-on-Cu. c, evaporated-etched Al-on-Cu. d, de-alloyed Cu-Al catalysts.	237
B-4	Morphologies of evaporated Cu catalysts on gas diffusion layers. a, Top-view SEM images before CO ₂ electroreduction. b, Top-view SEM images after 5 hours CO ₂ electroreduction in 1 M KOH at an applied current density of 600 mA cm ⁻² in a flow cell.	238
B-5	Morphologies of ion-implanted Al-on-Cu catalysts on gas diffusion layers. a, Top-view SEM images before CO ₂ electroreduction. b, Top-view SEM images after 5 h CO ₂ electroreduction in 1 M KOH at an applied current density of 600 mA cm ⁻² in a flow cell.	239
B-6	Morphologies of evaporated and evaporated-etched Al-on-Cu samples on gas diffusion layers. a, Top-view SEM images of the as-evaporated Al-on-Cu samples. b, Top-view SEM images of the evaporated-etched Al-on-Cu catalyst before CO ₂ electroreduction. c, Top-view SEM images of the evaporated-etched Al-on-Cu catalyst after 5 hours CO ₂ electroreduction in 1 M KOH at an applied current density of 600 mA cm ⁻² in a flow cell.	240
B-7	Schematic of the cathode side in a flow-cell configuration.	241

B-8	CO ₂ electroreduction performances on pure Cu, ion-implanted Al-on-Cu, and evaporated-etched Al-on-Cu catalysts. a, Faradaic efficiencies of gaseous products on pure Cu, ion-implanted, and evaporated-etched Al-on-Cu obtained from chronopotentiometry tests at an applied current density of 600 mA cm ⁻² . b, C ₂ H ₄ production partial current density versus potential on pure Cu, ion-implanted, and evaporated-etched Al-on-Cu.	241
B-9	Auger electron spectroscopic analysis of an ion-implanted Al-on-Cu catalyst. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.	242
B-10	Auger electron spectroscopic analysis of an evaporated-etched Al-on-Cu catalyst. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.	243
B-11	Pourbaix diagrams. Pourbaix diagrams of a, Cu and b, Al at ionic concentrations of 1 μ M. The potentials versus the standard hydrogen electrode (V vs. SHE) can be converted to the reversible hydrogen electrode scale (V vs. RHE) according to the Nernst equation, $V \text{ vs. RHE} = V \text{ vs. SHE} + 0.059 \times \text{pH}$. According to the Pourbaix diagrams, Cu metal is cathodically protected at potentials more negative than ca. -0.4 VSHE which corresponds to ca. 0.43 VRHE at pH 14 (1 M KOH) and ca. 0.48 VRHE at pH 15 (10 M KOH). Al metal is cathodically protected at potentials more negative than -2.3 VSHE, which corresponds to -1.47 VRHE at pH 14 and -1.41 VRHE at pH 15.	244
B-12	EDX analysis of an as-prepared ion-implanted Al-on-Cu sample before CO ₂ electroreduction. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations.	245

B-13 EDX analysis of an ion-implanted Al-on-Cu sample after 5 hours CO ₂ electroreduction at 600 mA cm ⁻² (-1.8 to -2.0 VRHE) in the 1 M KOH electrolyte. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Potassium is observed on the surface after the reaction. . . .	246
B-14 EDX analysis of an evaporated-etched Al-on-Cu sample before CO ₂ electroreduction. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Chlorine is observed on the surface due to the use of HCl solution in the etching process.	247
B-15 EDX analysis of an evaporated-etched Al-on-Cu sample after 5 hours CO ₂ electroreduction at 600 mA cm ⁻² (-1.8 to -2.0 VRHE) in the 1 M KOH electrolyte. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Potassium was observed on the surface after the reaction.	248
B-16 EDX analysis of a pure Cu catalyst after half-hour CO ₂ electroreduction in 1 M KOH with 1 mM Al(OH) ⁴⁻ . a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Fluorine was from the PTFE/carbon layer in the gas diffusion electrode. Al was electrodeposited on Cu at an applied current density of 600 mA cm ² . Therefore, dissolving Al into solution was avoided.	250
B-17 Auger electron spectroscopic analysis of a pure Cu catalyst after 0.5-hour CO ₂ electroreduction at 600 mA cm ² in 1 M KOH with 1 mM Al(OH) ⁻ . a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.	251
B-18 CO ₂ electroreduction performance. Faradaic efficiencies for gaseous products with a pure Cu catalyst at a constant current density of 600 mA cm ² in 1 M KOH with the presence of 1 mM Al(OH) ⁴⁻ obtained from Chronopotentiometry tests.	252

B-19	EDX analysis of a pure Cu after immersing in 1 M KOH with 1 mM Al(OH)_4^- for 0.5 hour. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. We manually selected Al which was automatically marked in red by EDX software (Esprit 2.1) with a large error of 43.81%, indicating Al was actually out of the EDX detection limit. In SEM image in Figure B-14a, the morphology was also changed. Most of Cu dissolved into 1 M KOH leaving thin Cu(OH)_2 nanowires to be dissolved.	253
B-20	Top-view SEM images of the de-alloyed Cu-Al. a-b, De-alloyed Cu-Al catalyst prepared on a C-GDL substrate by physical evaporation and chemical etching. c-d, De-alloyed Cu-Al catalyst prepared on a PTFE substrate by co-sputtering and chemical etching.	254
B-21	EDX analyses in TEM of de-alloyed Cu-Al catalysts. a, as-prepared de-alloyed Cu-Al catalyst. b, de-alloyed Cu-Al catalyst after 5 hours CO_2 electroreduction in 1 M KOH at 600 mA cm^2 in a flow-cell configuration.	254
B-22	EDX analysis in SEM of the as-prepared de-alloyed Cu-Al catalyst before CO_2 electroreduction. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Chlorine was detected because of the use of a 5 wt% HCl solution in the etching process to prepare a de-alloyed Cu-Al catalyst.	255
B-23	EDX analysis in SEM of a de-alloyed Cu-Al catalyst after 5 hours CO_2 electroreduction at 600 mA cm^2 (-1.8 to -2.0 VRHE) in 1 M KOH. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Potassium was detected because of the use of KOH electrolyte in CO_2 electroreduction.	256
B-24	HAADF-STEM with EELS mapping images of Cu and Al of a de-alloyed Cu-Al catalyst after 5 hours CO_2 electroreduction. A nanoporous structure was observed with the de-alloyed Cu-Al after the CO_2 electroreduction.	257

B-25 STEM-ADF image of the de-alloyed Cu-Al catalyst. The observation was performed in the same area where we performed EELS analysis (Figure 4-2c)	257
B-26 Auger electron spectroscopic analysis of an as-prepared de-alloyed Cu-Al catalyst. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.	258
B-27 Auger electron spectroscopic analysis of a de-alloyed Cu-Al catalyst after 5 hours CO ₂ electroreduction at 600 mA cm ² in 1 M KOH. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al. . . .	259
B-28 ML and DFT analyses of the (111) and (100) facets across varying ratios of Al concentrations. a, b, The distribution of ΔE_{CO} values for the adsorption sites on Cu (111) and (100) surfaces with different amounts of Al replacement on the top layer of atoms.	260
B-29 Reaction Gibbs free energy diagram. Reaction Gibbs free energy diagram from adsorbed CO ₂ to OC ₂ H ₄ , an intermediate to C ₂ H ₄ , and to CHOCH ₃ , an intermediate to C ₂ H ₅ OH, on the pure Cu (111) surface (blue lines) and ML predicted 12% Al incorporated Cu (111) surface (red lines). The ML predicted structure of 12% Al incorporated Cu (111) is inserted in the lower left.	260
B-30 Reaction Gibbs free energy diagram. Reaction Gibbs free energy diagram from adsorbed CO ₂ to OC ₂ H ₄ , an intermediate to C ₂ H ₄ , and to CHOCH ₃ , an intermediate to C ₂ H ₅ OH, on the pure Cu (100) surface (blue lines) and ML predicted 4% Al incorporated Cu (100) surface (red lines). The ML predicted structure of 4% Al incorporated Cu (111) is inserted in the lower left.	261
B-31 A possible pathway with intermediates in the CO ₂ electroreduction. Optimized geometries of CO ₂ reduction intermediates from adsorbed CO ₂ * to OC ₂ H ₄ and CHOCH ₃ over the 12% Al incorporated Cu (111) s	262

B-32	A possible pathway with intermediates in the CO ₂ electroreduction. Optimized geometries of CO ₂ reduction intermediates from adsorbed CO ₂ * to OC ₂ H ₄ and CHOCH ₃ over the 4% Al incorporated Cu (100) s	263
B-33	Initial and final intermediates of forming ethanol (red) and C ₂ H ₄ (blue) on the 12% Al incorporated Cu (111) and 4% Al incorporated (100) surfaces. A possible pathway with intermediates in the CO ₂ electroreduction. According to the previous work reported by Goddard et al.[182], surface water may play an important role to determine the key intermediates that branch toward ethanol and C ₂ H ₄ production especially at high pH conditions. We use their method to calculate the Gibbs free energy barriers of the key intermediates on the 12% incorporated Cu (111) and 4% Al incorporated Cu (100) surfaces. The energy barriers for forming ethanol are 0.343 eV and 0.223 eV larger than that for forming C ₂ H ₄ on the Al incorporated Cu (111) and Cu (100) surfaces, respectively.	264
B-34	De-alloyed Cu-Al with different surface Al concentrations. a, SEM images of de-alloyed Cu-Al with different etching time. b, Al concentrations on surfaces determined by surface-sensitive Auger spectroscopic analysis. Surface Al concentrations could be roughly varied from 4–28% by controlling the etching time. c, Faradaic efficiencies of C ₂ + and C ₂ + /C1 ration with de-alloyed catalysts at different surface Al concentrations.	265
B-35	SEM images of a de-alloyed Cu-Al catalyst soaked in 10 mM CuCl ₂ for 10 minutes. 10 minutes is an optimal time that we used to roughly maintain nanoporosity while replacing surface Cu with Al by the displacement reaction of $2\text{ Al} + 3\text{ CuCl}_2 \rightarrow 2\text{ AlCl}_3 + 3\text{ Cu}$. We also found that nanopores were gradually blocked by the precipitated Cu with a prolonged CuCl ₂ treatment.	265

B-36 Auger electron spectroscopic analysis of a de-alloyed Cu-Al catalyst soaked in 11 mM CuCl ₂ solution for 0.5 hour. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al. The Al concentration calculated from Al narrow-scan spectrum is already close to the AES detecting limit.	266
B-37 CO ₂ electroreduction performance. Faradaic efficiencies for all products at an applied current density of 600 mA cm ² obtained with of a de-alloyed Cu-Al catalyst after soaking in 10 mM CuCl ₂ solution for 10 minutes.	267
B-38 CO ₂ electroreduction stability of the de-alloyed Cu-Al/C-GDL electrode. The CO ₂ electroreduction activity of a de-alloyed Cu-Al/C-GDL electrode at an applied current density of 600 mA cm ² . Left axis: potential (V vs. RHE) vs. time (s), right axis: C ₂ H ₄ Faradaic efficiency (%) vs. time (s)	267
B-39 Configuration of de-alloyed Cu-Al catalysts on (a) C-GDL and (b) PTFE substrates for CO ₂ electrolysis in a flow cell system. Optical images were taken on the backside of the samples to show the hydrophobicity of a, Cu-Al/C-GDL before and after 5 hours stability test in 1 M KOH, and b, Cu-Al/PTFE before and after 50 hours stability test in 3 M KOH with 3M KI) solutions.	268
B-40 Co-sputtering, chemical etching and spray coating to make graphit/carbon NPs/de-alloyed Cu-Al /PTFE substrates.	268
B-41 CO ₂ electroreduction performance with graphite/carbon NPs/de-alloyed Cu-Al/PTFE samples in a 1 M KOH electrolyte. a, Faradaic efficiencies for all products at an applied current density of 400 mA cm ² with 10 distinct samples measured. b, Faradaic efficiency of gas products vs. time during the stability test.	269

B-42 CO ₂ electroreduction performance with graphite/carbon NPs/de-alloyed Cu-Al/PTFE samples in a 3 M KOH and 3 M KI electrolyte. a, Faradaic efficiencies for all products at an applied current density of 150 mA cm ² with 10 distinct samples measured. b, Faradaic efficiency of gas products vs. time during the stability test. c, The CO ₂ electroreduction stability at an applied current density of 150 mA cm ² . Left axis: potential (V vs. RHE) vs. time (s), right axis: C ₂ H ₄ Faradaic efficiency (%) vs. time (s)	269
B-43 ICP-AES results of Cu and Al ions in the testing solution with different the testing time. The de-alloyed Cu-Al catalysts on PTFE was measured at 400 mA cm ² in 1 M KOH solution and at 150 mA cm ² in 3 M KOH solution for 50 hours.	270
B-44 Reaction Gibbs free energy diagram. Reaction Gibbs free energy diagram from adsorbed CO ₂ to C ₂ H ₄ and C ₂ H ₅ OH on Cu (111) facet with (red lines) and without (blue lines) Al incorporation in the subsurface (the grey circle shows the Al atoms' positions). The model of Al incorporated Cu is described in lower left. The energy barrier of the rate determine step (OC-CO dimerization) decreases after Al incorporation.	271
B-45 Reaction Gibbs free energy diagram. Reaction Gibbs free energy diagram from adsorbed CO ₂ to C ₂ H ₄ and C ₂ H ₅ OH on Cu (100) facet with (red lines) and without (blue lines) Al incorporation in the subsurface. The model of Al incorporated Cu is described in lower left. The energy barrier of the rate determine step (OC-CO dimerization) decreases after Al incorporation.	272
B-46 Synchrotron analyses of the de-alloyed Cu-Al catalyst before and after 5, 10, 20 and 30 min. of the CO ₂ electroreduction reactions.	273

B-47	Ex situ and in situ synchrotron analyses of the high-performing de-alloyed Cu-Al catalyst and pure Cu catalyst. a-b, In situ X-ray absorption near-edge structure (XANES) spectra of de-alloyed Cu-Al catalyst before, during and after CO ₂ electroreduction tests and ex situ XANES spectra of Cu foil, Cu ₂ O, CuO. c-d, In situ X-ray absorption near-edge structure (XANES) spectra of pure Cu catalyst before and during CO ₂ electroreduction tests and ex situ XANES spectra of Cu foil.	274
B-48	Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO ₂ reduction. 4% Al incorporated Cu (100) surface with different amounts of Cu-O bonds as shown in configurations 1–5.	275
B-49	The distribution of Cu-O bond length in the Figure B-48.	275
B-50	Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO ₂ reduction. 4% Al incorporated Cu (100) surface with different amounts of Cu-O bonds as shown in configurations 1–5.	276
B-51	The distribution of Cu-O bond length in the Figure B-50.	276
B-52	Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO ₂ reduction. 4% Al incorporated Cu (100) surface with different amounts of Cu-O bonds as shown in configurations 1–5.	277
B-53	The distribution of Cu-O bond length in the Figure B-52.	277
B-54	Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO ₂ reduction. 4% Al incorporated Cu (100) surface with different amounts of Cu-O bonds as shown in configurations 1–5.	278
B-55	The distribution of Cu-O bond length in the Figure B-54.	278

B-56 Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO ₂ reduction. 12% Al incorporated Cu (111) surface with different amounts of Cu-O bonds as shown in configurations 1–5.	279
B-57 The distribution of Cu-O bond length in the Figure B-56.	279
B-58 Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO ₂ reduction. 12% Al incorporated Cu (111) surface with different amounts of Cu-O bonds as shown in configurations 1–5.	280
B-59 The distribution of Cu-O bond length in the Figure B-58.	280
B-60 Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO ₂ reduction. 12% Al incorporated Cu (111) surface with different amounts of Cu-O bonds as shown in configurations 1–5.	281
B-61 The distribution of Cu-O bond length in the Figure B-60.	281
B-62 Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO ₂ reduction. 12% Al incorporated Cu (111) surface with different amounts of Cu-O bonds as shown in configurations 1–5.	282
B-63 The distribution of Cu-O bond length in the Figure B-62.	282
B-64 Activity, selectivity, and coverage maps at -0.5 V applied potential, based on (211) metal scaling relations and microkinetic models. a, Activity of CH ₄ and H ₂ production (in units of log(TOF [1/s]) as a function of ΔE_{CO} vs. ΔE_H . b, Selectivity towards CH ₄ and H ₂ production as a function of ΔE_{CO} vs. ΔE_H . c, Surface coverages of CHO, CO, H, and OH adsorbates as a function of ΔE_{CO} vs. ΔE_H	283

C-1	Corner plot of the residuals of all models. Each subfigure shows the parity between the residuals of pairs of models. Solid contour lines delineate quartiles of the point distribution. Single, faded points indicate parity points in the fourth, least dense quartile of points. Shaded pixels indicate the highest density of points with darker shading indicating a higher density. The figures along the diagonal show histogram distributions of the residuals for each model. All units are in eV.	286
C-2	Corner plot of the residuals of all models. Each subfigure shows the parity between the estimated standard deviations of pairs of models. Solid contour lines delineate quartiles of the point distribution. Single, faded points indicate parity points in the fourth, least dense quartile of points. Shaded pixels indicate the highest density of points with darker shading indicating a higher density. The figures along the diagonal show histogram distributions of the predicted standard deviations for each model. All units are in eV.	287

List of Tables

2.1	Various strengths and weaknesses of various informatics tools alongside the corresponding use-cases that we recommend for each.	71
5.1	Performance metrics for all methods used in this study, which include: Median Absolute Error (MDAE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Relative Percent Difference (MARPD), R^2 correlation coefficient (R^2), miscalibration area (MisCal), calibration error (CalErr), sharpness (Sha), coefficient of variation (C_v), and negative log-likelihood (NLL). The units of Median Absolute Error (MDAE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and sharpness are all in electron volts (eV). The units of Mean Absolute Relative Percent Difference (MARPD) are in %. The miscalibration area, calibration error, coefficient of variation (C_v), and negative log-likelihood (NLL) are unitless.	126
A.1	List of possibly active surfaces for CO_2 reduction. All of the surfaces whose minimum CO adsorption energies were between -0.87 and -0.57 eV. MPID represents the Materials Project ID. Apparently duplicate entries for MPID/Miller index combinations are caused by different, nonidentical cuts along a Miller plane, which yields multiple surfaces for a single Miller index. All energies reported here were calculated by DFT, not with machine learning.	194

A.2	List of possibly efficient surfaces for H₂ evolution. All of the surfaces whose minimum H adsorption energies were between -0.47 and -0.17 eV. MPID represents the Materials Project ID. Apparently duplicate entries for MPID/Miller index combinations are caused by different, nonidentical cuts along a Miller plane, which yields multiple surfaces for a single Miller index. All energies reported here were calculated by DFT, not with machine learning.	202
B.1	The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-48.	275
B.2	The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-50.	276
B.3	The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-52.	277
B.4	The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-54.	278
B.5	The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-56.	279
B.6	The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-58.	280
B.7	The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-60.	281
B.8	The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-62.	284
B.9	Zero-point energy and specific heat of different reactants, intermediates and products on the Al incorporated Cu (111) surface. It corresponds to the data used in Figure B-32.	284
B.10	Entropy and Gibbs free energy of different reactants, intermediates and products on the Al incorporated Cu (111) surface. It corresponds to the data used in Figure B-32.	284

B.11 Zero-point energy and specific heat of different reactants, intermediates and products on the Al incorporated Cu (100) surface. It corresponds to the data used in Figure B-32.	284
--	-----

Chapter 1

Introduction

1.1 Motivation: climate change & sustainability

As humanity's populations continue to grow,[120] our energy demands also rise.[38, 6, 5] The majority of this energy demand has been met through the combustion of fossil fuels, which has increased atmospheric CO₂ concentrations significantly.[114] This increase in CO₂ has exacerbated a global greenhouse effect, increasing average global temperatures at rates that the world's environments, biological ecosystems, and economies have struggled to adapt to.[148, 38] Such destabilization raises concerns regarding the sustainability of our current, fossil-fuel-based energy economy.[114, 43, 148] We need a way to meet our increasing energy demands while mitigating the consequential destabilization of the world's environment, climate, and economies. Thus alternative methods of energy generation and storage are desired to reduce dependence on fossil fuels.

The most abundant source of sustainable energy is the sun.[140] Unfortunately, solar energy is intermittent and must be stored for on-demand use. One possible solution is to store the energy in battery stations that are connected to electrical grids.[31] Energy storage in batteries could address the needs of our power grids, but it does not address the needs of the transportation sector—e.g., automobiles, trains, or planes. This sector requires fuels that are energy dense and easily transported.

One promising strategy is to store sustainable energy in the form of chemical

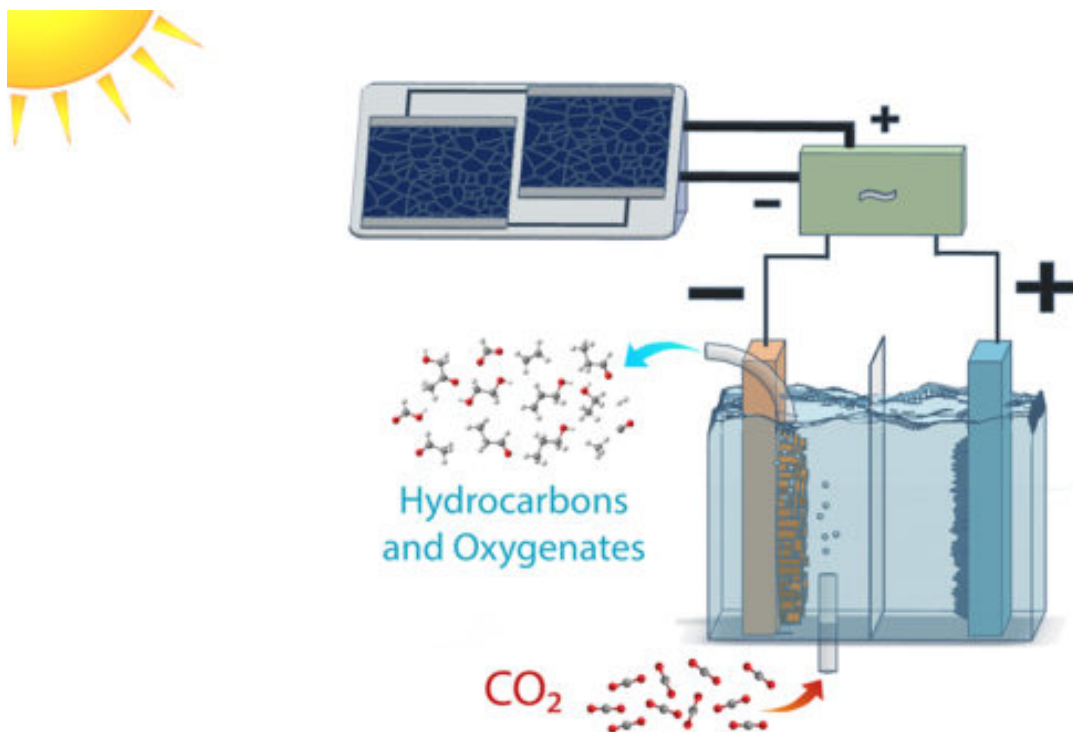


Figure 1-1: Example of a electrochemical cell that uses solar energy to convert CO₂ into hydrocarbon fuel. (Credit: Clarissa Towle/Berkeley Lab)

bonds. For example, we could split water into hydrogen and oxygen and then use the hydrogen to feed fuel cells. Or we could reduce CO₂ with water to create synthetic hydrocarbon fuels (Figure 1-1), which could be fed directly into the current transportation infrastructure for a net-zero carbon emission cycle. These are called solar fuels,[98, 160] and they represent the idea of storing solar/sustainable energy in chemical bonds for on-point use.

This idea of using sustainable energy to create chemical bonds could also be extended beyond the energy storage sector. Why not create solar chemicals as well? We currently rely on fossil fuels as the chemical feed stocks for the chemical industry.[97] This includes production of ammonia, which is used to create fertilizer that supports the world's food supplies,[67] further exemplifying our reliance on a limited resource. Reducing our dependence on fossil fuels by enabling solar fuel and solar chemical production may improve the long-term sustainability of humankind. Thus the goal of this thesis is to contribute to the research community's goal of enabling commercial

production of solar fuels and chemicals.

1.2 The role of catalysis in sustainable energy

Commercial production of solar fuels and chemicals needs to be economically feasible. Such feasibility is governed by the capital and operating costs of the chemical facilities that would produce them. In turn, these capital and operating costs are governed by the rate, selectivity, and thermodynamic efficiency of the reactions. For the reactions that we are interested in—e.g., CO₂ reduction, H₂ evolution, or ammonia synthesis—we do not yet have reaction catalysts that are sufficiently active, selective, efficient, and stable. We have found catalysts that meet some of these criteria for some of these reactions, but we have not yet found catalysts that meet all of the criteria.[160] Thus our current goal is to discover catalysts that are active, selective, efficient, and stable enough to enable commercial production of solar fuels and chemicals.

Originally, catalyst discovery was performed purely through experimentation. But advances in quantum mechanical understanding and computational power and have enabled more sophisticated *ab initio* studies—e.g., DFT.[137] These studies have improved our understanding of catalytic processes and accelerated our ability to discover new catalysts.[134] For example, DFT can be used to calculate the adsorption energies of reaction intermediates for CO₂ reduction to methane (Figure 1-2).[143] These adsorption energies can then be paired with transition state calculations[62] and/or Brønsted-Evans-Polanyi relationships[17] to evaluate microkinetic models,[118] which reveal elementary chemical reaction rates. These reaction rates can elucidate phenomena such as the Sabatier relationship[119] (Figure 1-3), which can be used to quantify reaction performance. Put simply: We can use DFT to estimate catalyst activity, selectivity, efficiency, or stability *ab initio*[133, 119].

Unfortunately, transition-state calculations and subsequent microkinetic modeling are prohibitively expensive to do for new catalysts, both in terms of computational cost and human overhead costs. Approximations can be made to reduce these costs, such as assuming particular reaction pathways or assuming scaling relationships be-

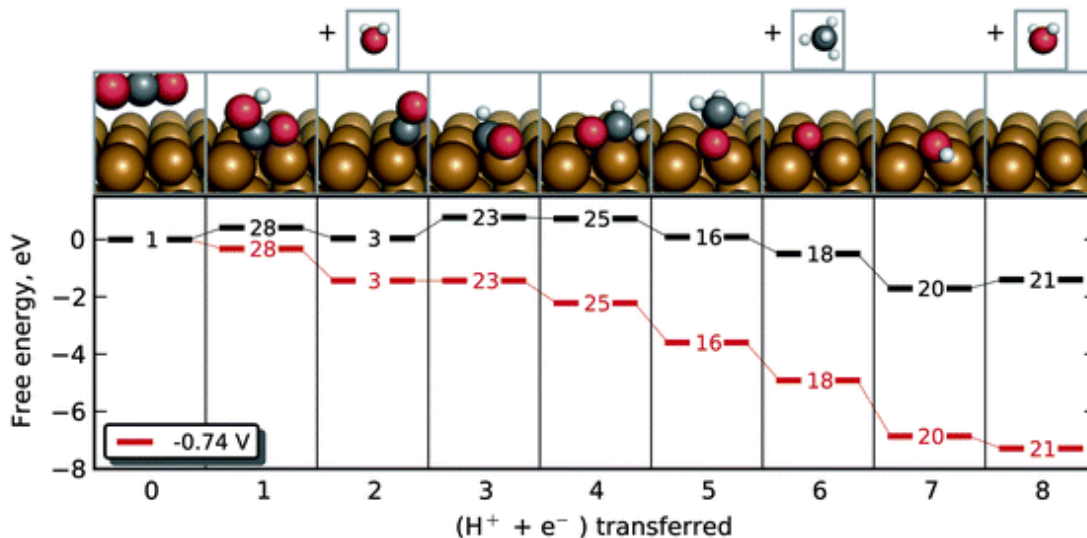


Figure 1-2: Illustration of how DFT can be used to calculate adsorption energies of reaction intermediates. These adsorption energies are useful for performing microkinetic modeling, which provides vital reaction information. Figure obtained directly from Peterson *et al.*[143].

tween intermediate energies and/or transition states.[134, 17] These approximations reduce the fidelity of the results though. Thus there is a tension between the fidelity of the DFT predictions and the breadth at which we can perform them. In other words, we need to balance our ability to exploit the accuracy of DFT and our ability to explore broader search spaces. Given that the solar fuel research community is still in the initial stages of process development, I propose that we focus on exploration rather than exploitation. Therefore we should focus on generating relatively large volumes of coarse data so that we can more quickly generate a wider range of catalysts to screen experimentally.

1.3 Catalysis informatics

Serendipitously, the advancements in computers in the past two decades led to a rise of large-scale data generation and storage across several industries. This proliferation of data has led to a surge in research aimed at analyzing large amounts of data—e.g., ML or informatics. This surge is exemplified by the success of technology companies

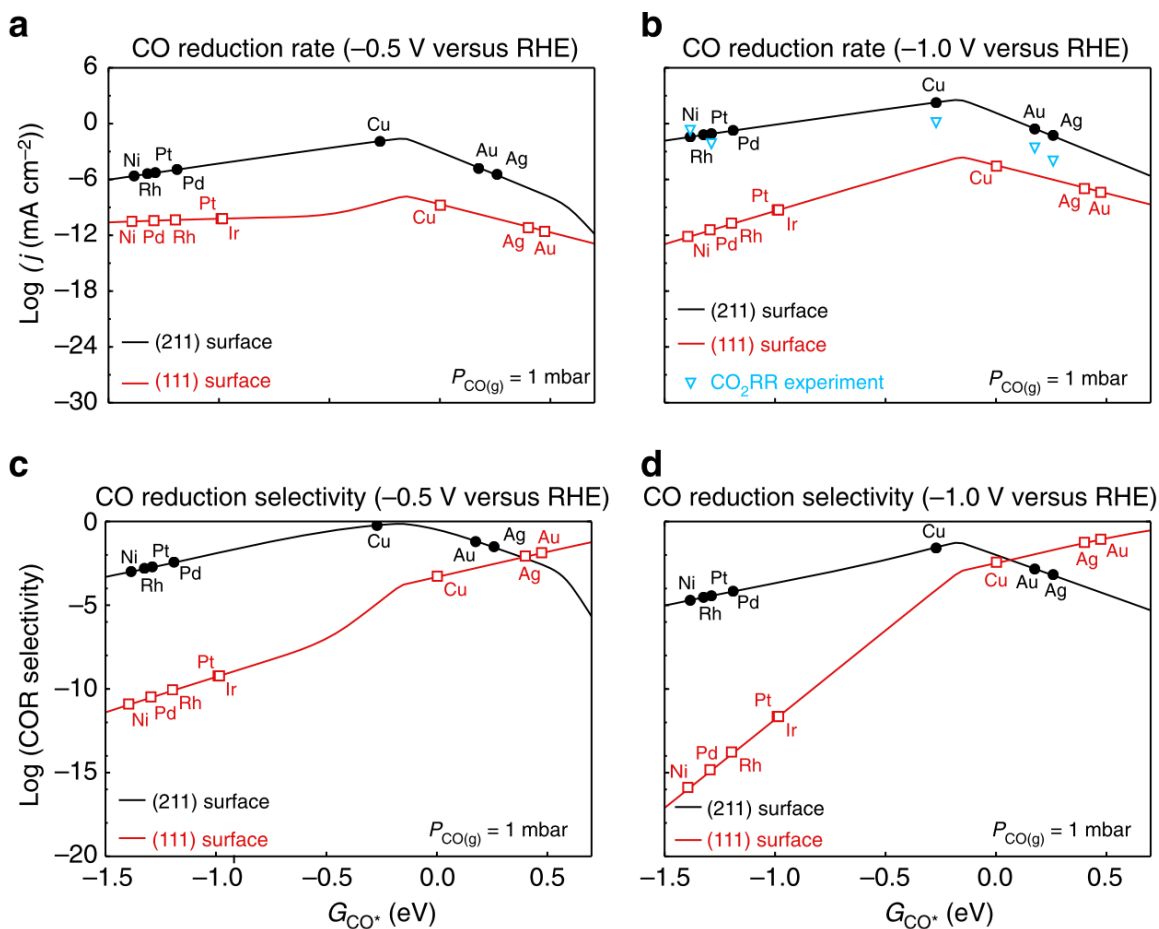


Figure 1-3: A set of activity and selectivity volcanoes for CO₂ reduction as a function of CO adsorption energy (G_{CO^*}). These are used to predict catalyst performance from relatively simple descriptors (e.g., G_{CO^*}) instead of performing entire microkinetic modeling analyses. Figures obtained directly from Liu *et al.*[105].

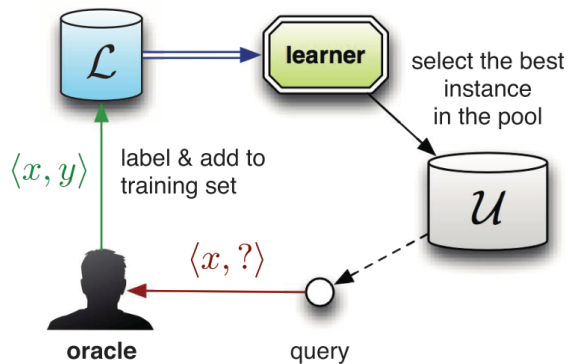


Figure 1-4: Example of an “active” method whereby a dataset is used to decide which datum to query next, after which the datum is obtained and added to the dataset. This cyclical process is continued iteratively until some stopping criteria are met that indicate some goal has been achieved (e.g., a surrogate model has been built, a function has been optimized, or some data were discovered). Reproduced from Settles.[163]

such as Google, Amazon, or Facebook as well as the explosion in the number of participants in conferences such as Neural Information Processing Systems (NeurIPS), International Conference on Learning Representations (ICLR), or International Conference on Machine Learning (ICML). The use of tools and insights generated from the ML and technology communities is called “catalysis informatics”.[117]

There are several applications of catalysis informatics. They include software engineering methods for data generation; database management tools; data analysis; surrogate modeling; and every combination thereof.[147, 117, 155, 150, 55, 156] Note that these methods have enabled automatic data generation, which has then enable a subfield of research on “active” methods. Active methods are defined by their cyclical procedures that involve analyzing data, choosing which data to search for next, obtaining it, re-analyzing the updated dataset, and then continuing this process iteratively (Figure 1-4). Thus “active discovery” is the iterative and automated process of discovering new data, which is well-suited to the problem of computational catalyst discovery.

1.4 Research objective

The goal of my thesis research is to help stabilize the global environment and economies. I do this by mitigating climate change through advancement of sustainable energy storage technologies. Specifically, I create and implement methods for discovering catalysts that may enable commercial production of solar fuels and chemicals.

Chapter 2

Dynamic Workflows for Routine Materials Discovery in Surface Science

This work originally appeared as: Kevin Tran, Palizhati Aini, Seoin Back, and Zachary W Ulissi. Dynamic Work-flows for Routine Materials Discovery in Surface Science. Journal of Chemical Information and Modeling, 58(12):2392—2400, 2018.

2.1 Abstract

The rising application of informatics and data science tools for studying inorganic crystals and small molecules has revolutionized approaches to materials discovery and driven the development of accurate machine learning structure/property relationships. We discuss how informatics tools can accelerate research, and we present various combinations of workflows, databases, and surrogate models in the literature. This paradigm has been slower to infiltrate the catalysis community due to larger configuration spaces, difficulty in describing necessary calculations, and thermodynamic/kinetic quantities that require many interdependent calculations. We present our own informatics tool that uses dynamic dependency graphs to share, organize, and schedule calculations to enable new, flexible research workflows in surface science.

This approach is illustrated for the large-scale screening of intermetallic surfaces for electrochemical catalyst activity. Similar approaches will be important to bring the benefits of informatics and data science to surface science research. Lastly, we provide our perspective on when to use these tools and considerations when creating them.

2.2 Introduction

Humans use a substantial amount of fossil fuels to meet increasing energy demands,[4, 3] but this fossil fuel consumption may be causing green house gas emissions and pollution that are changing global environments and climates[114, 43, 2] at rates that may be causing an unsafe drop in global biodiversity and ecosystem stability.[148, 151] To combat this, various alternative energy economies have been proposed.[160, 28] These alternative economies include the use of photovoltaics, batteries, solar fuels and chemicals, fuel cells, and other rising technologies.

A common theme among these technologies is that researchers still need to perform additional materials development to improve economic viability.[160, 146, 24] Some researchers have turned to *ab initio* simulations to aid in materials screening and development. For example: Density Functional Theory (DFT) is used to predict electrocatalyst performance for hydrogen evolution, oxygen reduction/evolution, and nitrogen reduction;[160] DFT is also used to predict electrolyte stability for batteries;[20] and both DFT and molecular dynamics (MD) are used to predict material band gaps for photovoltaics[77, 132]; among many others. These simulations can be computationally expensive and time consuming to set up, especially when dealing with large search spaces. This is why many researchers have begun building or using information science tools such as workflow management, database creation/management, or surrogate modeling to accelerate their materials development. These types of informatics tools are plentiful in software engineering fields and are proliferating in materials informatics fields, but they are less common in the field of surface science and catalysis.

In this perspective, we first discuss how informatics tools can augment and improve

computational research for materials screening and development. We then describe our own informatics tool, GASpy, that we built on top of previous tools to fill a new niche of data-driven catalyst science: fully automated materials screening. We show how GASpy can accelerate two common surface science calculation types: low coverage adsorbate thermodynamics and surface energy calculations. We also show the recent use of GASpy for the first large-scale screen of intermetallic surfaces for electrochemical catalyst activity. Finally, we discuss inherent challenges and trade-offs in developing informatics-based approaches and provide our perspective on their strengths, their weaknesses, and other considerations for future research efforts.

2.3 Review of informatics tools for inorganics and surfaces

2.3.1 Databases (DBs)

Computational collaborators often share data in order to avoid performing redundant calculations, especially when we are in the same research group. We have historically shared data by simply storing our results in file systems on shared computer resources and then reporting the location of the files for our collaborators to search and parse manually. This method is easy to set up and is flexible, but it becomes time-consuming and intractable as the amount of data that we share increases. This method also requires that our collaborators have the same access to computing resources that we do, which is a barrier for other research teams and especially experimentalists. As a community, some of us have addressed this issue by making new article formats that contain the underlying data[84] or by making databases, which allow us to store and share data systematically. Using databases also allows us to reduce the amount of redundant calculations that we perform as a community.

There are various examples of databases being used in the surface science and materials community. One widely used databasing tool for small datasets in surface science is the database module in the open-source Atomic Simulation Environment

(ASE) which stores simulations as ASE atoms objects in a custom SQL schema.[64] Many examples of such databases for small calculations have been collected at the Computational Materials Repository.[95] ASE atoms objects can also be stored in flexible Mongo schema using the open-source Vasp module,[83] or via the more specialized open-source pymatgen-db Python module.[136] Other examples of large-scale databases include (but are not limited to) OQMD,[82] The Materials Project,[72] Aflowlib,[34] and NOMAD,[1] and we recommend previous reviews for extensive discussion of their datasets and methods.[63]

The increased use of databases in materials and surface science has drastically accelerated material discovery by reducing the barrier for collaborative data sharing. The availability of large, trustworthy databases has also been a primary driver of surrogate models discussed below. However, these databases are complex systems and require training for end users to properly populate, query, and manage. For the most part there are not consistent API's or schemes to access information or calculations across databases, in large part due to the overwhelming number of possible calculation types, levels of theory, and materials classes. Commercial efforts such as Citrination [63] have attempted to address this API challenge. Jain et. al. also offer a more detailed review on the impact of databases on the computational research community.[73]

2.3.2 Workflow Managers (WFMs)

Computational research often involves iterative studies using well-established or derivative methods, such as calculating bulk formation energies across thousands of materials or calculating potential energy surfaces across dozens of surfaces. Each of these calculations takes time for a researcher to configure and manage, and the time required to repeat the configurations dozens, hundreds, or thousands of times can sometimes become the limiting factor in these studies. Workflow managers (WFMs) coordinate and automate these common tasks to allow researchers to spend more time on research and less time performing repetitive tasks.

One of the most established set of WFMs was developed by the same team

that manages the Materials Project. Their WFMs include Python Materials Genomics (pymatgen),[136] FireWorks,[71] custodian,[136] and Atomate.[116] Pymatgen is open-source Python library for materials analysis; it has analysis tools available such as phase diagram generators or adsorption site finders.[136] FireWorks is software that manages calculations across multiple computing clusters, which is known in the computer science community as pilot abstraction. It provides extensive support for data-driven science computing such as failure detection and duplication elimination, long-running project reporting, and dynamic workflows. Custodian manages calculations for bulk crystal structures in select *ab initio* codes and can correct some setting errors.[136] Atomate, which is built on top of pymatgen, FireWorks, and custodian, provides a number of pre-built and customizable workflows for materials science applications. Once these tools are installed and learned, performing a larger number of calculations becomes much easier and performing them consistently becomes more natural. Other WFMs besides the ones developed by the Materials Project team exist and are necessary for populating many of the other large database systems. The main drawback to these WFM tools is that they take time to create and implement, and so researchers who are developing new calculation methods or dataset classes may not be able to wait for the construction of the workflow managers for their use-case.

2.3.3 Surrogate Models (SMs)

Even with the development of databases and workflow managers, some processing and search spaces are still too wide to be able to analyze fully by brute force. This has spurred the development of numerous surrogate models (SMs). Surrogate models are substitute models that produce the same predictions as their incumbent models, but they do so in a fraction of the time and less accurately. For example: Adsorption energies of intermediates can be correlated using d-band models[58] or scaling relations[8] instead of DFT. Adsorption properties of surfaces can be predicted using geometric descriptors, such as the generalized coordination number.[23] Both potential energy surfaces[13, 60] and reaction mechanisms[172] can be constructed/identified with artificial neural networks (ANNs). Compounds with low thermal conductivity can be

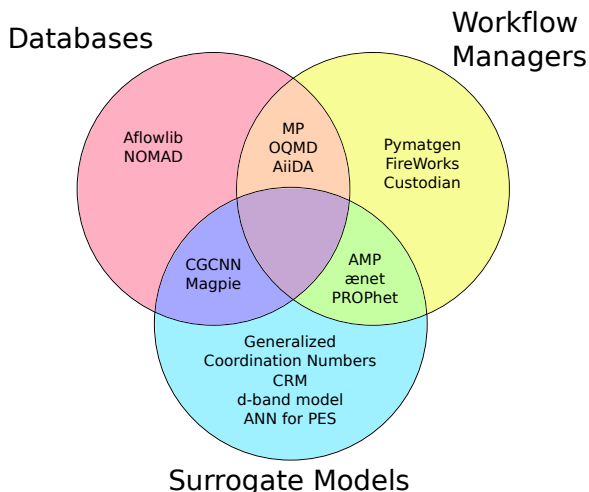


Figure 2-1: Literature approaches of various combinations of databases, workflow managers, and surrogate models.

discovered with Bayesian optimization.[161]

Surrogate models can also be used to discover insights and to supplement experimental designs. For example: Meredig and Wolverton introduced the Cluster-Ranking-Modeling (CRM) method to discover chemical descriptors that could be used to model materials’ behaviors; these descriptors were then used to develop underlying insights into the systems being studied.[123] There is an extensive number of examples of surrogate modeling showing success, and we refer the reader to textbooks with further details.[106] The examples we show here are only a small subset of the amount of published work regarding surrogate modeling.

The main downside of surrogate modeling is that it is inherently less accurate than classical *ab initio* models or experiments, and so surrogate model results should either be used with more caution or verified with classical models and experiments.

2.3.4 Combining DBs, WFMs, and SMs

DBs, WFMs, and SMs can also be combined to yield new combinations of benefits.

The first combination we discuss is the simultaneous use of WFMs with SMs. An example of this combination can be found in the Atomistic Machine-learning Package (AMP).[80] AMP is used mainly to create potential energy surfaces with

machine learning models instead of DFT, but it also has a feature that allows it to combine these machine-learning-predicted potential energy surfaces with DFT to perform nudged elastic band (NEB) calculations. AMP’s NEB calculations are performed at drastically accelerated speeds, but the minimum energy pathways that it yields are still supported by full-accuracy DFT calculations. Other examples of tools that combine ANNs with DFT workflows are the Atomic Energy Network (`ænet`)[11] and PROPhet.[85] Although these tools provide accelerated calculations with full-DFT-accuracy, they are not protected against data redundancy because they do not have internal databases to reference.

Another combination is to use DBs with SMs. The natural way to combine these tools is to use databases of calculations as training data for creating SMs. For example, Crystal Graph Convolutional Neural Networks (CGCNN) is one of many packages that uses calculated properties from the Materials Project database and the Inorganic Crystal Structure Database to provide material property predictions and atomic level chemical insights.[184] Materials-Agnostic Platform for Informatics and Exploration (Magpie) uses data from OQMD to create machine learning based models and predict properties of crystalline and amorphous materials.[176]

WFMs can also be combined with DBs. For example: the Materials Project uses its supercomputing resources and workflow tools such as `custodian` and `FireWorks` to perform data processing and analysis for millions of related calculations on bulk materials. AiiDA is another infrastructure that manages both data and calculations.[145] This combination of databasing and workflow management helps researchers create and share large amounts of data while avoiding duplication of calculations. However, frameworks like these require a relatively high amount of code and data maintenance.

2.4 Generalized Adsorption Simulations in Python: GASpy

2.4.1 Dynamic Workflows via Database Queries

The aforementioned databasing/workflow managing frameworks were built mainly to automate workflows and then applying them to many systems. This is functionally different from a standard, manual, expert-driven approach. If an expert wants to calculate a property of interest for a new system, they typically start by asking teammates or collaborators for prerequisite data and then checking system folders for the data. They then submit the smallest set of calculations to fill the gap in knowledge. This expert-driven approach is more flexible in that as long as pre-requisite data exists—regardless of source or precise method—then the data can be used. Of course, this method requires the expert to verify calculation context and accuracy, and there is an inherent trade-off in how verifiable the final calculation is.

For practical, day-to-day scientific explorations we needed a framework that was more closely aligned with this expert-driven approach, and so we constructed a Generalized Adsorption Simulator for Python (GASpy).[171] GASpy is an open-source repository that combines various workflow management packages with databasing software to create a “smart database” of DFT simulations and results. A user can query this database and it will return pertinent results that are already stored. If a user queries the database for results that are not yet stored though, then the “smart” aspect of the database automatically queues and executes all prerequisite calculations and stores the results for later reference.

The framework for the smart database is constructed around tasks, which are singular operations that can be combined dynamically into various pipelines to perform different calculations. We built our first set of tasks to calculate the adsorption energy between an adsorbate and a catalyst surface. These tasks are:

- Use DFT to calculate the equilibrium bond lengths within the bulk form of the catalyst so that the resulting electronic structure and energies can be stored for

later use. This process is known colloquially as a “relaxation”. GASpy currently uses bulk structures from The Materials Project.[72]

- Enumerate the various surface facets that may result from the bulk structure and then relax those surfaces. GASpy uses pymatgen[136] to enumerate the surfaces of a bulk.
- Enumerate the adsorption sites on the surface. GASpy uses Delauney triangulation as implemented in pymatgen.[128]
- Place the adsorbate on an adsorption site and then relax the adsorbed structure.
- Subtract the gas-phase adsorbate energy and the energy of the surface from the energy of adsorbed structure to calculate the adsorption energy.[134]

These tasks are then combined with various workflow managers and databasing software. Some tasks are dependent on others—e.g., a surface needs to be enumerated from a bulk before any adsorption sites can be identified on the surface. This task interdependency was managed by Luigi,[15] a software package that simplifies task execution by managing task prerequisites and interdependencies for the user. All Luigi tasks that required DFT relaxation are executed through FireWorks.[71] All of these tasks are summarized in Figure 2-2. New tasks are written by defining: (1) requirements for the task—usually database queries for results from a previous task, (2) methods to generate the result from the requirements and specifications, and (3) a description of where the output will be located so upstream tasks can verify the task completed and use the results.

All DFT results are stored in Mongo database collections. Mongo is used because of its ability to change data storage structure easily, which provides flexibility that is beneficial to a constantly-changing research environment. Note that one Mongo database collection is dedicated for intermediate DFT results, such as bulk energies. Another Mongo collection is dedicated for processed results (combining results from multiple calculations), such as adsorption energies. This separation of intermediate

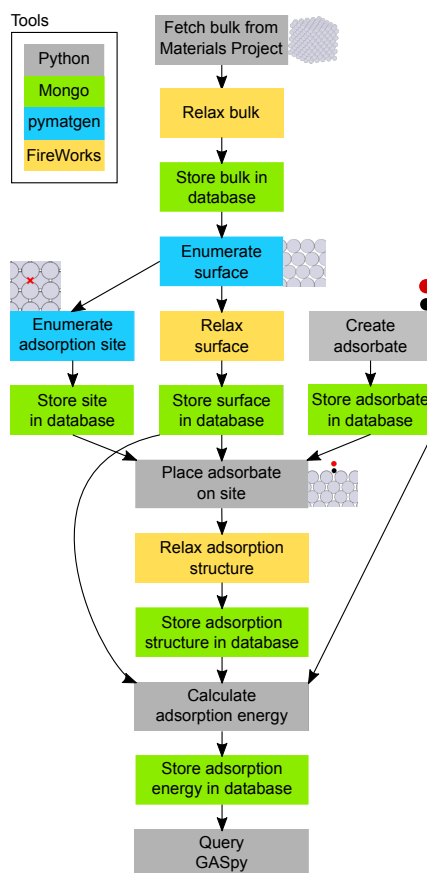


Figure 2-2: GASpy workflow in the form of a directed acyclic graph. Each box is a task, and all tasks are managed by Luigi. The colors of the tasks indicate which package that Luigi uses to execute the task. Triggering any single task will cause Luigi to trigger all necessary upstream tasks. Thus querying the database for the adsorption energy of a specific site will automatically trigger any previously incomplete tasks.

data from processed data provides the flexibility to create a new collection of processed data without needing to unnecessarily parse through extraneous data. For example: We were able to easily create new tasks that use our intermediate DFT results from surface calculations to populate a new database of surface energies without ever needing to parse adsorption energy results. This separation of results from intermediate data allowed us to reuse old data for new a purpose while minimizing query slow-down from data processing and analysis.

GASpy’s workflow management and databasing framework provides various benefits for computational materials and catalysis informatics. GASpy inherits the main strength of workflow management frameworks: reduced overhead for performing calculations. Traditional workflows involve training researchers for weeks or months and then requiring them to spend hours or days managing hundreds or thousands of DFT relaxations. GASpy users are trained for only a few hours and are able to spend only minutes to query and queue an arbitrary number of DFT relaxations. GASpy also inherits the main strength of databases: calculations are pooled together and never repeated. This includes intermediate calculations as well, which means that a single bare-surface relaxation queued by one user will be used by all other users when calculating adsorption energies on that surface.

GASpy’s framework does have some drawbacks though. For small numbers of calculations, the time required to construct the infrastructure and then use it is longer than the time required to simply perform and analyze results manually. Thus, creating and using a GASpy-like framework is efficient when it is based on established calculation workflows, but it is not efficient when based on non-established calculation workflows. This drawback is inherent in the framework and unavoidable. There are other drawbacks to GASpy that could be fixed in the future but have not yet been addressed. For example: GASpy has no way of handling errors automatically. If a DFT relaxation fails because a server crashes, then a user must re-query GASpy to re-queue the relaxation. Or if a DFT relaxation fails because it was unable to converge, then a user must submit another query to specify different DFT settings and re-queue the relaxation. There is also no notification system to help users monitor

specific DFT calculations. If a user purposefully queues 20 simulations on some specific copper surfaces, then they only know that their copper surface calculations are done when they appear in the database. These issues that are unaddressed but fixable are indicative of one of the main weaknesses of tools such as GASpy: it takes time to automate everything.

2.4.2 Combining workflow managers, databases, and surrogate modeling

Despite these drawbacks, the data querying system of GASpy provides a substantial advantage: automated simulation frameworks can now be used and implemented with a few lines of Python code, such as active machine learning. Active machine learning, also known as optimal design of experiments, is an iterative method for creating surrogate models. First a surrogate model is trained on a small set of data; the resulting model is then used to select additional training data.[163, 42] Thus the model queries its own training data, retraines itself, and repeats this process iteratively until the model converges. Selection of training data can be performed using various methods. One example is uncertainty sampling whereby the model queries data at locations where its predictions are most uncertain. Another example is minimization of expected error whereby the model queries data at locations that are most likely to yield useful information that can be used to reduce model error. Most of these selection criteria are intended to help the surrogate model explore a search space and thus build a training set to yield a more robust surrogate model.

The querying selection criteria could also be changed to exploit the surrogate model’s incumbent knowledge. This is the goal of surrogate-based optimization,[59, 32] which is the optimization of a specific objective function based on a simpler, more inexpensive surrogate model instead of the incumbent, more costly model. For example: Instead of optimizing the amount of carbon captured by an adsorber by using a system of partial differential equations, a surrogate model can be used to perform the same optimization in a fraction of the time.[32]

2.4.3 Application of GASpy

In the context of surrogate-based optimization, the process of actively and automatically querying new training data is called sample-point refinement.[59] The difference between active learning and surrogate-based optimization with sample-point refinement is a subtle but important difference: active learning seeks to construct a robust surrogate model, but surrogate-based optimization with sample-point refinement seeks to optimize an objective function. This difference in goals leads to different methods for querying new data. An example querying method for optimization is the searching surrogate model (SSM) method where new data is selected based solely on the surrogate model’s prediction of the new data point’s proximity to the optimum.[59] The fundamental difference between the methods used in surrogate-model optimization and the methods used in active learning is that the optimization methods tend to exploit the surrogate model’s knowledge about a search space while active learning methods tend to explore the search space instead.

In the context of catalysis informatics, the only example of fully automated surrogate-based optimization with sample-point refinement that we could find is implemented in GASpy. GASpy seeks to find adsorption sites whose adsorption energies are optimal for particular reactions as determined by literature-sourced volcano scaling relationships.[133, 105] This search/optimization is done by training a surrogate model on the database of DFT-calculated adsorption energies and then using a combination the SSM sampling method and random/uncertainty sampling to select new data points. These new data points are simply queried to GASpy’s smart database framework to queue the DFT calculations. This workflow is illustrated in Figure 2-3, which we currently use to perform 200–500 DFT calculations per day across three different computing clusters. These calculations are fully automated and require no user input at all.

Note that we are able to pause these machine-led DFT calculations and simply use GASpy as a smart database. Thus we can still use our database of results to search for both chemical insights and candidate materials, and then we can perform

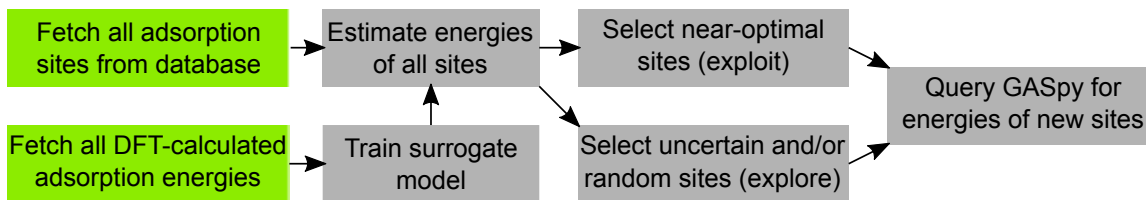


Figure 2-3: Active learning/sample-point refinement process for GASpy. Green tasks are performed with Mongo and gray tasks are performed with Python. The final task here, “Query GASpy”, is the same final task shown in Figure 2-2. This active learning/sample-point refinement process is repeated daily.

human-led follow-up calculations in a high-throughput fashion. This is the main strength of GASpy: we are able to accelerate human-led studies while simultaneously supplementing these studies with machine-led calculations that ensure that we use 100% of our computing capacity at all times.

We used GASpy in such a fashion to screen intermetallic alloys of 31 different elements for near-optimal CO adsorption energies, which can be indicative of CO₂ reduction activity.[105] GASpy performed approximately 20,000 DFT calculations of CO adsorption energies over time and found 131 different intermetallic surfaces with low-coverage CO adsorption energies near the optimal value for CO₂ reduction, as shown in Figure 2-4a. We used the same process to also search for intermetallics with near-optimal hydrogen bonding energies[133] to search for hydrogen-generating electrocatalysts; GASpy performed another 20,000 H adsorption calculations to find 304 candidate surfaces.[171]

In addition to identifying candidate surfaces, the DFT results were also used to identify candidate compositions via t-SNE [113] analysis (Figure 2-4b). This allows us to gain insights into the potential intermetallic performance based solely on the number and distribution of potential active sites instead of activity of one surface at a particular alloy ratio. In this plot, the clusters indicate similarity in site coordination and elemental combinations. Colors indicate the binding strength between the adsorbate (CO) and active sites. Black points represent sites that bind too strongly; pink points indicate sites that bind too weakly; and purple points represent sites with optimal binding. The clusters that are nearly uniformly purple are further labeled in

Figure 2-4b; they represent candidate materials for experimental investigation.

We have shown that we can fully automate adsorption energy calculations and have already begun automating surface energy calculations. It follows that this type of automation scheme could be extended to other applications, such as transition state calculations or nanoparticle calculations. Automating such calculations requires a deep understanding of both the application and software engineering though. This means that it will be more difficult to create informatics tools for applications and calculation types that are less understood, such as oxide adsorptions or single-atom-catalysts (Figure 2-5). It will also be more difficult to create informatics tools without proper software engineering training, which is another field of expertise itself.

2.5 Perspective on informatics tools

2.5.1 Considerations when using informatics tools

Workflow managers, databases, and surrogate models are all useful informatics tools that have the potential to accelerate scientific discovery in catalysis. These tools can be integrated in various combinations to yield new frameworks with new strengths and weaknesses, summarized in Table 2.1 and Figure 2-6. In general: Both workflow managers and database tools accelerate and improve informatics-based research at the cost of increased programming overhead, a longer time for implementation, and reduced reliability (compared to expert computational chemists studying new systems). Surrogate models can be used to screen large search spaces when accuracy is not crucial, or they can be used to gain general insights into important features/factors. One of Table 2.1’s most interesting insights is that combining workflow managers with surrogate modeling provides the acceleration benefits surrogate models without the drawback of reduced accuracy, because the workflow manager is able to automatically verify the surrogate model accuracy at critical points. To use GASpy as an example: An adsorption site is not considered to be optimal just because the surrogate model predicts it to be. It is only considered optimal when the calculation is

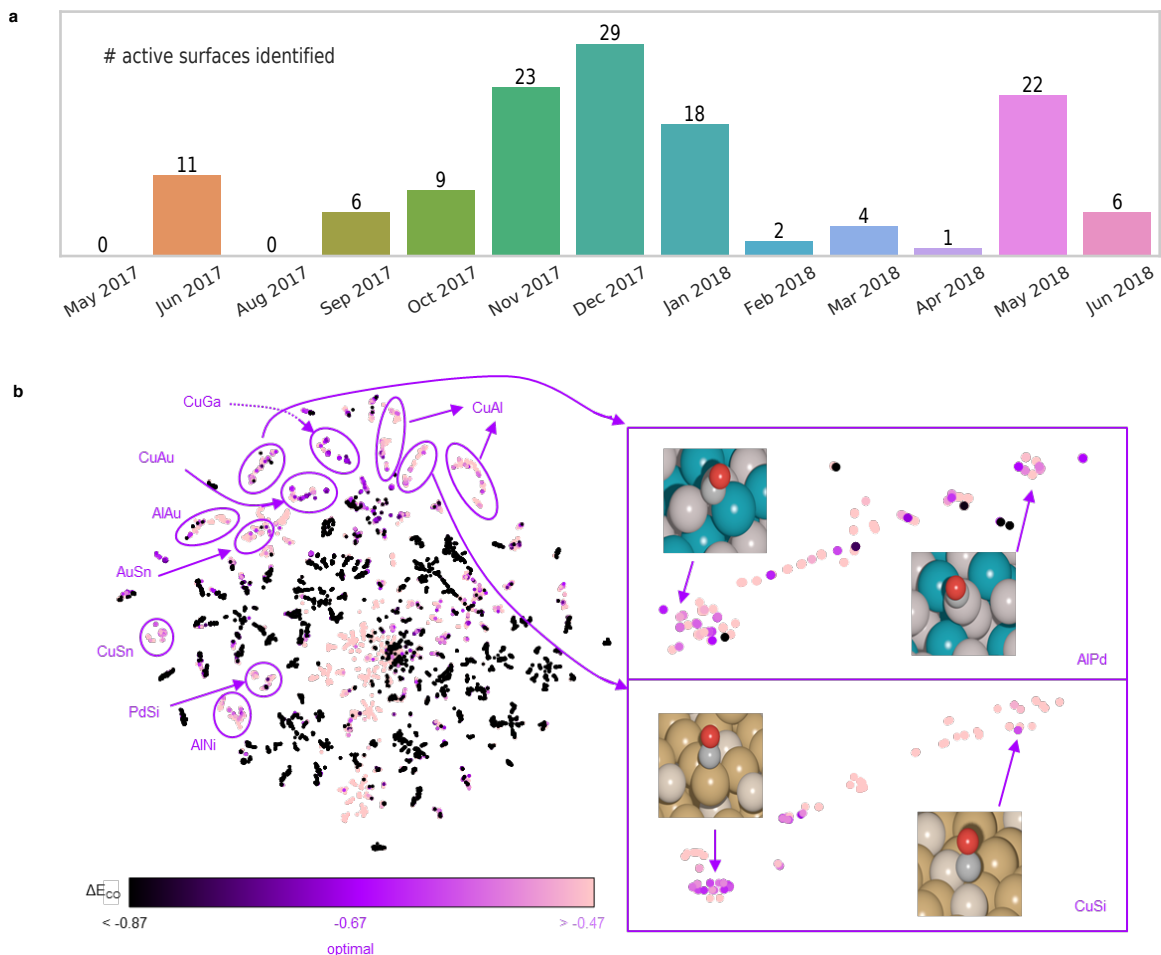


Figure 2-4: Summary of CO₂ reduction results from GASpy. **a**, The number of intermetallic surfaces with near optimal ΔE_{CO} values for CO₂RR identified as a function of time. **b**, Latent space visualization using t-SNE [113] of all the adsorption sites simulated with DFT. Proximity in this reduced space indicates similarity in the structures of the adsorption sites. Both figures were reproduced with permission from Tran & Ulissi.[171]

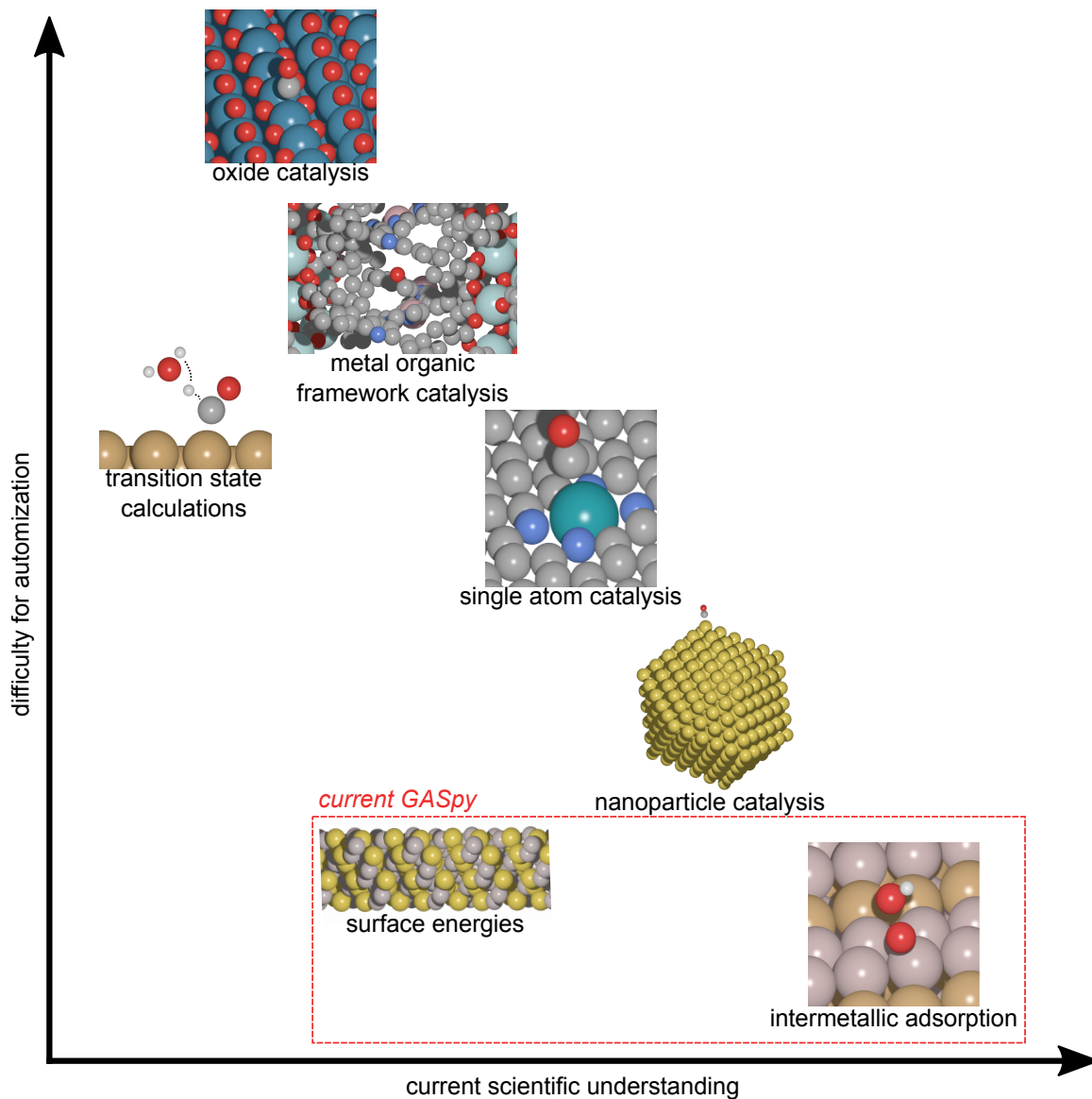


Figure 2-5: Estimated difficulty for automating a type of task vs. the field's current scientific understanding. In general, it is more difficult to create informatics tools for applications where we do not yet have systematic and generic methods developed.

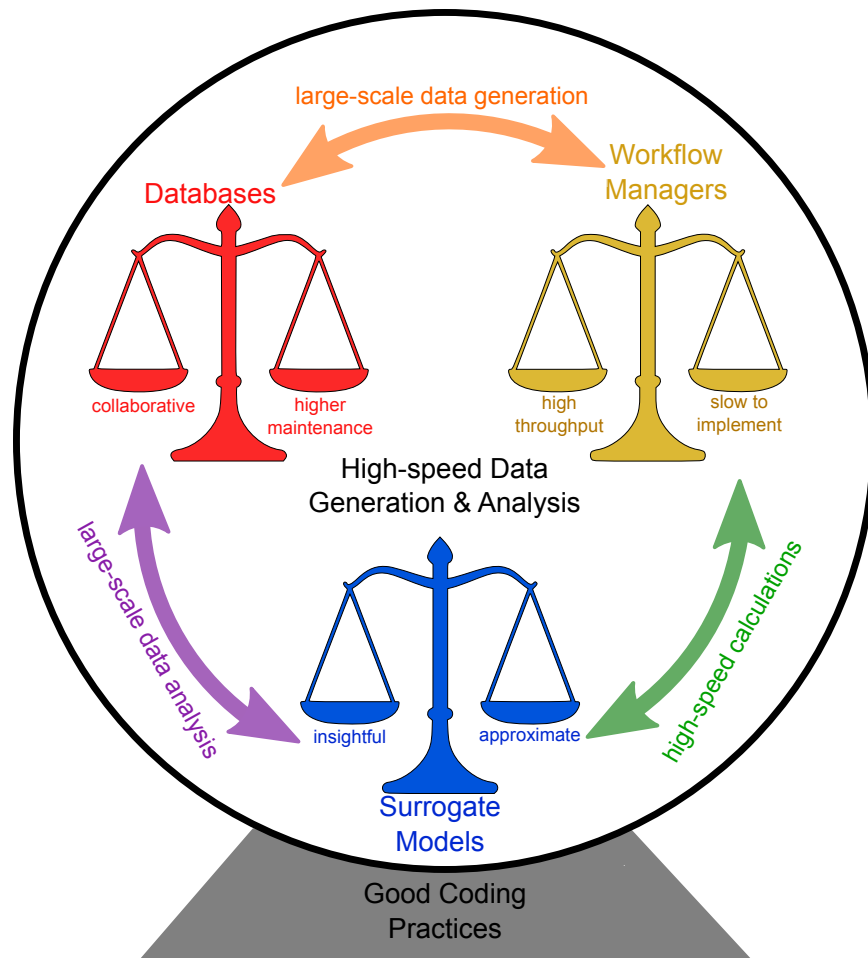


Figure 2-6: Various informatics and data science tools and how they interact with each other. All of these tools are supported by a foundation of good coding practices.

verified by DFT.

The main drawback to many of these tools is the increase in programming overhead and longer times to implementation. We do not recommend using workflow managers if you are under severe time constraints, do not have systematically enumerated design spaces, and/or are still experimenting with the feasibility of the workflows you are trying to automate. We also note that the management of many of these informatics tools requires much more strict programming skills and practices.

2.5.2 Considerations when building informatics tools

We encourage any research teams who intend to build their own informatics tools to educate themselves on good code documentation, version control, unit testing, efficient coding, and other “good coding practices”.[115, 181, 159] Good code documentation—e.g., clear and concise naming conventions, document strings, comments, and even formatting—are crucial for code readability. Code readability is essential because new team members need to be able to understand and modify old code quickly and easily. This is especially important for research groups that employ students, since new students are often working with code written by others. Version control is crucial for documentation of changes and for managing/minimizing the risks of introducing new software features. Unit testing is crucial for ensuring that code works and that new changes do not interfere with other functionalities of the software. The importance of unit testing scales with the complexity of the software being built, because there are more opportunities for unintended interactions to cause errors. All of these skills are usually taught to computer scientists and not usually taught to classically trained research scientists. Therefore research groups must ensure that their software managers learn these programming skills from the numerous available resources, such as summary articles,[181] textbooks[115, 159] that were written specifically for computational scientists, online video tutorials, workshops, or colleagues. Failure to learn these skills could result in code bases that are not easily transferable, reproducible, or sustainable.

We also encourage all researchers who intend to create informatics tools to open their source code publicly. This helps ensure reproducibility while also allowing others to accelerate their research. We could not have made GASpy if the dozens of researchers and industrial software engineers did not provide their source code publicly. And we publish our source code with the hope that others may be able to accelerate their work with our code and ideas.

Lastly: We encourage those who are thinking about making their own informatics tools to first consider whether or not there is already a tool available for you to use.

If there is an appropriate tool already available, then you will be able to save yourself the time and effort required in creating something new. If you find a tool that will “almost work” for your needs, then we encourage you to collaborate with the tool’s authors to improve that tool. This type of open-source collaboration benefits everyone and also helps reduce the proliferation of redundant code bases.

2.6 Conclusions

We have both seen and created cases in the materials and catalysis fields where informatics tools have been used successfully to conduct research. Many of these tools have different strengths and weaknesses; they have different use-cases; and they require different skills to use effectively. We have both seen and created cases in the materials and catalysis fields where informatics tools have been used successfully to conduct research. Many of these tools have different strengths and weaknesses; they have different use-cases; and they require different skills to use effectively. But when they are used correctly and in the right situations, they complement and accelerate incumbent research workflows substantially and are well worth the effort. We encourage more researchers in the computational materials and catalysis fields to consider how they can use informatics tools, and we also encourage them to use these tools carefully.

Table 2.1: Various strengths and weaknesses of various informatics tools alongside the corresponding use-cases that we recommend for each.

Informatics tools	Strengths	Weaknesses	Recommended use-case
WFM	high-throughput consistent methods	longer time to implement	repetitive calculations derivative methods
DB	reduced redundancy	increased overhead	collaborative efforts
SM	accelerated calculations	reduced accuracy	large search/processing spaces when searching for general insights
WFM+DB	high-throughput consistent methods reduced redundancy	longer time to implement increased overhead	repetitive calculations derivative methods collaborative efforts
DB+SM	reduced redundancy accelerated calculations more robust training sets	increased overhead reduced accuracy	collaborative efforts large search/processing spaces when searching for general insights
WFM+SM	high-throughput consistent methods accelerated calculations no reduction in calculation accuracy	longer time to implement	expensive/repetitive calculations
WFM+DB+SM	high-throughput consistent methods reduced redundancy accelerated calculations enables active workflows	largest overhead longest time to implement	large screenings long-term, collaborative projects

Chapter 3

Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution

This work originally appeared as: Kevin Tran and Zachary W. Ulissi. Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution. Nature Catalysis, 1:696–703, 2018. It has been edited to include the supplementary information in Appendix A

3.1 Abstract

Electrochemical reduction of CO₂ or H₂ evolution from water can be used to store renewable energy that is produced intermittently. Scale up of these reactions requires the discovery of effective electrocatalysts, but the electrocatalyst search space is too large to explore exhaustively. Here we present a theoretical, fully automated screening method that uses a combination of machine learning and optimization to guide density functional theory calculations, which are then used to predict electrocatalyst performance. We demonstrate feasibility of this method by screening various alloys of 31 different elements, thereby performing a screening that encompasses 50% of the d-block elements and 33% of the p-block elements. This method has thus far

identified 130 candidate surfaces across 54 alloys for CO₂ reduction and 258 surfaces across 102 alloys for H₂ evolution. We use qualitative analyses to prioritize the top candidates for experimental validation.

3.2 Introduction

Global energy demands have increased over time and are likely to continue increasing.[4, 3] Meeting these demands using only fossil fuels may not be possible because doing so may negatively impacts the world’s environments, climate, and biodiversity.[114, 43, 148, 2] Alternative methods for energy production and storage include solar fuels, which are synthetic fuels created with photovoltaic energy or with photoelectrochemical cells.[98, 160] Examples include H₂ created by the electrochemical reduction of water, which can be combusted or used in hydrogen fuel cells, and synthetic hydrocarbons created by reducing CO₂. Solar fuel production is currently inhibited by a lack of sufficiently active, efficient, selective, stable, and low-cost catalysts.[129]

High-performing catalysts can be discovered using *ab initio* methods such as Density Functional Theory (DFT) to predict catalyst properties. A common approach is to use DFT to predict thermodynamic energy descriptors that correlate with detailed microkinetic model results or experimental measurements of catalyst activity and selectivity. For the CO₂ reduction reaction (CO₂RR), the CO adsorption energy is a common descriptor for predicting activity towards hydrocarbon production.[105] This method of using adsorption energies to predict performance is general and has been applied to many reaction systems,[134] including the Hydrogen Evolution Reaction (HER).[53] Of course, single descriptors cannot completely describe the electrocatalytic performance of an intermetallic, which requires analyses of surface stability, high coverage thermodynamics, and electrochemical kinetics. These analyses take considerable resources, and full theoretical studies cannot keep pace with accelerating experimental studies of intermetallics. Thus, a screening method for prioritizing experimental and theoretical studies is valuable.

Screening large search spaces is nontrivial. Some methods address this by fo-

cusing on limited search spaces, such as bimetallic A_3B crystals[61] or simple cubic lattice intermetallics.[101] Other approaches include creating databases of electronic structure calculations.[68, 154] Such wide-scoped datasets are necessary to perform robust screening studies, but creating the results in these databases has required collaboration between dozens of DFT experts. A consequence of this approach is that researchers spend a substantial portion of their time configuring, managing, and waiting for DFT calculations, which are inherently time consuming. The personnel cost of configuring and managing DFT calculations has been addressed by creating computer-science-based solutions. Generalized frameworks exist that are able to enumerate surfaces and adsorption sites on arbitrary intermetallic bulk structures.[136, 128] Software exists that manages computationally intensive calculations across multiple computing clusters.[71] These solutions are vital for performing high-throughput material screenings.

The computational cost of theoretical materials screenings has been partially addressed by the development of machine learning (ML) methods to accelerate DFT calculations. ML regressions on DFT data can yield formation energy predictions at a fraction of the computational cost[121, 177] or accelerate screenings within individual bimetallics.[173] Neural networks can accelerate nudged elastic band studies of reaction kinetics[142] or enable the study of disorder on intermetallic configurations.[19] One of the shortfalls of many of these approaches is that they require experts to manually prioritize new materials to screen and study. Many also rely on DFT-computed surface features such as the d-band center, or predict activity only at the surface level without distinguishing between surface sites.[101]

There are some methods that have been used to accelerate computational screenings that may not have been applied to electrocatalyst discovery yet. Surrogate-based optimization[59] is a method where a surrogate model is built to replace a more computationally expensive model, and then an objective function is optimized on the cheaper surrogate model. Another accelerative method is active machine learning—also known as optimal design of experiments.[163] With this method, a surrogate model is created from a given dataset, and then the model is used to select which

data should be obtained next. The selected data is added to the original dataset and then used to create an updated surrogate model. The process is repeated iteratively such that the surrogate model is improved continuously. This method of iterative surrogate model screening has already been used in other fields such as discovery of light emitting diodes,[51] drug discovery,[178] or molecular property prediction.[56]

We created a workflow that borrows ideas from both surrogate-based optimization and active machine learning. We used this workflow to screen a search space of 1,499 intermetallics for potentially selective catalysts for CO₂RR and HER. The workflow identified 54 intermetallics that have surfaces with near-optimal descriptors for CO₂RR and 102 intermetallics for HER, including both already-discovered and undiscovered catalysts. We then use a qualitative, heuristic method for down-selecting these compounds further, resulting in a shortened list of 10 intermetallics for CO₂RR and 14 intermetallics for HER. Viewing the intermetallic design space holistically also allows trends and design rules to be identified that are difficult to see from small case studies. This workflow can be extended easily to other reaction chemistries for which ideal thermodynamic descriptors are known.

3.3 Results

3.3.1 Framework Construction

Our workflow uses machine learning models to search an arbitrarily large design space of intermetallic crystals and surfaces for near-optimal activity (Figure 3-1). Surfaces are searched for ideal CO and H adsorption energies, which are indicative of catalyst performance for CO₂ reduction[105] and H₂ evolution,[53] respectively. The workflow verifies the adsorption energies of these sites by performing DFT calculations automatically. DFT results are stored in a database, which is used to retrain the machine learning models. This yields a closed feedback loop of ML screening, DFT verification, and ML retraining that produces a database of DFT results that grows continuously, systematically, and without the need for user interaction. It is important

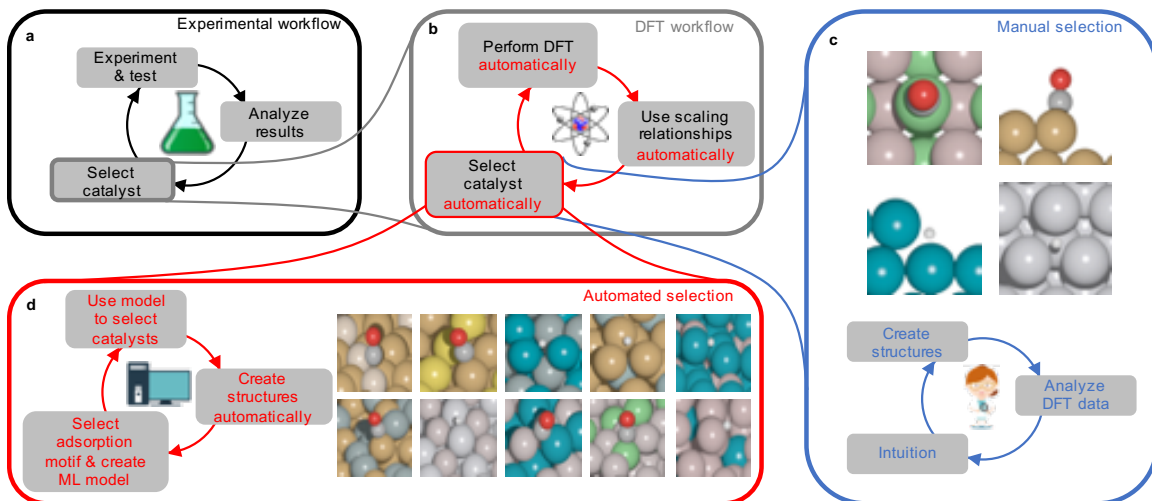


Figure 3-1: **Workflow for automating theoretical materials discovery.** **a**, The experimental workflow for finding catalysts is accelerated by **b**, a DFT workflow for screening the catalysts *ab initio*. **c**, Conventional workflows (blue) require scientific intuition to select candidates for DFT screenings, while **d**, our workflow (red) uses machine learning to select candidates systematically and automatically. Red text outlines the framework that this study used to perform active machine learning/surrogate-based optimization.

to note that this workflow does not use ML to accelerate calculations of user-supplied systems. Instead, it uses ML to guide full-accuracy DFT screenings. Thus we use DFT to perform a surrogate-based optimization, and we use an active learning feedback loop as an optimization guide.

3.3.2 Enumeration of search space

To generate a search space of adsorption sites, we obtained 1,499 different intermetallic combinations from the Materials Project[72] across 31 different elements (Supplementary Figure A-1). We enumerated the surfaces on each intermetallic using pymatgen[136] and then used Delaunay triangulation[128] to enumerate all adsorption sites on each surface; see the Methods section below for additional details. This search space encompassed 50% of the d-block elements and 33% of the p-block elements. Every possible adsorption site on each surface was considered for potential activity by our ML models.

3.3.3 Active learning optimization of catalyst descriptors

We enabled the machine learning of catalyst descriptors by developing a fingerprinting method to represent an intermetallic adsorption site numerically (Figure 3-2). For each site, the types of elements coordinated with the adsorbate were tabulated. Each element type was described with a vector of four numbers: the atomic number of the element (Z), the Pauling electronegativity of the element (χ), the number of atoms of the element coordinated with the adsorbate (CN) as determined by Voronoi tessellation performed by pymatgen,[136] and the median adsorption energy between the adsorbate and the pure element ($\Delta\tilde{E}$). $\Delta\tilde{E}$ values were calculated from our own database of adsorption energies, and χ values were obtained from the Mendeleev database[110]. We repeated this vector creation process on the second shell of atoms that are bonded to the coordination atoms. One issue with this method is that it yields a variable number of features. We addressed this issue using a method found in literature.[35] See the Methods section below for additional details. Note also that the illustration in Figure 3-2 this is a simplification. The real fingerprint vector has 4 items per element and 4 elements per shell, which yields a total of 32 items per fingerprint vector.

These fingerprints were chosen using a combination of intuition, trial-and-error, and success in other surrogate modeling studies. To account for bulk steric effects, atomic radii have been used as features.[35] Atomic radii may change depending on the local environment though, so elemental periods and groups may be appropriate substitutes for atomic radius. Initial, heuristic investigations showed negligible difference in performance between using period/group and atomic number though, so we used the atomic number because of its relatively small dimensionality. To account of electronic affinity effects, Pauling electronegativity has been shown to be a successful feature.[101] To account for both sterics and environmental electronic effects, the coordination number has been shown to be a successful feature.[22] To improve predictive capability, crude estimates of properties have been shown to be successful.[190] In this setting, the crude estimate of adsorption energy on a specific site is $\Delta\tilde{E}$.

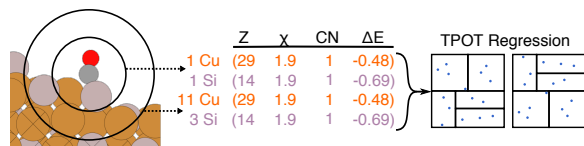


Figure 3-2: **Fingerprint of coordination site.** Adsorption sites are reduced to numerical representations, or fingerprints, and these fingerprints are used as model features by TPOT[135] to predict ΔE_{CO} . Z is the atomic number of an element; χ is the Pauling electronegativity; CN is the number of atoms of an element within a shell; and $(\Delta \tilde{E})$ is the median adsorption energy between the adsorbate and the pure element.

An automated machine learning package, TPOT,[135] was then used to select a machine learning regression method to predict adsorption energies from the site fingerprints. TPOT’s recommended modeling pipeline changed regularly due to the stochastic nature of TPOT and the constantly changing training dataset. To aid prediction we used a preprocessing pipeline to shift and scale each feature across all data points so that the averages and variances for each feature were zero and one, respectively. We also performed a principal component analysis on the fingerprints to orthogonalize the feature space. Supplementary Note 1 outlines the TPOT settings used, and the Supplementary Methods section outlines other regression techniques and feature representations we tested during development.

The regression methods illustrated in Figure 3-2 were combined with all available DFT data to train and update surrogate models daily. These models were used to aid in selecting adsorption sites for DFT calculation. First, the trained models were used to estimate ΔE_{CO} and ΔE_H for all the adsorption sites that we enumerated. These estimates were pooled with the explicit DFT results stored in the database, and ML-estimated adsorption energies were removed if the DFT result existed for that exact site/surface/adsorbate, leaving only one prediction/estimate per adsorption site. Then we defined the strongest binding energy on each surface as the low-coverage adsorption energy of that surface. The surfaces whose low-coverage adsorption energies were predicted to be near-optimal (as established in the Methods section) were selected for DFT calculation with a Gaussian probability defined by the distance of the predicted site from the optimal values with a standard deviation of 0.2 eV.

The targeting of specific adsorption energies with near-optimal values allows us to exploit the knowledge that the surrogate models have learned thus far, while the addition of Gaussian noise is a heuristic method that allows us to explore search spaces that the surrogate model normally would not suggest. Approximately 80% of the calculations were dedicated to this descriptor optimization goal. The remaining 20% of our resources were dedicated to simulating all of the sites on surfaces whose low-coverage adsorption energies were closest to the top of the volcano—i.e., nearest to a ΔE_{CO} of -0.67 eV and a ΔE_H of -0.27 eV for CO₂RR and HER, respectively. This mitigated the chances of finding a false minimum adsorption energy on a surface. The regression and surrogate model prediction was performed once per day, and the Gaussian selection of DFT calculations was performed four times per day. In total, 42,785 DFT calculations of adsorption energies were completed by this study at a rate of approximately 200–300 calculations per day. See the Methods section below for details regarding these DFT calculations.

3.3.4 Performance of active learning optimization

Due to the iterative nature of the surrogate modeling, we calculated prediction errors via evaluation on a rolling forecasting origin.[69] Specifically: We retrospectively trained a surrogate model on the first 200 data points that we obtained and then calculated the prediction errors between the next 200 data points and this first model’s predictions of these points. We then trained a second model on the first 400 points and then calculated the prediction errors between the next 200 data points and this second model’s predictions of these points. We performed this iteratively until we obtained a single prediction error for every data point, excluding the first 200 points. All prediction errors are plotted against time in Figure 3-3a and Supplementary Figure A-2 for ΔE_{CO} and ΔE_H , respectively, along with a record of the number of near-optimal surfaces identified over time. The root-mean-squared-error (RMSE), mean absolute error (MAE), and median absolute deviation (MAD) across all of the time-dependent ΔE_{CO} predictions are 0.46, 0.29, and 0.17 eV, respectively. The RMSE, MAE, and MAD of ΔE_H predictions are 0.41, 0.24, and 0.16 eV, respectively. Note that we

chose 200 as the step size because our framework was able to perform at least 200 calculations per day, and so a step size of 200 served as a proxy for surrogate model updates.

The profile of the prediction errors over time provide us with practical insights into our framework. Between November 2017 and January 2018, fellow users of our computing clusters reduced their usage during the holiday season. This allowed our automated framework to effectively consume their unused capacity, thereby increasing our calculation throughput. This temporary increase in throughput is the likely cause of the relatively high rate of surface identification, and the improvement in prediction errors at this time may have been caused by an improved sampling of the search space. Then in February 2018, we expanded the number of elements in our search space from ca. 20 elements to the 31 elements that we are searching currently. We seeded this new search space by manually queueing calculations with the new elements. This expansion in search space and subsequent seeding may have caused the decline in both predictive performance and identification rate. In May 2018, we refined our zero-point energy, entropic, and solvation correction calculations to what is now shown in the Supplementary Methods section. This refinement may have caused the increase in the number of surfaces identified during that month. Thus the trends in the prediction errors and identification rates are confounded with both the methods we used and various managerial events, such as changes in throughput capacity, changes in search space, or changes in optimization targets.

Performance metrics and plots that are typically used to judge static surrogate models are shown in the Supplementary Note 2, such as train/test errors, parity plots, and learning curves. Notably, the errors calculated via rolling forecasting origin are generally larger than the errors calculated from the classical train/test split and learning curve methods. This is because classical train/test splitting methods allow the training sets to share the same sampling space as the test sets. Our workflow often searches unexplored sampling spaces though, which are more difficult to predict. Thus the rolling forecasting origin method of evaluating error is more representative of our use scenario because it restricts the models' from seeing data that they would not

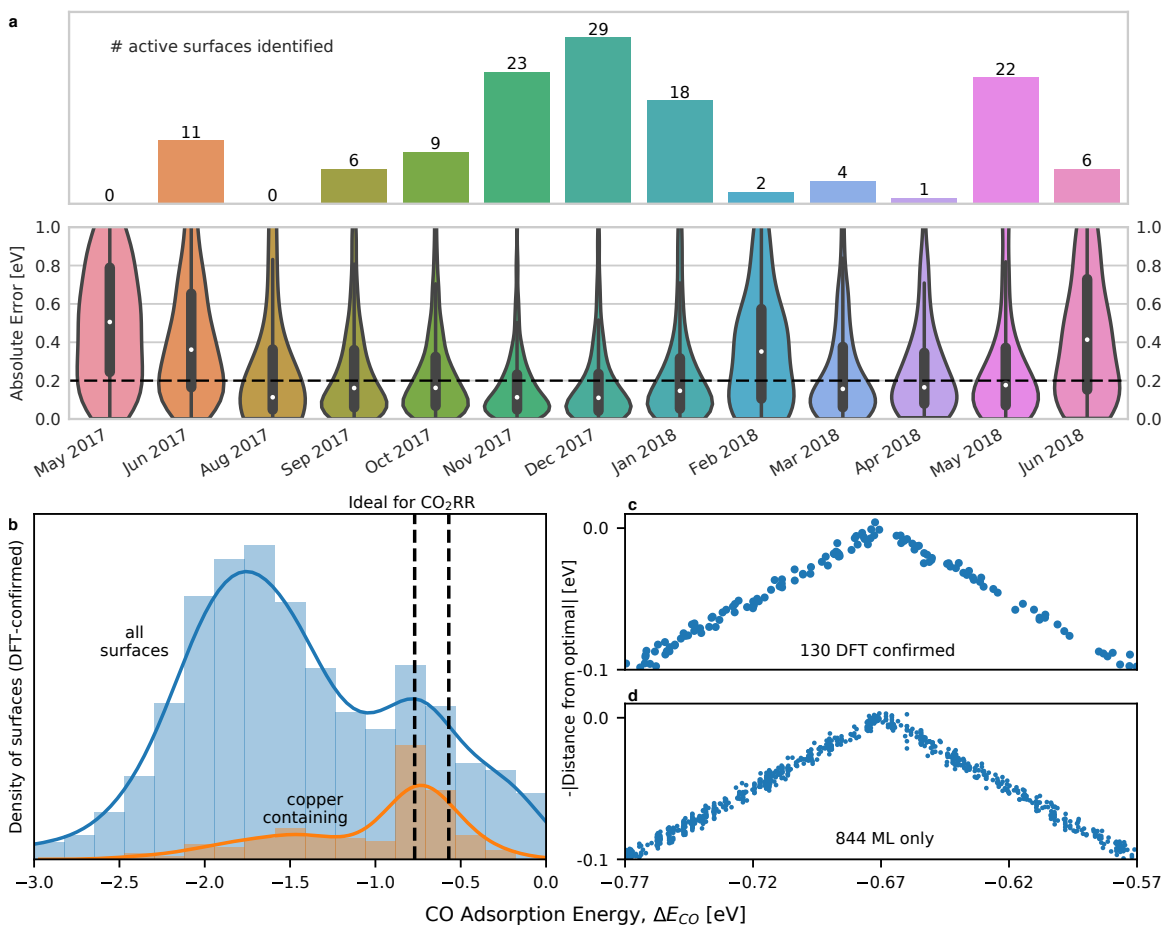


Figure 3-3: **Identification of surfaces with near-optimal ΔE_{CO} values for CO₂RR.** **a**, Both the number of near-optimal surfaces identified and violin plots of the absolute error in predicting ΔE_{CO} as a function of time. The outer shells of the violins bound all data; narrow vertical lines bound 95% of the data; thick vertical lines bound 50% of the data; and white dots represent medians. Apparently missing months are not shown because no data were collected during those months. **b**, The normalized distribution of the low-coverage, DFT-calculated CO adsorption energies of all of the DFT-analyzed surfaces in this study. The sub-distribution for copper is also illustrated in orange. Dashed lines indicate the ± 0.1 eV range around the optimal ΔE_{CO} value of -0.67 eV. **c**, Surfaces whose low-coverage CO adsorption energies have been calculated and verified with DFT. **d**, Surfaces whose low-coverage CO adsorption energies have been calculated only by the machine learning models.

normally see in practice. We hypothesize that the rolling forecasting origin errors and the train/test errors would converge if we had sampled the search space sufficiently.

3.3.5 Discovering potential intermetallic catalysts for CO₂ reduction

This framework discovered 130 different intermetallic facets with near-optimal ΔE_{CO} as confirmed by DFT (Figure 3-3, Supplementary Table A.1). These surfaces correspond to 54 different intermetallic combinations and are recommended for experimental verification of activity. Some of these intermetallics have already been investigated. For example, Cu/Sn blends have been shown to reduce CO₂ to either CO or formate at high Faradaic efficiencies[130]. Ni/Ga intermetallics have been shown to be active for CO₂ reduction,[167, 173]. Pd/Au bimetallics have been shown to be active for CO₂ reduction to C1-C5 products[86], and ML results in this study suggest that single Pd atoms surrounded by Au atoms may be the most likely bimetallic active site with ΔE_{CO} ca. -0.8 eV, in contrast to the hypothesized Pd-rich Pd/Au site. Cu/Al bimetallics, which have not been previously studied, also show promising experimental results in current ongoing work whose results will be published in due course.

3.3.6 Discovering trends in CO₂ reduction

In addition to simply finding potential active surfaces, the data that this study generated can be used to gain insight into the chemistry of CO adsorption. Figure 3-4 illustrates the fraction of enumerated surfaces that have near-optimal values for ΔE_{CO} for various bimetallic combinations. Elements in Figure 3-4 are rank ordered by the average pure-surface ΔE_{CO} values as calculated by this study with DFT. Forming bimetallics with two elements that both have stronger binding than Ge generally led to inactive materials, but silicon broke this trend, suggesting that it has destabilizing properties when alloyed into an intermetallic.

Other trends can also be found when analyzing different elemental pairings. A

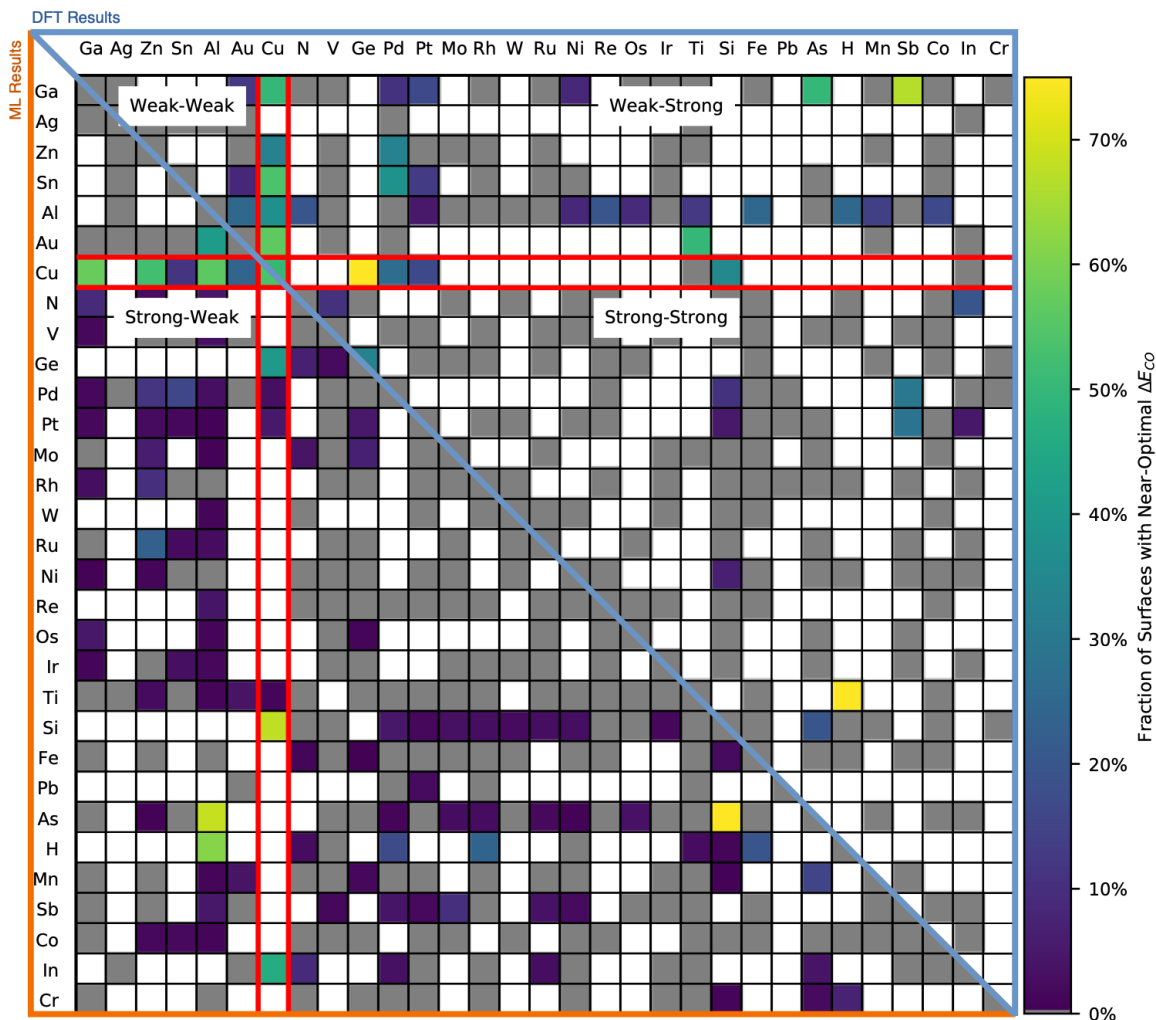


Figure 3-4: **CO₂ reduction activity map for bimetallics.** Visualization of two component intermetallics whose surfaces have low-coverage CO adsorption energy (ΔE_{CO}) values inside the range of $[-0.77, -0.57]$ eV. White shading indicates an absence of any enumerated surfaces; grey shading indicates that all ΔE_{CO} values are outside the range of $[-0.77, -0.57]$ eV; and colored shading indicates possible activity. The ΔE_{CO} values used to create the upper half of this figure were calculated by DFT, and the values used to create the bottom half were calculated by the surrogate ML model.

number of strong-weak elemental pairings yield possibly active surfaces, including combinations of strong-binding elements like Pd, Pt, Ni, or Os with weak-binding elements like Al, Sn, Ga, or Sb, showing that the strong/weak Ni/Ga motif found in previous work is more general than previously known.[173] Interestingly, combining two weak-binding elements can lead to possibly active surfaces. For example: The strongest binding Ga surfaces and Au surfaces are approximately -0.44 eV and -0.53 eV respectively, based on a combination of DFT and ML predictions. However, a Ga-Ga bridge site on AuGa₂(100) leads to a near-ideal binding energy of -0.57 eV.

Although the volume of data generated by this study is arguably intractable to study in detail, the size of the data enables certain methods of data analysis. For example: We are now able to assess potential intermetallic performance based solely on the number and distribution of potential active sites instead of the activity of one particular surface at one particular alloying ratio. Figure 3-5 illustrates this point by showing all 19,644 sites where we performed DFT calculations of ΔE_{CO} . The X and Y axes in this figure are a reduced 2-dimensional feature space[113]; reference the Supplementary Methods for details. Clusters of points in this reduced space share similarities in site coordination and elemental combinations. Sites are colored by their ΔE_{CO} so that regions of strong-binding (black), weak-binding (light purple) or near-ideal binding (dark purple) can be identified.

Clusters that are nearly uniformly dark purple are robust combinations are labeled in Figure 3-5. These are not the only possible active alloys; instead they represent combinations that are most likely to yield a higher fraction of active adsorption sites than other alloy combinations investigated thus far. This is especially important when matching theory with polycrystalline experiments where the precise active surface may not be known a priori or where there is little control over the surfaces created. Clusters that contain weak-binding sites alongside active binding sites may still be active as the CO will prefer the stronger binding, more active sites. Presence of strong binding sites are the more likely to hide an active site on a surface and should be avoided. Within a cluster, the embedding shows how active site coordination or alloying ratios may affect the activity. For example: The bottom-right panel

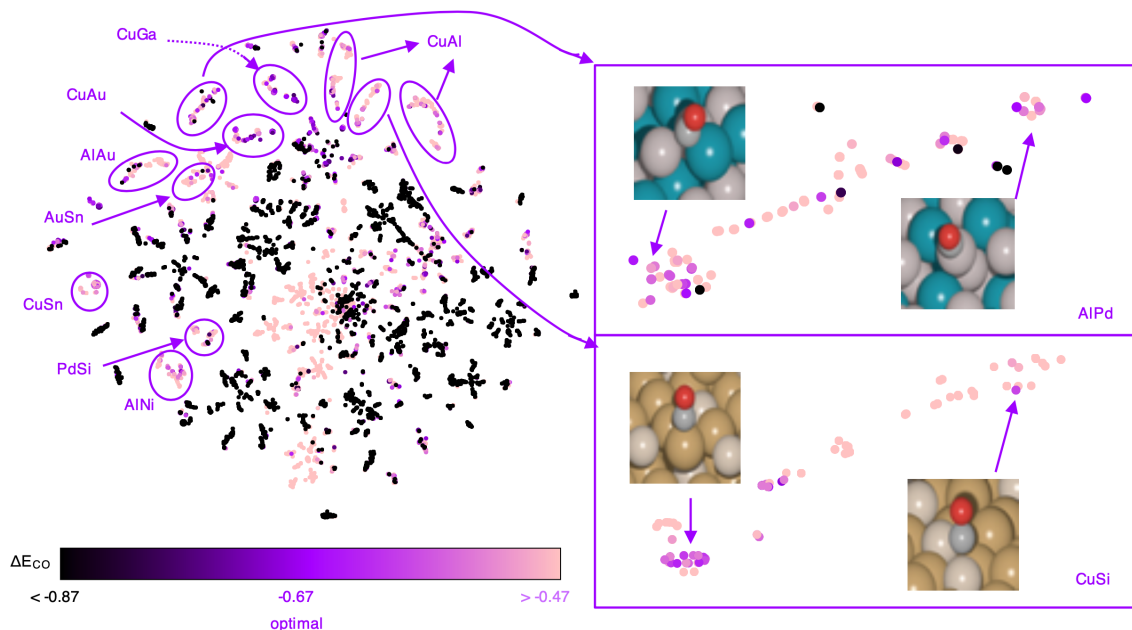


Figure 3-5: **Active site motif analysis.** Latent space visualization using t-SNE [113] of all the adsorption sites simulated with DFT. Proximity in this reduced space indicates similarity in the structures of the adsorption sites. Black points represent sites that bind too strongly; dark purple points represent sites with optimal binding; and purple points indicate sites that bind too weakly. Stronger binding sites are overlaid on top of weaker binding sites to indicate that the stronger sites have a greater influence on activity than weaker sites due to their greater thermodynamic stability. We labeled dark purple clusters/materials, because we expect them to be better candidates for further investigation and experimentation.

in Figure 3-5 shows that Si sites in CuSi alloys tend to bind too weakly, thereby suggesting that higher ratios of Al to Si may improve activity.

The machine learning model in this work also provides activity estimates of surfaces without explicit DFT calculations. The model predicts that approximately 81% of surfaces have non-ideal ΔE_{CO} values, defined as outside of the range $[-0.8 - \text{MAE}, -0.5 + \text{MAE}]$ eV. This considerably narrows the potential experimental search space. Likewise, the search space for bimetallic combinations can be reduced. If at least one surface must be predicted to be near-optimal for the CO_2RR , the search space can be reduced by 72%. If at least 10% of surfaces must be active (similar to the robust determination from above), then the search space is reduced by 93%.

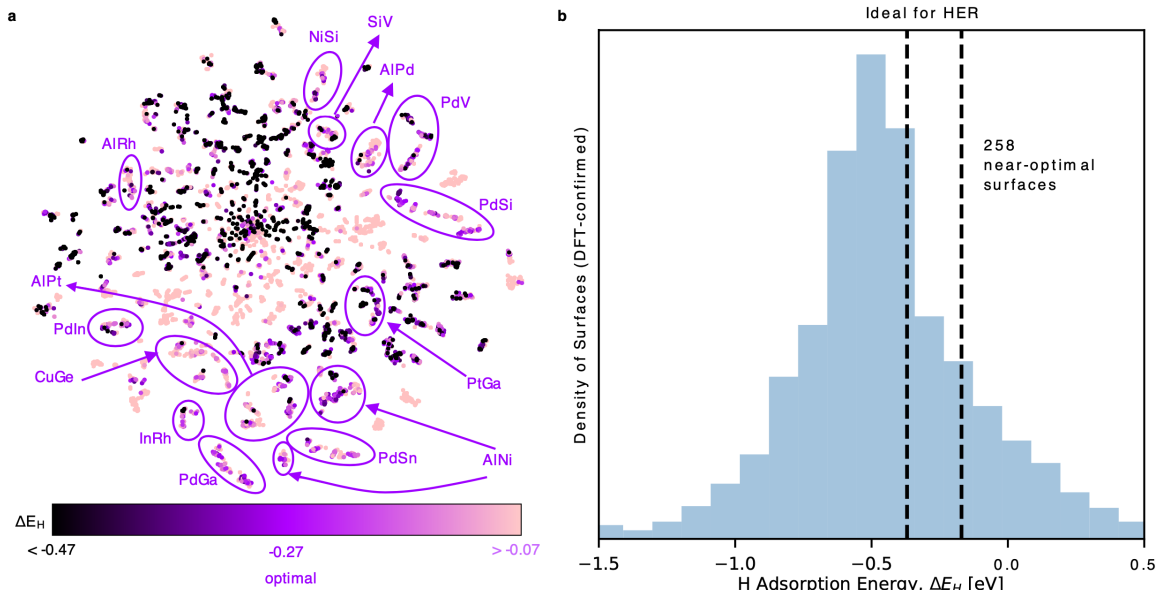


Figure 3-6: **Analysis of results for HER performance.** **a**, t-SNE [113] visualization of all the adsorption sites simulated with DFT. Similar to Figure 3-5, stronger binding sites are overlaid on top of weaker binding sites, and dark purple clusters/materials are labeled because we expect them to be better candidates for further investigation and experimentation. **b**, Normalized distribution of low-coverage ΔE_H values calculated by our DFT workflow. Dashed lines indicate the ± 0.1 eV range around the optimal ΔE_H value of -0.27 eV.

3.3.7 Discovering potential catalysts for H_2 evolution

The same types of analyses performed for CO_2RR can also be performed for HER. Figure 3-6a illustrates the t-SNE representation of the 23,141 adsorption sites where we used DFT to calculate ΔE_H . Figure 3-6b shows the distribution of DFT-calculated ΔE_H values, where we found 258 different surfaces with low-coverage ΔE_H values within 0.1 eV of the optimal value of -0.27 eV. All of these surfaces are listed in Supplementary Table A.2, and the bimetallic map of HER performance is shown in Supplementary Figure A-3. Similar to our analysis for CO_2RR , a number of the intermetallics that our screening study identified as having surfaces with near-optimal ΔE_H values have already been verified by various literature studies.[27, 185, 102]

Supplementary Figure A-3 shows that, in addition to Pt, there is a band of elements with comparable monometallic adsorption energies that tend to yield intermetallic surfaces with near-optimal ΔE_H values: As, Al, Si, Sb, Rh, and Pd. Many

of these elements also appear in the t-SNE diagram for HER as well (Figure 3-6a), suggesting that these elements warrant further study and experimentation.

3.4 Conclusions

We created a framework that produces and stores DFT data continuously and without the need for user intervention. This framework combines task and calculation management software with active machine learning and surrogate-based optimization to enable the automated, systematic selection and execution of DFT calculations. The framework produced 42,785 adsorption energy calculations to identify 130 candidate surfaces across 54 intermetallics with potentially high CO₂ reduction activity and 258 candidate surfaces across 102 intermetallics for hydrogen evolution. A number of the candidate surfaces found here have already been validated by literature experiments,[130, 167, 86, 27, 185, 102] suggesting that the unstudied candidates found in this screening warrant further study. The full list of potential surfaces is shown in Supplementary Tables A.1 and A.2, and shortened lists of candidate intermetallics are illustrated in Figures 3-5 and 3-6.

Our workflow for generating DFT data offers a combination of benefits that we have not yet seen in other literature frameworks. Our task and calculation management systems reduce the amount of time required to configure and process DFT calculations; our database of DFT results enables holistic analyses across numerous adsorption sites, surfaces, and material spaces; and our active machine learning/surrogate-based optimization workflow guides the discovery of candidate catalysts without the need for expert intuition. The flexibility of the framework also allows for expert-assisted guidance, allowing us to use the high-throughput DFT workflow to study specific sites, surfaces, or systems if needed. The combination of flexibility, automation, and machine learning guidance accelerates the theoretical discovery and study of catalysts for CO₂ reduction, H₂ evolution, or any other chemistry with a descriptor-performance scaling relationship.

A shortfall of our workflow is its heavy reliance on descriptor-performance rela-

tionships, which are used to guide the active learning algorithms. For example: this method will have issues with predicting CO₂ reduction activity for materials/surfaces that yield reaction mechanisms where ΔE_{CO} is independent of activity. Additionally, this method does not address other important aspects of catalyst performance, such as surface stability or catalyst cost. These issues are acceptable because this framework is used primarily as a tool to screen for candidate catalysts from a relatively large search space and to supplement experts’ intuitions with machine-derived suggestions. Our framework does not replace robust theoretical and experimental studies; it accelerates them by reducing search spaces to more tractable sizes and focusing expensive studies to systems that are more likely to yield interesting results. Future work could still be done to address the issues of diverse reaction mechanisms or multiple aspects of catalyst performance.

3.5 Methods

3.5.1 Enumerating search space

For each of the 1,499 intermetallic crystals we obtained from the Materials Project,[72] we used pymatgen[136] to enumerate symmetrically distinct facets with Miller indices between -2 and 2. Many intermetallic facets contained asymmetric top/bottom surfaces, and in those cases both were analyzed as well as distinct surfaces arising from the absolute position of the surface cut. In total 1,499 crystal structures were considered resulting in 17,507 unique surfaces and 1,684,908 unique adsorption sites. Surfaces were enumerated using ideal structures from the Materials Project instead of relaxed structures. This can cause differences in the number of enumerated facets, but it allows the enumeration to be completed without DFT relaxations for every bulk structure.

3.5.2 Addressing models with a variable number of features

One issue with our fingerprinting method is that it yields a variable number of features. For example, two vectors are needed to represent the first shell of a Cu-Al bridge site, but only one vector is needed to represent a Cu-Cu bridge. This issue can be addressed with zero-padding, but can be better modeled using a literature method[35] to make dummy features to replace features that are not populated naturally. To continue with the previous example, the first shell of the Cu-Al site would be represented by a vector of 4 numbers for the Cu element and 4 more numbers for the Al element, but a Cu-Cu site would be represented by 4 numbers for the Cu element and 4 dummy features. These dummy features are the average atomic number ($\bar{Z}$), the average Pauling electronegativity ($\bar{\chi}$), the average median adsorption energy of all elements we studied ($\bar{E}$), and a CN value of zero. Using averages of valid feature values reduces the bias induced from these dummy features while the CN value of zero ensures that no valid configuration could be confounded with the dummy features.

3.5.3 Calculating optimal adsorption energies

A descriptor/activity relationship[105] was used to predict catalyst activity and selectivity for CO₂RR given a free energy change, ΔG_{CO} . This relationship shows that a ΔG_{CO} of -0.17 eV yields an optimal activity and selectivity, corresponding to a ΔE_{CO} of -0.67 eV; reference the Supplementary Methods for more details. Similarly, literature relations were adopted to predict HER performance[53]. This relation predicted an optimal ΔG_{H} of -0.03 eV corresponding to a target ΔE_{H} of 0.27 eV.[133]

3.5.4 Calculating adsorption energies with DFT

The adsorption energy calculation workflow used in this study mimicked typical computational chemistry methods for calculating adsorption energies for sites of interest in the catalog.[134] Crystal structures from the Materials Project were relaxed using DFT. Relaxed crystal structures were used to generate facets of interest which were then relaxed with free surface atoms and fixed subsurface atoms. Finally the

adsorbate was placed on the surface at the relevant site and a final relaxation was completed. Final relaxed structures and their energies were comparable to traditional expert-made structures, as shown in Supplementary Note 3.

We performed all DFT calculations using: the *Vienna Ab initio Simulation Package*[91, 92, 89, 90] (*VASP*) implemented in *ASE*:[64] the RPBE functionals[57]; k-point grids of $4 \times 4 \times 1$; an energy cutoff of 350 eV; and the default pseudopotentials supplied by *VASP* version 5.4. Bulk relaxations were performed with a $10 \times 10 \times 10$ k-point grid and a 500 eV cutoff and only isotropic relaxation were allowed. Surfaces were replicated in the X/Y directions so that each cell vector was at least 4.5 Å. No spin magnetism or dispersion corrections were included. Slabs were replicated in the Z direction to a minimum of 7 Å and at least 20 Å of vacuum was included in between slabs. For some facets this led to slabs with large depth due to constraints in how the facet could be formed. Generally the bottom layers were fixed and defined as those atoms more than 3 Å from the top of the surface in the scaled Z direction. Adsorption energies were calculated relative to gas-phase CO(g) for CO, and relative to gas-phase $\frac{1}{2}\text{H}_2(\text{g})$ for H.

3.5.5 Automating DFT calculations

DFT calculations and other calculational tasks were coordinated in parallel and in an automated, high-throughput fashion. Each type of calculation and task was encoded as an interdependent task, and then dependency management software (Luigi[15]) was used to manage the tasks in parallel. For example, an adsorption energy calculation depends on a single surface relaxation which depends on a single bulk relaxation. Requesting an adsorption energy calculation automatically triggers the prerequisite bulk and slab relaxations and then adds the results to a database. When a new adsorption energy calculation is triggered that requires the same surface, the prerequisite bulk and slab results are read from the database instead of being regenerated. This differs from a fixed pipeline approach,[116] because intermediate tasks such as slab relaxations can be shared across multiple pipelines and at different times. DFT tasks were managed by a central FireWorks [71] database that distributed DFT relaxation

tasks across multiple computing clusters. This combined Luigi and FireWorks framework enabled high-throughput DFT calculations, because adsorption energies could be queried for any of the 1,684,908 enumerated sites without the need for human management of the intermediate tasks. All DFT relaxations were stored in a Mongo database that contains DFT calculation settings, the identity of the original crystal structure, the Miller indices of the slab, the exact, Cartesian location of the adsorption site, chemical information about the adsorption site such as local coordination, and the adsorption energy.

Our simulations occasionally yielded abnormal relaxations arising from desorptions, dissociations, surface reconstructions, or DFT non-convergence. These abnormalities were omitted from our regressions by excluding data from simulations that met any of the following criteria: simulations whose final maximum equilibrium force between any two atoms exceeded $0.5 \text{ eV}/\text{\AA}$, where the absolute value of the adsorption energy exceeded 4 eV , where any atom moved more than $0.5 \text{ \AA}$ during bare slab relaxation, where the adsorbate moved more than $1.5 \text{ \AA}$ during adsorption relaxation, and where any slab atom moved more than $1.5 \text{ \AA}$ during adsorption relaxation. These exclusion criteria were used as heuristics to reduce outliers. This approach may induce bias in the dataset if systematic portions of the search space are missing because they often fail for these reasons. Some DFT errors may be treated by automatically tuning the DFT calculation settings,[136] but these approaches are not robust across the full range of calculation errors in adsorption simulations.

The adsorption energy database, which was required to train machine learning models, was initially seeded with ΔE_{CO} and ΔE_{H} calculations for every unique adsorption site on a variety of surfaces, including the (100), (111), and (211) facets of the most stable crystal form of each element included in Supplementary Figure A-1. We also added all unique coordinations types (on-top Ni, on-top Fe, bridge Ni-Fe, etc.) up to a coordination number of two and then selected the corresponding surfaces with the smallest number of atoms. This ensured that the original machine learning model contained at least some data to begin with.

Chapter 4

Accelerated discovery of CO₂ electrocatalysts using active machine learning

This work originally appeared as: Miao Zhong, Kevin Tran,* Yimeng Min,* Chuanhao Wang,* Ziyun Wang, Cao-Thang Dinh, Phil De Luna, Zongqian Yu, Armin Sedighian Rasouli, Peter Brodersen, Song Sun, Oleksandr Voznyy, Chih-Shan Tan, Mikhail Askerka, Fanglin Che, Min Liu, Ali Seifitokaldani, Yuanjie Pang, Shen-Chuan Lo, Alexander Ip, Zachary Ulissi, and Edward H. Sargent. Accelerated discovery of CO₂ electrocatalysts using active machine learning. Nature, 581(7807):178–183, 2020. It has been edited to include the supplementary information in Appendix B.*

**These authors contributed equally.*

My contribution in this work involved performing the active learning portion of the research and writing the corresponding sections in the manuscript. I was also a primary editor of the entire manuscript.

4.1 Abstract

The rapid increase in global energy demand and the need to replace carbon dioxide (CO₂)-emitting fossil fuels with renewable sources have driven interest in chemi-

cal storage of intermittent solar and wind energy.[103, 158] Particularly attractive is the electrochemical reduction of CO₂ to chemical feedstocks, which uses both CO₂ and renewable energy.[41, 99, 109, 47, 100, 75] Copper has been the predominant electrocatalyst for this reaction when aiming for more valuable multi-carbon products,[66, 186, 143, 87, 125, 111, 39] and process improvements have been particularly notable when targeting ethylene. However, the energy efficiency and productivity (current density) achieved so far still fall below the values required to produce ethylene at cost-competitive prices. Here we describe Cu-Al electrocatalysts, identified using density functional theory calculations in combination with active machine learning, that efficiently reduce CO₂ to ethylene with the highest Faradaic efficiency reported so far. This Faradaic efficiency of over 80 per cent (compared to about 66 per cent for pure Cu) is achieved at a current density of 400 milliamperes per square centimetre (at 1.5 volts versus a reversible hydrogen electrode) and a cathodic-side (half-cell) ethylene power conversion efficiency of 55 ± 2 per cent at 150 milliamperes per square centimetre. We perform computational studies that suggest that the Cu-Al alloys provide multiple sites and surface orientations with near-optimal CO binding for both efficient and selective CO₂ reduction.[105] Furthermore, in situ X-ray absorption measurements reveal that Cu and Al enable a favourable Cu coordination environment that enhances C-C dimerization. These findings illustrate the value of computation and machine learning in guiding the experimental exploration of multi-metallic systems that go beyond the limitations of conventional single-metal electrocatalysts.

4.2 Introduction

To accelerate catalyst discovery, we developed a machine-learning-accelerated, high-throughput DFT framework[171] to screen materials ab initio. We provided this framework 244 different copper-containing crystals from The Materials Project[72], and it enumerated 12,229 surfaces and 228,969 adsorption sites. We then performed DFT simulations on a subset of these sites to calculate their CO adsorption energies

(see Appendix B). These data were then used to train an ML model, which was used to predict CO adsorption energies on all the enumerated adsorption sites. The framework then combined the ML-predicted CO adsorption energies with volcano scaling relationships[105] to predict the most catalytically active sites, which have CO adsorption energies (ΔE_{CO}) near to -0.67 eV—a value predicted to produce near-optimal activity in the volcano scaling relationship (Appendix B and Figures B-1 and B-2 for details on calculating the optimal ΔE_{CO} of -0.67 eV). These optimal sites were then simulated using DFT to provide additional training data for the ML model. Cycling among DFT simulation, ML regression, and ML prioritization yielded an automated framework that systematically searched for surfaces and adsorption sites with near-optimal CO adsorption energies. In total, the framework performed *ca.* 4,000 DFT simulations, yielding a set of candidates for experimental testing.

Among the candidate materials that the framework identified, we found that Cu-Au, Cu-Sn, Cu-Al, and Cu-In alloys were promising for active and selective CO₂ reduction. We determined this by creating a 2-dimensional activity and selectivity volcano plots for CO₂ reduction (Figures 4-1a and 4-1b) using the same method that was used in previous theoretical studies²⁸ (see Appendix B and Figure B-64 for details). As expected, Figure 4-1a shows that a CO binding energy near -0.67 eV is required for relatively high activity. It also shows that, given a CO binding energy *ca.* -0.67 eV, a H binding energy above *ca.* -0.5 eV is required for activity and that a H binding energy above *ca.* -0.2 eV is required for selectivity towards CO₂ reduction instead of H₂ evolution (Figures 4-1a and 4-1b).

Since these criteria were met by multiple copper alloy candidates, we pared the list of candidates by visualizing and analyzing them in a t-SNE diagram[113] (Figure 4-1c). Each point on this diagram represents one adsorption site for which we performed a DFT calculation. Points near each other tend to have similar coordination atoms and surface compositions (Appendix B). Thus, clusters of sites tend to represent different archetypes of adsorption sites (Figure 4-1d). Figure 4-1c shows that Cu-Al exhibits an abundance of adsorption sites and site types with near-optimal ΔE_{CO} values, suggesting that Cu-Al alloys may be active across a relatively wide range of

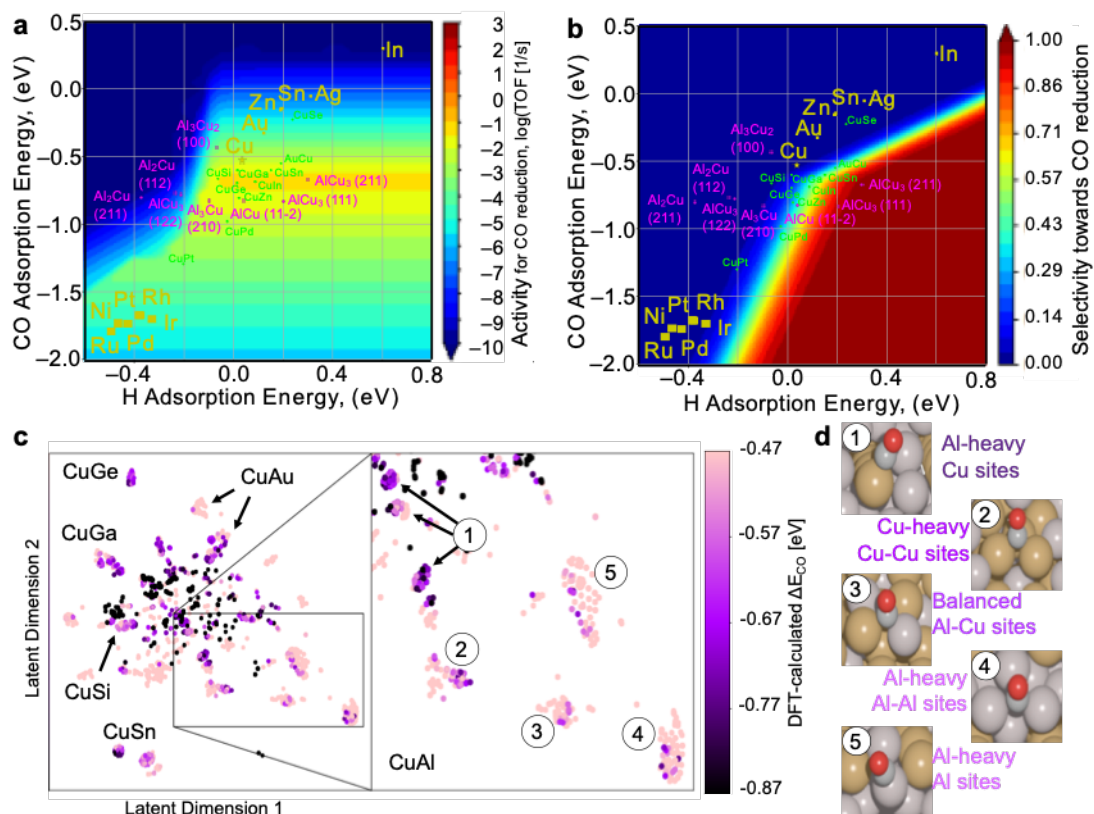


Figure 4-1: Screening of Cu and Cu based compounds using computational methods. a, A 2-dimensional activity volcano plot for CO₂ reduction. b, A 2-dimensional selectivity volcano plot for CO₂ reduction. CO and H adsorption energies in Figs. a and b were calculated with DFT. Points in yellow are average adsorption energies of monometallics; points in green are average adsorption energies of copper alloys; and points in magenta are average, low-coverage adsorption energies of Cu-Al surfaces. c, t-SNE[113] representation of *ca.* 4,000 adsorption sites that we performed DFT calculations on with Cu-containing alloys. The Cu-Al clusters are labeled numerically. d, Representative coordination sites for each of the clusters labeled in the t-SNE diagram. Each site archetype is labeled by the stoichiometric balance of the surface, i.e., Al heavy, Cu heavy, or balanced, and the binding site of the surface.

surface compositions and site types. The zoomed-in t-SNE diagram with example adsorption sites (Figure 4-1d) reveals that Al sites tend to bind CO too weakly; Cu sites surrounded by mostly Al atoms may bind CO too strongly; and Cu-Al bridge sites surrounded mostly by Cu atoms are predicted to be active. Cu-Al alloys’ low abundance of low ΔE_{CO} sites also suggests that Cu-Al may be resistant to CO over-binding issues. Therefore, Cu-Al alloys with a higher Cu content than Al may be active for CO₂ reduction.

4.3 Methods

4.3.1 Experimental methods

To test these hypotheses, we prepared experimentally a suite of Cu-Al model catalysts: ion-implanted Al-on-Cu and evaporated-and-etched Al-on-Cu (Appendix B and Figure B-3). Each catalyst shows a morphology similar to that of an evaporated pure Cu catalyst (Figures B-4–B-6). Compared with the pure Cu catalyst, which attained a C₂H₄ Faradaic efficiency (FE) of 35% at a current density of 600 mA cm⁻² in a 1 M KOH electrolyte in a flow-cell configuration (Figure B-7), both ion-implanted and evaporated-etched Al-on-Cu catalysts exhibited higher C₂H₄ FEs of ~60% under the same testing conditions. The CO FEs on both Cu-Al catalysts were suppressed to ~10%, one-third of that obtained using pure Cu (Figure B-8). Incorporating Al on Cu thus increased selectivity toward C₂H₄. Tafel slopes of C₂H₄ production (Figure B-8) for pure Cu, ion-implanted, and evaporated-and-etched Al-on-Cu are 180, 147 and 145 mV dec⁻¹, respectively, further highlighting the faster C-C dimerization kinetics with Al-on-Cu catalysts.

To estimate quantitatively the amount of Al incorporated near the Cu surface, we used surface-sensitive Auger electron spectroscopic (AES) analysis (Figures B-9 and B-10). AES typically provides compositional information on the top 1–3 nm of the samples and does so over a relatively large area (100 *μm*² in our studies)[36]. We estimated Al concentrations on surfaces of 4.5% and 25% for the ion-implanted and

evaporated-and-etched Al-on-Cu, respectively. SEM and EDX analyses confirmed no major change of morphologies nor Al concentrations for the ion-implanted and evaporated-and-etched Al-on-Cu samples before and after 5-hour reaction (Figures B-5, B-6, and B-12–B-15). (Detailed operating stability information in Supplementary Information and Figures B-16–B-19). Pourbaix diagrams[141] (Figure B-11) explain that both Cu and Al are cathodically protected at potentials more negative than their oxidation potentials of -1.4 V vs. RHE in a pH 14 electrolyte.

We therefore sought to develop an optimized and robust Cu-Al catalyst. We explored both thermal evaporation and co-sputtering followed by chemical etching where we synthesized de-alloyed nanoporous Cu-Al catalysts (see fabrication details in Supplementary Information). As shown in the scanning electron spectroscopy (SEM) and high angle angular dark field-scanning transmission electron microscopy (HAADF-STEM) images in Figures 4-2b and B-20, a nanoporous structure with pore sizes of 5–20 nm was formed. Compared to ion-implanted and evaporated-etched Al-on-Cu catalysts, the de-alloyed nanoporous Cu-Al catalysts may offer more catalytically active sites for adsorption and electroreduction of CO₂. Following 5 hours CO₂ electroreduction at a current density of 600 mA cm⁻², the morphology remained similar indicating a stable catalyst (Figure 4-2b). The grain size of the catalyst increased following reaction, potentially due to the surface reconstruction of Cu and Al in the electrolyte during the reaction. Energy-dispersive X-ray spectroscopy (EDX) analyses in TEM and SEM, electron energy loss spectroscopy (EELS) spectra, and elemental mapping in STEM, all confirmed a homogeneous distribution of Al and Cu in de-alloyed catalysts before and following 5 hours reaction (Figures 4-2c and B-21–B-24). We performed HAADF-STEM analysis and found that Cu (111) and (200) facets were observed with interplanar spacings of 0.211 and 0.182 nm (Figure B-25). AES analysis further revealed ~9% Al on the surface following the reaction respectively (Figures B-26 and B-27).

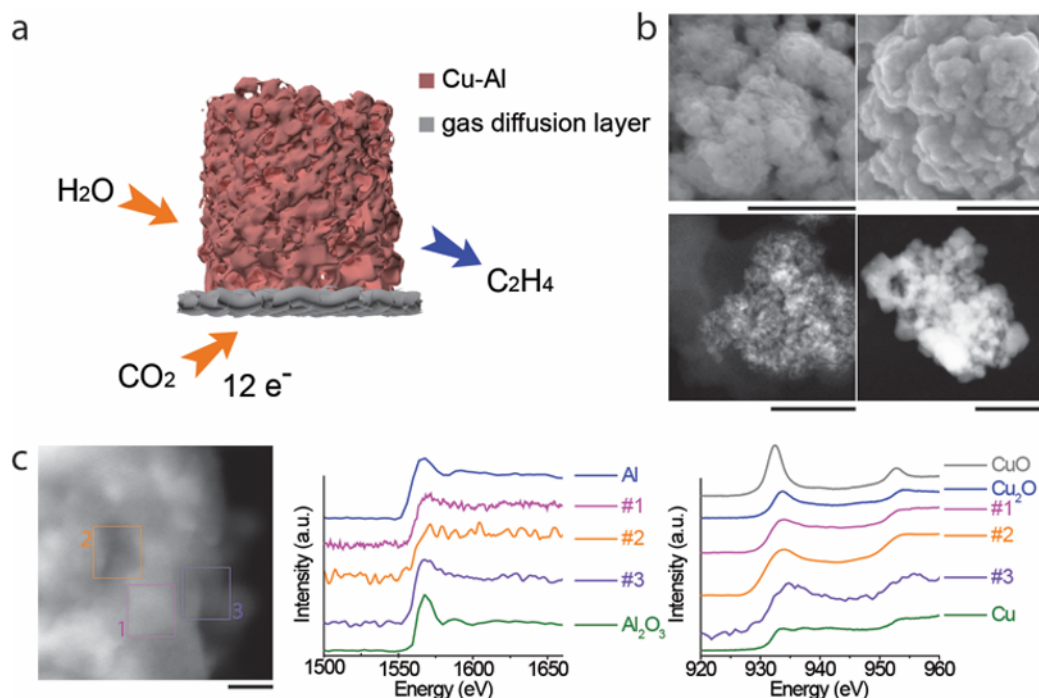


Figure 4-2: Schematic and characterization of de-alloyed Cu-Al catalyst. a, Schematic of a de-alloyed nanoporous Cu-Al catalyst on a gas diffusion layer for CO₂ electroreduction. b, Scanning electron microscopy (SEM) and high angle angular dark field-scanning transmission electron microscopy (HAADF-STEM) images of de-alloyed Cu-Al catalyst before (left) and after (right) 5 hours CO₂ electroreduction at an applied current density of 600 mA cm⁻² in flow cells. The scale bars for SEM images are 500 nm (top-left) and 200 nm (top-right). The scale bars for TEM images are 200 nm (bottom-left) and 100 nm (bottom-right). c, HAADF-STEM image, and electron energy loss spectroscopy (EELS) spectra of the de-alloyed Cu-Al catalyst. #1, 2, 3 curves in the EELS spectra represent the EELS results measured at #1, 2, 3 areas in the corresponding HAADF-STEM image. Al, Al₂O₃, CuO, Cu₂O and Cu EELS results are plotted as references. The scale bar is 5 nm.

4.3.2 Computational methods

Given the presence of Cu (111) and (100) surfaces, we used our ML model and DFT calculations to analyze how the ratio of Al to Cu on these surfaces affects ΔE_{CO} . First, we enumerated (using Delaunay triangulation[128]) the range of adsorption sites on the Cu (111) surfaces having different Al concentrations; and then predict ΔE_{CO} for these sites using the ML model, creating thereby a distribution of ΔE_{CO} values. We repeated this operation for the Cu (100) surfaces at different Al concentrations. The resulting distributions (Figures B-28a and B-28b) show that adding $\sim 12\%$ Al to the Cu (111) surface maximizes the density of sites with ΔE_{CO} values near the optimum of -0.67 eV and that adding 4–12% Al maximizes the density of optimal sites for the Cu (100) surface.

We performed density functional theory (DFT) calculations over the best ML predicted structures to characterize the changes in energy barriers in the major steps during CO₂ reduction. The reaction energy in the rate-determining step of C-C bond-making¹² decreased from 1.5 eV to 0.9 eV on Cu (111) and from 0.6 eV to 0.4 eV on Cu (100) with the benefit of Al incorporation (Figures B-28–B-31), which is consistent with ML predictions of increased C₂+ production with Al-containing Cu. The DFT results further showed that the reaction energy for forming HO(CH)CH, an intermediate of ethanol[182], was higher than that for forming CCH, an intermediate of C₂H₄[182] with Al-containing Cu (Figure B-32). Water near the Al atoms may act as a proton donor for the electrochemical dehydration reduction of HOCCH to CCH instead of hydrogenation of HOCCH to HO(CH)CH[182]. Thus, oft-produced alcohol was suppressed and the C₂H₄ production was promoted.

4.4 Results

4.4.1 Initial catalyst exploration

We tested the CO₂ electroreduction activity of de-alloyed Cu-Al catalysts with different Al concentrations on the surfaces. A C₂+ /C1 ratio of ~ 30 was obtained with

$\sim 10\%$ Al on the surface which is in line with the ML and DFT predictions (Figure B-33).

We then systematically evaluated the CO_2 electroreduction performance of the de-alloyed Cu-Al catalysts on the conventional carbon-based gas diffusion layer (C-GDL) substrates with $\sim 10\%$ Al at the surfaces at current densities from 200 to 800 mA cm^{-2} in 1 M KOH in flow cells (Figure 4-3a and 4-3b). To quantify FEs for each product, we carried out CO_2 electroreduction in a chronopotentiometry mode. As shown in Figure 4-4a, we achieved C_2H_4 FE of 80% at a current density of 600 mA cm^{-2} . This is a 2-fold increase compared to the 35% FE of pure Cu measured under the same conditions. An electricity-to-ethylene half-cell power conversion efficiency (C_2H_4 PCE) of 34% was achieved (Figure 4-3d), which is similar to the previously published highest half-cell C_2H_4 PCE of $\sim 30\%$ using a plasma-activated copper electrocatalyst¹³ with a C_2H_4 FE of 60% at a relatively smaller applied current density of $\sim 12 \text{ mA cm}^{-2}$ under the same electrolyte pH condition. An average C_2H_4 FE of $75\% \pm 4\%$ was obtained over 17 de-alloyed distinct Cu-Al on C-GDL samples ($\sim 10\%$ Al on the surfaces) under the same current density of 600 mA cm^{-2} . The overall C_2+ product was 85%–90% when we used the de-alloyed Al-based catalyst, appreciably higher than the 55–60% using the flat Cu catalyst (Figures 4-3c and B-8).

We further designed control catalysts—nanoporous Cu on C-GDL with a very limited amount of Al on the surface and having similar nanoporosity to that of the de-alloyed Cu-Al catalyst—to clarify the role of morphology (Appendix B and Figure B-34). AES analysis revealed that surface Al was decreased from 10% to 2–3% (Figure B-35). Considering the penetration depth of 1–3 nm in the AES, the surface Al concentration on nanoporous Cu should be lower than 2–3%. The C_2H_4 FE was decreased to 53% at the same current of 600 mA cm^{-2} (Figure 4-4a and B-36). Therefore, we conclude from these studies that incorporating Al on the Cu surface is crucial to promote C_2H_4 production.

The Cu-Al on C-GDL catalysts exhibited stable potentials between -1.8 and -2.1 V vs. RHE and a C_2H_4 FE of 75% over 5 hours of continuous operation at 600 mA cm^{-2} (Figures B-37 and B-38). After 5 hours, C-GDL gradually lost its hydrophobicity

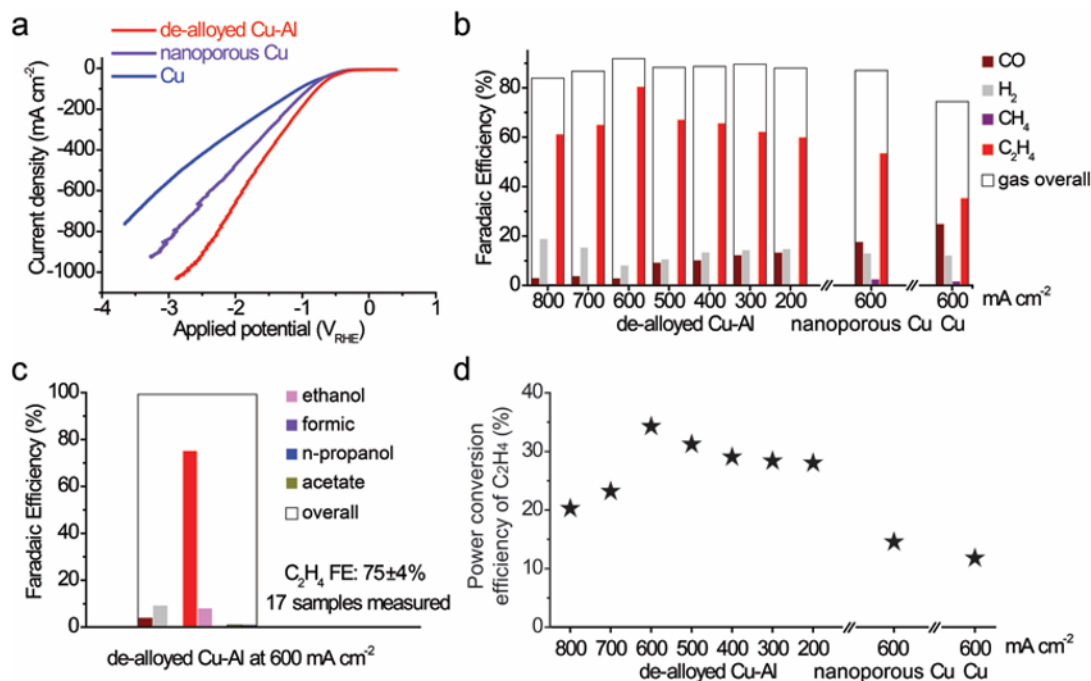


Figure 4-3: CO₂ electroreduction performance on de-alloyed Cu-Al, porous Cu and deposited Cu catalysts on C-GDL substrates in a 1 M KOH electrolytes. a, C₂H₄ production partial current density vs. potential with de-alloyed Cu-Al, nanoporous Cu and evaporated Cu catalysts. b, Faradaic efficiencies for gaseous products with de-alloyed Cu-Al catalysts at different applied current densities and with nanoporous Cu and evaporated Cu catalysts at a constant current density of 600 mA cm⁻² obtained using chronopotentiometry. c, Faradaic efficiencies for all products at an applied current density of 600 mA cm⁻² with 17 de-alloyed Cu-Al samples measured. d, Half-cell power conversion efficiency of C₂H₄ with de-alloyed Cu-Al catalysts at different applied current densities and with nanoporous Cu and evaporated Cu catalysts at a constant current density of 600 mA cm⁻² obtained using chronopotentiometry.

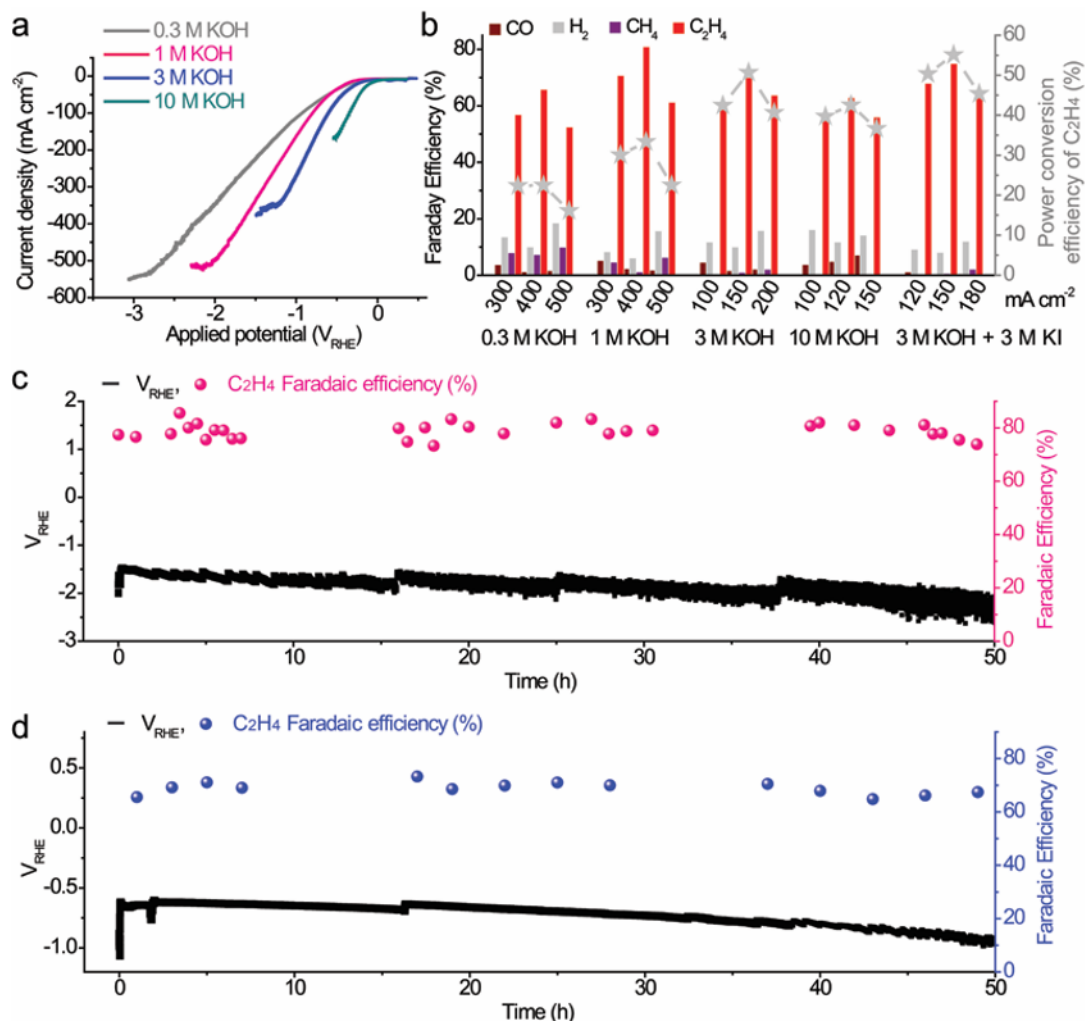


Figure 4-4: CO₂ electroreduction performance on de-alloyed Cu-Al catalysts on PTFE substrates in alkaline electrolytes at different pH values. a, C₂H₄ production partial current density vs. potential with de-alloyed Cu-Al in 0.3, 1, 3 and 10 M KOH electrolytes. b, Faradaic efficiencies for gaseous products with its corresponding C₂H₄ power conversion efficiencies of the de-alloyed Cu-Al catalysts in the different electrolytes and at different applied current densities. c, The CO₂ electroreduction stability of the carbon NPs/de-alloyed Cu-Al/PTFE electrode in a 1 M KOH electrolyte at an applied current density of 400 mA cm⁻². Left axis: potential (V vs. RHE) vs. time (s), right axis: C₂H₄ Faradaic efficiency (%) vs. time (s). d, The CO₂ electroreduction stability of the carbon NPs/de-alloyed Cu-Al/PTFE electrode in a 3 M KOH electrolyte at an applied current density of 150 mA cm⁻². Left axis: potential (V vs. RHE) vs. time (s), right axis: C₂H₄ Faradaic efficiency (%) vs. time (s).

and became flooded with 1 M KOH electrolyte[41]. Therefore, CO₂ could no longer diffuse to the catalyst surface for CO₂ reduction.

4.4.2 System optimization

To improve the device stability, we fabricated de-alloyed Cu-Al catalysts on the polytetrafluoroethylene (PTFE) substrates whose hydrophobicity was proved to be stable during operation in a strong alkaline electrolyte³ (Appendix B and Figures B-20, B-39, and B-40). Carbon NPs/graphite was coated on the de-alloyed Cu-Al surface to create a sandwich structure, which distributed the current uniformly over the catalyst to stabilize its surface during reaction[41]. As shown in Figures 4-4b, 4-4c, and B-41, we achieved C₂H₄ FEs over 80% in 1 M KOH at a current density of 400 mA cm⁻² over 50 hours.

To improve device efficiency, we studied Cu-Al performance at different pH conditions, which has been shown to affect C-C dimerization[175]. Experimentally, we found that 3 M KOH (pH 14.5) was optimal to achieve *ca.* 50% half-cell C₂H₄ PCE with C₂H₄ FE of 70–72% at a current density of 150 mA cm⁻² over 50 hours (Figures 4-4b and 4-4d). We then further optimized the cation concentration by adding an additional 3 M KI into the electrolyte. KI was chosen because K⁺ cation and I⁻ anion are known to increase CO₂ reduction activity by accelerating the hydrogenation of the key adsorbed CO intermediate.[41, 104] This further diminished the CO FE to below 0.3% and reduced H₂ production by 2–3%, leading to a 2–3% increase of C₂H₄ FE to 72–75%. As a result, we achieved 55±2% half-cell C₂H₄ PEC (over 10 distinct samples) at 150 mA cm⁻² over 50 hours in a 3 M KOH and 3 M KI solution (Figures 4-4b and B-42).

4.4.3 System characterization

No obvious leaching of Al and Cu into the solution was observed by inductively coupled plasma atomic emission spectroscopy (ICP-AES) analysis (Figure B-43), indicating that Cu-Al catalysts remained stable during the tests. Pourbaix diagrams

suggest that Cu and Al metals are cathodically protected at potentials more negative than 1.47 VRHE at pH 14, which corresponds to a current density above *ca.* 320 mA cm⁻² for Cu-Al catalysts in 1 M KOH (Figure 4-4a). We also did not observe obvious leaching of Al at a lower potential of *ca.* -0.4 VRHE at pH 14.5 in the ICP-AES analysis. The Cu-Al catalyst remained physically intact after reaction. This may suggest that the Pourbaix diagram is not suited to explain the stability of alloys with nanostructured crystal edges and dangling bonds on their surfaces[189]. As reported[189], materials at nanoscale are usually more stable than its bulk under harsh electrochemical conditions beyond the predictions of the Pourbaix diagram. Another hypothesis is that Al in the subsurface could be protected by the stable Cu on the top surface, thereby improving Cu-Al stability

To explore this hypothesis, we used DFT to analyze the reaction energy changes when Al is placed in the subsurface of either Cu (111) or Cu (100). We added a total concentration of 12% and 4% Al for Cu (111) and Cu (100), respectively, which were the ML-predicted optimum Al concentrations for each surface. The reaction energy in the rate-determining steps is lower with Al in the subsurface compared to that of pure Cu (Figures B-44 and B-45).

To further understand the reasons for the improved performance of our Cu-Al electrocatalyst, we performed both theoretical and experimental studies of the local coordination information during CO₂ electroreduction. Experimentally, we performed in situ synchrotron X-ray absorption near-edge structure (XANES) analysis under these same testing conditions. Al-Al and Al-O bonds were observed in soft X-ray absorption spectra (Figure B-46). Cu-O bonding was observable via both ex situ and in situ XANES analyses with the de-alloyed Cu-Al catalyst before, during and after the reaction. The intensity of the in situ Cu-O peak is smaller than that of the ex situ ones, indicating Cu-O in the de-alloyed Cu-Al catalyst was partially reduced to Cu during the reaction. In contrast, no obvious Cu-O bonding was observed for the pure Cu during the ex situ and in situ analyses under the same testing conditions (Supplementary Information and Figure B-47). We propose that a mixture of Cu-Cu and Cu-O was formed in the de-alloyed Cu-Al catalysts during the reaction.

This may contribute to the favourable coordination environment that accelerates C-C dimerization, which is in line with the previous report[183]. DFT simulations suggest that the reaction energies in the CO₂ reduction energy diagrams are reduced with O on the top surface or in the subsurface to the Cu-Al model (Figures B-48–B-63).

4.5 Conclusions

To conclude, we developed a new de-alloyed Cu-Al catalyst for highly active and selective CO₂ electroreduction to C₂H₄. We successfully demonstrate the discovery of electrocatalysts by combining volcano relationships, DFT, and active ML to optimize catalyst performance. We further optimized the theoretically-discovered electrocatalyst by developing a multi-layered device and then optimizing reaction conditions for multi-carbon production. The findings suggest avenues to multi-metal catalysts that outperform single-component catalysts by utilizing an intermediate-binding-optimization and reaction-electrolyte-optimization strategy for multi-carbon production via CO₂ electroreduction.

Chapter 5

Methods for comparing uncertainty quantifications for material property predictions

This work originally appeared as: Kevin Tran, Willie Neiswanger,* Junwoong Yoon, Qingyang Zhang, Eric Xing, and Zachary W Ulissi. Methods for comparing uncertainty quantifications for material property predictions. Machine Learning: Science and Technology, 1(025006), 2020. It has been edited to include the supplementary information in Appendix C. *These authors contributed equally.*

My contribution in this work involved defining the evaluation methods and procedures alongside Willie. I then codified the procedures and tested the neural network, neural network ensemble, Gaussian process, and convolution-fed-Gaussian-process. Zack and I co-invented the convolution-fed-Gaussian-process. All authors wrote the sections that corresponded to the work that they did. I wrote all other sections (e.g., abstract, introduction, conclusions).

5.1 Abstract

Data science and informatics tools have been proliferating recently within the computational materials science and catalysis fields. This proliferation has spurred the

creation of various frameworks for automated materials screening, discovery, and design. Underpinning these frameworks are surrogate models with uncertainty estimates on their predictions. These uncertainty estimates are instrumental for determining which materials to screen next, but the computational catalysis field does not yet have a standard procedure for judging the quality of such uncertainty estimates. Here we present a suite of figures and performance metrics derived from the machine learning community that can be used to judge the quality of such uncertainty estimates. This suite probes the accuracy, calibration, and sharpness of a model quantitatively. We then show a case study where we judge various methods for predicting density-functional-theory-calculated adsorption energies. Of the methods studied here, we find that the best performer is a model where a convolutional neural network is used to supply features to a Gaussian process regressor, which then makes predictions of adsorption energies along with corresponding uncertainty estimates.

5.2 Introduction

The fields of catalysis and materials science are burgeoning with methods to screen, design, and understand materials.[117, 55, 155, 9] This research has spurred the creation of ML models to predict various material properties. Unfortunately, the design spaces for these models are sometimes too large and intractable to sample completely. These under-sampling issues can limit the training data and therefore the predictive power of the models. It would be helpful to have an UQ for a model so that we know when to trust the predictions and when not to. More specifically: UQ would enable various online, active frameworks for materials discovery and design (e.g., active learning,[163] online active learning,[29] Bayesian optimization,[45] active search,[49] or goal oriented design of experiments[78]).

Such active frameworks have already been used successfully in the field of catalysis and materials informatics. For example: Peterson[142] has used a neural network to perform online active learning of nudged elastic band (NEB) calculations, reducing the number of force calls by an order of magnitude. Torres *et al.*[168] have also used

online active learning to accelerate NEB calculations, but they used a GP model instead of a neural network. Jinnouchi *et al.*[76] have used online active learning to accelerate molecular dynamics simulations. These methods are all underpinned by models with UQ, which have garnered increasing attention.[144, 131]

The goal of UQ is to quantify accurately the likelihood of outcomes associated with a predicted quantity. For example, given an input for which we wish to make a prediction, a predictive UQ method might return a confidence interval that aims to capture the true outcome a specified percentage of the time or might return a probability distribution over possible outcomes. Performance metrics for predictive UQ methods aim to assess how well a given quantification of the probabilities of potential true outcomes adheres to a set of observations of these outcomes. Some of the performance metrics for predictive UQ are agnostic to prediction performance—they provide an assessment of the uncertainty independent of the predictive accuracy (i.e. a method can predict badly, but could still accurately quantify its own uncertainty).

We have seen few[74, 153] comparisons of different methods for UQ within the field of catalysis and materials informatics. Here we examine a protocol[93, 96] for comparing the performance of different modeling and UQ methods (Figure 5-1). We then illustrate the protocol on a case study where we compare various models’ abilities to predict DFT calculated adsorption energies. We also offer anecdotal insights from our case study. We acknowledge that such insights may not be transferable to other applications, but we find value in sharing them so that others can build their own intuition.

5.3 Methods

5.3.1 Dataset information

All regressions in this paper were performed using a dataset of 47,279 DFT calculated adsorption energies created with GASpy[171, 169]. Within this dataset, there were 52 different elements within the 1,952 bulk structures used as bases for the adsorption

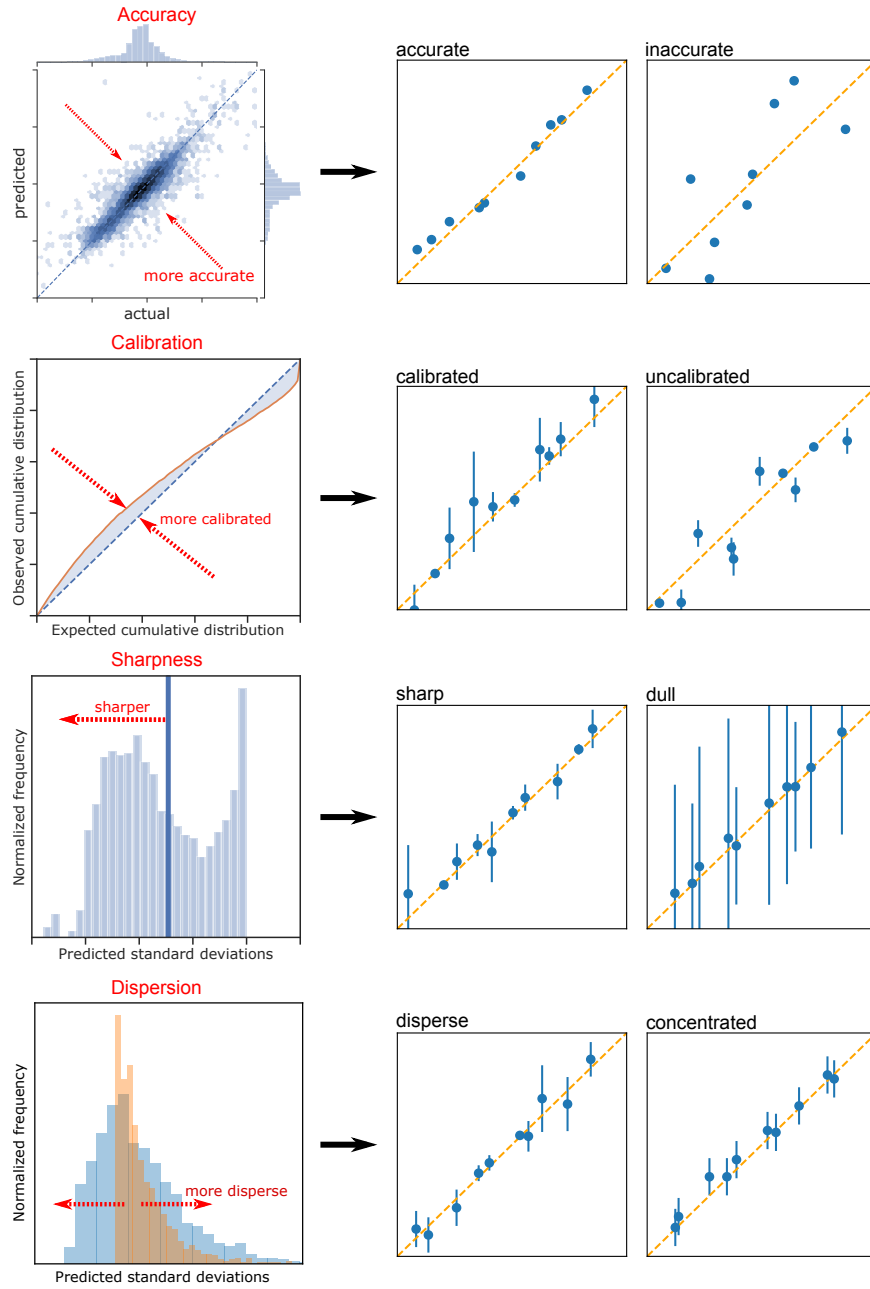


Figure 5-1: Overview of proposed procedure for judging the quality of models with uncertainty estimates. First and foremost, the models should be accurate. Second, the models should be “calibrated”, which means that their uncertainty estimates should be comparable with their residuals. Third, the models should be “sharp”, which means that their uncertainty estimates should be low. Lastly, the models should be “disperse”, which means that the distribution of the uncertainty estimates should be wide. This study demonstrates how to visualize and quantify these characteristics so that different methods of UQ can be compared objectively.

surfaces. The 61 bulk structures that contained one element encompassed 5,844 of the adsorption calculations; the 1,057 bulk structures that contained two elements encompassed 31,651 of the calculations; the 774 bulk structures that contained three elements encompassed 9,139 of the calculations; and the 60 bulk structures that contained four or five elements encompassed 645 of the calculations. The dataset also comprised 9,102 symmetrically distinct surfaces and 29,843 distinct coordination environments (as defined by the surface and the adsorbate neighbors). Lastly, the dataset comprised 21,269 H adsorption energies; 18,437 CO adsorption energies; 3,464 OH adsorption energies; 2,515 O adsorption energies; and 1,594 N adsorption energies.

GASpy performed all DFT calculations using the Vienna Ab-initio Simulation Package (VASP)[91, 92, 89, 90] version 5.4 implemented in the Atomic Simulation Environment (ASE)[64]. The revised Perdew-Burke-Ernzerhof (rPBE) functionals[57] were used along with VASP’s pseudopotentials, and no spin magnetism or dispersion corrections were used. Bulk relaxations were performed with a $10 \times 10 \times 10$ k-point grid and a 500 eV cutoff, and only isotropic relaxation were allowed during this bulk relaxation. Slab relaxations were performed with k-point grids of $4 \times 4 \times 1$ and a 350 eV cutoff. Slabs were replicated in the X/Y directions so that each cell was at least 4.5 Å wide, which reduces adsorbate self-interaction. Slabs were also replicated in the Z direction until they were at least 7 Å thick, and at least 20 Å of vacuum was included in between slabs. The bottom layers of each slab were fixed and defined as those atoms more than 3 Å from the top of the surface in the scaled Z direction.

To split the data into train/validate/test sets, we enumerated all adsorption energies on monometallic slabs and added them to the training set manually. We did this because some of the regression methods in this paper use a featurization that contains our monometallic adsorption energy data[171], and so having the monometallic adsorption energies pre-allocated in the training set prevented any information leakage between the training set and validation/test sets. After this allocation, we performed a 64/14/20 train/validate/test split that was stratified[166] by adsorbate. We then used the validation set’s results to tune various hyperparameters manually. After tuning, we calculated the test set results and present them in this paper exclusively.

Note that the test results were obtained using models that were trained only using the training set, not the validation set. This is acceptable because we only seek to compare methods here, not to optimize them.

Note that random splits such as this may yield overly optimistic model results. If a model created with the training set is meant to make extrapolative predictions in feature domains outside of the training set, then it may be appropriate to use a train/validate/test split using k-means clustering[122] rather than random splitting. If the model is meant to be used in an online and iterative fashion, then it may be appropriate to use a time-series split[69]. If the model is meant to be used to interpolate within a given feature space, then the basic random split may be appropriate. We chose to use a basic random split in this work to simplify the results for illustrative purposes. Future work for different applications should use splitting methods that align with the intended use of the models to be generated.

5.3.2 Regression methods

We explore various methods that aim to quantify the uncertainty for regression procedures where the predicted quantity is a continuous variable. To standardize the assessment of performance, we ensure that each UQ method returns predictive uncertainty results in a consistent format: a distribution over possible outcomes of the predicted quantity for any specified input point. This result format allows us to compute all the predictive uncertainty performance metrics which we introduce in subsequent sections. Figure 5-2 illustrates all of the methods we investigate in this study, and we describe each method in detail below.

NN: To establish a baseline for predictive accuracy, we re-trained a previously reported NN[184, 12] on this study’s training set. This NN model projects a three-dimensional atomic structure into a graph, which is then fed into convolutional layers to extract local atomic information for predicting global target properties. In this case, we predict DFT-calculated adsorption energies, ΔE . The graph consists of nodes representing atoms and edges representing distances between atoms. The NN updates the node features using the local information extracted in the convolutional layers,

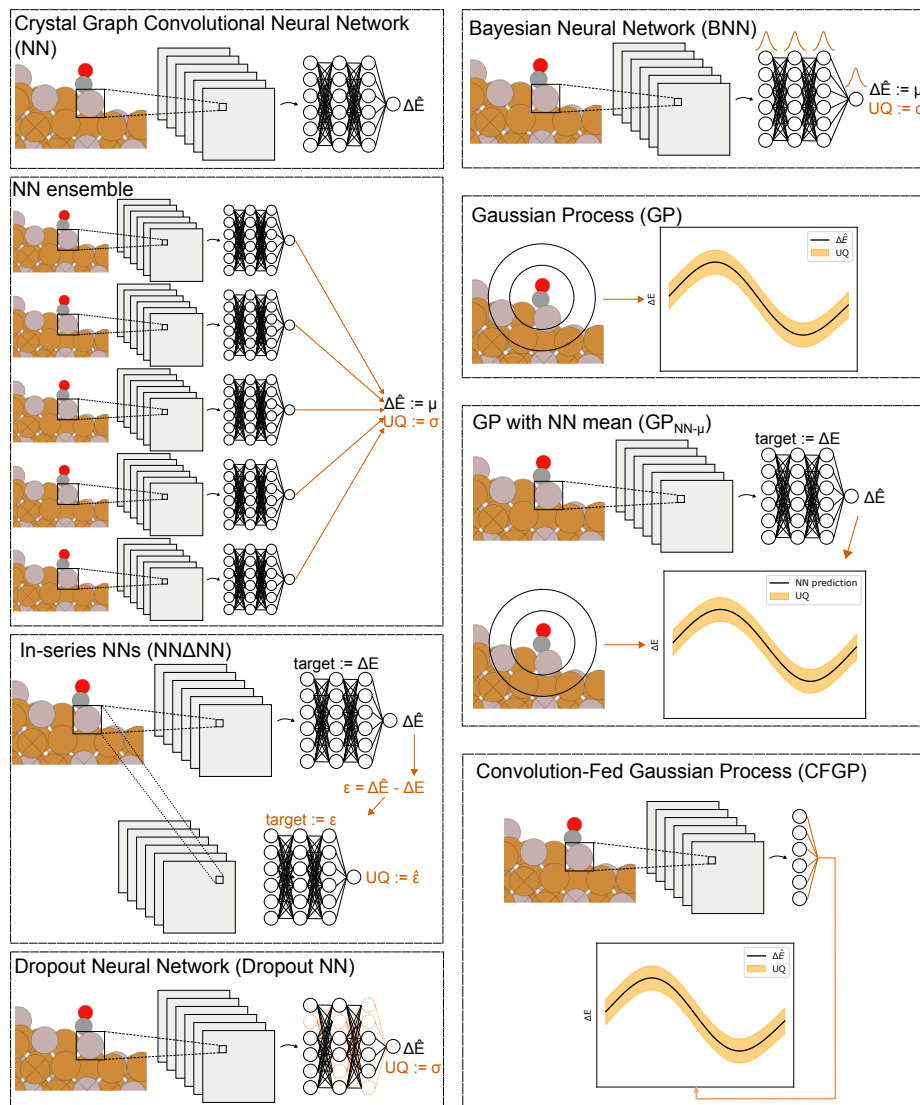


Figure 5-2: Overview of the various UQ methods we investigated in this study. ΔE represents DFT-calculated adsorption energies; $\Delta \hat{E}$ represents ML-predicted adsorption energies; UQ represents ML-predicted uncertainty quantifications; μ represents the mean of a sample of points; σ represents the standard deviation of a sample of points; ϵ represents the residuals between DFT and ML; and $\hat{\epsilon}$ represents the residuals between ML-predicted ϵ and the actual ϵ .

then hidden layers in the NN maps the node features to the adsorption energies. Reference Back *et al.*[12] for additional details.

NN Ensemble: We created an ensemble of NNs by 5-fold subsampling the training data and then training individual NN models on the 5 folds. Each individual NN model’s architecture is identical to the base NN architecture outlined previously. The only differences are their training sets and their individually randomized initial weights. For the final prediction of the ensemble we computed the mean of the set of models’ predictions, and for the ensemble’s estimate of uncertainty we computed the standard deviation of the set of predictions.

BNN: The aim of Bayesian Neural Network (BNN) is to determine the posterior distribution of model parameters rather than a single optimal value of the parameters. In practice, inferring true posterior distributions is very difficult and even infeasible in most cases. Thus, we approximate the model posterior to be as close as possible to the true posterior. The same NN architecture was used, but we converted the NN into BNN by assigning posterior distributions to all model parameters in the hidden layers in the NN model. The BNN then approximated the true posterior distributions using variational inference so that it could use the approximated posterior to predict the adsorption energies. We sampled the model parameters 20 times from the approximated posterior distributions, and used the mean of these predictions as the final prediction and the standard deviation of these predictions as the estimation of uncertainty. We implemented the BNN and performed variational inference using Pyro.[16]

Dropout NN: Dropout Neural Networks (Dropout NN) have been shown to approximate Bayesian models.[46] We created a Dropout NN by first replicating the exact architecture used to create the convolutional NN outlined previously. Then we enforced a random dropout rate of 30% in the dense hidden layers that followed the convolutional layers. The nodes were randomly dropped out during both training and prediction. To make predictions, we sampled the Dropout NN 20 times. The mean of the predictions was used as the final prediction of the Dropout NN, and the standard deviation of the predictions was used as the estimation of uncertainty.

NN Δ NN: Suppose we have trained a NN. We may aim to empirically fit an additional mapping that predicts the error of the first NN. Here we show in-series NNs (NN Δ NN), which trains a secondary NN to predict the residuals of the initial NN. When training the first NN, we hold out 10% of the training data. Afterwards, we use the residuals of the initial NN on the held-out portion as training data for the second NN. After the secondary training, this second NN can predict residuals for the first NN on some new set of input data. The predictions of the second NN can then be used as uncertainty estimates. Note that both the NNs included within the NN Δ NN were constructed using the same convolutional architecture outlined previously.

GP: GPs are one of the most common regression methods for producing UQs, and so we use them here as a baseline. We fit a standard GP using the same exact features that we used in previous work.[171] These features are defined by the elements coordinated with the adsorbate and by the elements of its next-nearest neighbors. Specifically: We use the atomic numbers of these elements, their Pauling electronegativity, a count of the number of atoms of each element near the adsorbate, and the median adsorption energy between the adsorbate and the elements. To ensure that these features interacted well with the GP’s kernel, we normalized each of the features to have a mean of zero and standard deviation of one. Reference Tran & Ulissi[171] for additional details. To define the GP, we assumed a constant mean and used a Matern covariance kernel. We trained the length scale of the Matern kernel using the Maximum Likelihood Estimation (MLE) method. All GP training and predictions were done with GPU acceleration as implemented in GPyTorch.[48].

GP_{NN- μ} : GPs are Bayesian models in which a prior distribution is first specified and then updated given observations to yield a posterior distribution. The mean of this posterior distribution is used for regression, and the covariance matrix is used for UQ. Typically, in lieu of any additional prior knowledge, practitioners will take the prior distribution to have zero-mean. However, we could instead supply an alternative curve for the prior mean, and then perform the usual Bayesian updates to compute the posterior of this GP given observations. Here, for the GP prior mean, we supply the prediction given by a single pre-trained NN. We call this method GP with NN

mean ($\text{GP}_{\text{NN}-\mu}$). For the input features of this GP, we used the same exact features we used for the plain GP—i.e., the vector of atomic numbers, electronegativity, etc. For the covariance kernel of this GP, we used a Matern kernel where we fit the kernel hyperparameters using MLE. All GP training and predictions were done with GPU acceleration as implemented in GPyTorch.[48].

CFGP: A limitation of using this formulation of a GP with NN-predicted mean is that it requires the use of hand-crafted features for the GP. This requirement reduces the transferability of the method to other applications where such features may not be readily available. To address this, we formulated a different method where we first train a NN (as described previously) to predict adsorption energies and then fix the network’s weights. Then we use the 46 pooled outputs of the convolutional layers of the network as features in a new GP. The GP would then be trained to use these features to produce both mean and uncertainty predictions on the adsorption energies. We call this a CFGP. Note that we normalized the 46 convolution outputs of the NN so that each output would have a mean of zero and a standard deviation of one across the training set. To define the GP, we assumed a constant mean and used a Matern covariance kernel. We trained the length scale of the Matern kernel using the MLE method. All GP training and predictions were done with GPU acceleration as implemented in GPyTorch.[48].

5.3.3 Performance metrics

We used five different metrics to quantify the accuracy of the various models: MDAE, RMSE, MAE, MARPD, and R^2 correlation coefficient (R^2). We used MDAE because is insensitive to outliers and is therefore a good measure of accuracy for the majority of the data. We used RMSE because it is sensitive to outliers and is therefore a good measure of worst-case accuracy. We used MAE because it lies between MDAE and RMSE in terms of sensitivity to outliers. We used MARPD and R^2 because they provide normalized measures of accuracy that may be more interpretable for those unfamiliar with adsorption energy measurements in eV. MARPD values were calculated with Equation 5.1:

$$MARPD = \frac{1}{N} \sum_{n=1}^N \left| 100 \cdot \frac{\hat{x}_n - x_n}{|\hat{x}_n| + |x_n|} \right| \quad (5.1)$$

where n is the index of a data point, N is the total number of data points, x_n is the true value of the data point, and $\hat{x}_n$ is the model’s estimate of x_n . In this case, x_n is a DFT-calculated adsorption energy and $\hat{x}_n$ is the surrogate-model-calculated adsorption energy. The ensemble of these metrics provide a more robust view of accuracy than any one metric can provide alone.

To assess the calibration (or “honesty”) of these models’ UQs, we created calibration curves. A calibration curve “displays the true frequency of points in each interval relative to the predicted fraction of points in that interval”, as outlined by Kuleshov *et al.*[93]. In other words: We used the standard deviation predictions to create Gaussian random variables for each test point and then tested how well the residuals followed their respective Gaussian random variables. Thus “well-calibrated” models had residuals that created Gaussian distributions whose standard deviations were close to the model’s predicted standard deviations. We discuss calibration curves in more detail in the Results section alongside specific examples. We also calculated the calibration errors[93] of our models, which is a quantitative measure of calibration.

As Kuleshov *et al.*[93] also pointed out, well-calibrated models are necessary but not sufficient for useful UQs. For example: A well-calibrated model could still have large uncertainty estimates, which are inherently less useful than well-calibrated and small uncertainty estimates. This idea of having small uncertainty estimates is called “sharpness”, and As Kuleshov *et al.*[93] define it with Equation 5.2:

$$sha = \frac{1}{N} \sum_{n=1}^N var(F_n) \quad (5.2)$$

where $var(F_n)$ is the variance of the random variable whose cumulative distribution function is F at point n . This is akin to the average variance of the uncertainty estimates on the test set. Here we propose and use a new formulation (Equation 5.3) where we add a square root operation. This operation gives the sharpness the same units as the predictions, which provides us with a more intuitive reference. In other

words: Sharpness is akin to the average of the ML-predicted standard deviations.

$$sha = \sqrt{\frac{1}{N} \sum_{n=1}^N var(F_n)} \quad (5.3)$$

Another consideration is the dispersion of the uncertainty estimates. If a model predicts a constant value for uncertainty, it may still be able to perform well with regards to calibration or sharpness. Constant values for uncertainty are likely to fail when models are used to make predictions outside the bounds of the training data. One way to address this issue is to calculate the C_v [96]. See Equation 5.4:

$$C_v = \frac{\sqrt{\frac{\sum_{n=1}^N (\sigma_n - \mu_\sigma)^2}{N-1}}}{\mu_\sigma} \quad (5.4)$$

where σ_n is the predicted standard deviation of point n , μ_σ is the average value of σ_n , and N is the total number of test points. Low values of C_v indicate a narrow dispersion of uncertainty estimates, which may suggest poor performance in out-of-domain predictions. Thus a higher C_v may indicate more robust uncertainty estimates. But as Scalia *et al.*[153] point out, the optimal dispersion is a function of the validation/test data distribution. Therefore, C_v should be used as a secondary screening metric rather than a primary performance metric.

We also assessed the performance of each predictive uncertainty method by comparing their NLL values the test set. For each test point, we established a Gaussian probability distribution using the mean and uncertainty predictions of each UQ model. Then we calculated the conditional probability of observing the true value of the test point given the probability distribution created from the UQ; this is the likelihood of one test point. We then calculated the product of all the likelihoods of all test points, which yielded the total test likelihood. It follows that better UQ methods yield higher total likelihood values. Equivalently, we could calculate the natural logarithms of each likelihood, sum them, and then take the negative of this value; this is NLL. Equation 5.5 shows how we calculated NLL:

$$NLL = - \sum_{i=1}^n \ln P(y_i | N(\hat{y}_i, \hat{\sigma}_i^2)) \quad (5.5)$$

where y_i is the true value of a test point, $\hat{y}_i$ is a model’s predicted mean value at that test point, $\hat{\sigma}_i^2$ is the model’s predicted variance at that test point, n is the set of all test points, and $N(x, y)$ is a normal distribution with mean x and variance y . Note how the NLL value depends on the size and location of the test set. This means that the absolute value of NLL changes from application to application, and so a “good” NLL value must be contextualized within a particular test set. Within a test set, a lower NLL value indicates a better fit. We also note that we assumed Gaussian distributions for our UQ methods’ predictions. This assumption does not necessarily need to be applied, meaning that the normal distribution in Equation 5.5 may be replaced with any other appropriate distribution.

We use NLL because it provides an overall assessment that is influenced by both the predictive accuracy of a method as well as the quality of its UQ. Previous work [50, 37] has shown the NLL to be a strictly proper scoring rule, which intuitively means that it provides a fair quantitative assessment (or score) for the performance of the UQ method, and that it can be decomposed into terms that relate to both calibration and sharpness. NLL is also a popular performance metric that has been used to quantify uncertainty in a variety of prior work [94] and provides an additional single score for UQ methods.

5.4 Results

5.4.1 Illustrative examples

Let us first discuss the results of our NN ensemble for illustrative purposes. Figure 5-3 contains a parity plot, calibration curve, and predicted-uncertainty distribution of our NN ensemble model. The parity plot shows the accuracy of the model; the calibration curve shows the honesty of the model’s uncertainty predictions; and the uncertainty distribution shows the sharpness of the model’s uncertainty predictions.

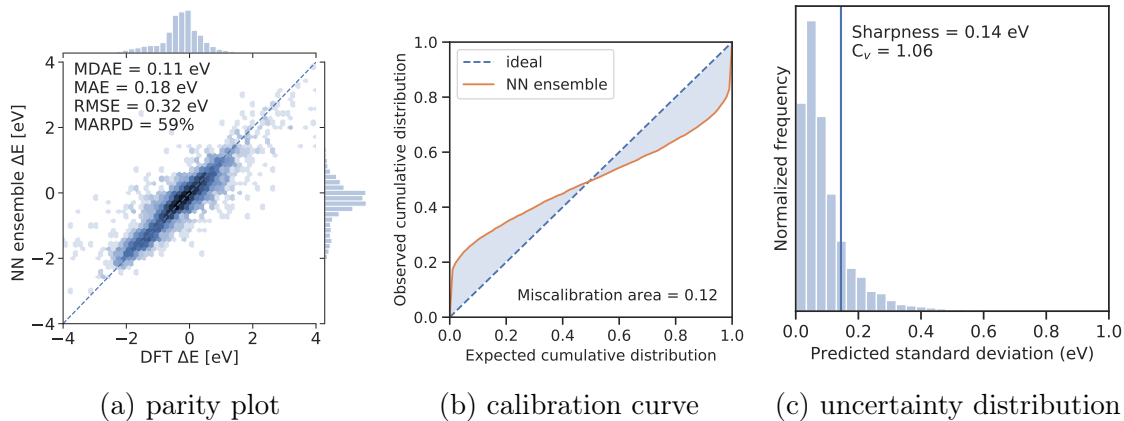


Figure 5-3: Results of the NN ensemble. Each figure here was created with the test set of 8,289 points.

Accurate models have parity plots whose points tend to fall near the diagonal parity line. Calibrated models have calibration curves that approach the ideal diagonal line. Sharp models have uncertainty distributions that tend towards zero. Note that sharpness should not be won at the cost of calibration.

The calibration curve was created by first establishing Gaussian random variables for each test point where the means were the model's predictions and the variances were the model's predicted variances. The test residuals could then be compared against their respective random variables. For simplification purposes, we divided each of the test residuals by their corresponding standard deviations so that we could test all residuals against the same unit Gaussian distribution. Thus if the normalized test residuals followed a unit Gaussian distribution, then the model's uncertainty predictions could be considered well-calibrated. We tested this by calculating the theoretical cumulative distribution of points within the intervals $(-\infty, x] \forall x \in (-\infty, \infty)$ and then compared it against the observed cumulative distributions. A plot of the observed cumulative distributions against the theoretical cumulative distributions is called a calibration curve. A perfectly calibrated model would have normalized residuals that are Gaussian, which would yield a diagonal calibration line. Therefore, models' calibration could be qualified by the closeness of their calibration curves to this ideal, diagonal curve. We quantified this closeness by calculating the area between the calibration curve and the ideal diagonal. We call this the miscalibration

area, and smaller values indicate better calibration. We also calculated the calibration error,[93] which is the mean square difference between the expected cumulative distributions and observed cumulative distributions.

The shape of a calibration curve could also yield other insights. If a model’s UQs were too low/confident, then the normalized residuals would be too large and they would fall outside their distributions too frequently. This would result in a lower observed cumulative distributions compared to the expected cumulative distributions, which would correspond to a calibration curve that falls below the ideal diagonal. Therefore, overconfident models yield calibration curves that fall under the ideal diagonal, and underconfident models yield calibration curves that fall over the ideal diagonal. Figure 5-4 illustrates this point by plotting calibration curves of various models alongside their parity plots that contain error bars corresponding to ± 2 standard deviations. Note that when we say a calibration curve “falls under the diagonal”, we allude to curves whose right-hand-side fall under the diagonal.

5.4.2 Summary results

Figure 5-5 contains parity plots for all UQ methods studied here; Figure 5-6 contains all calibration curves; and Figure 5-7 contains all distribution plots of the ML-predicted UQs. These figures illustrate the accuracy, calibration, and sharpness of the different UQ methods, respectively. Table 5.1 lists their performance metrics.

Regarding accuracy: All methods’ MDAE results are virtually identical, and their MAE results are within 10% of each other. This suggests that all methods have comparable predictive accuracies for inliers. The plain GP has a higher RMSE value than the rest of the methods, indicating that it has the worst predictive accuracy for outliers. Correlations between residuals and uncertainty estimates are discussed in the Supplementary Information section briefly.

Regarding calibration: The NN ensemble, BNN, and Dropout NN are overconfident; the GP is underconfident; and the NN Δ NN, GP_{NN- μ} , and CFGP models are relatively calibrated. The three more calibrated methods all share a characteristic that the other methods do not: They all start with a NN that is dedicated for prediction

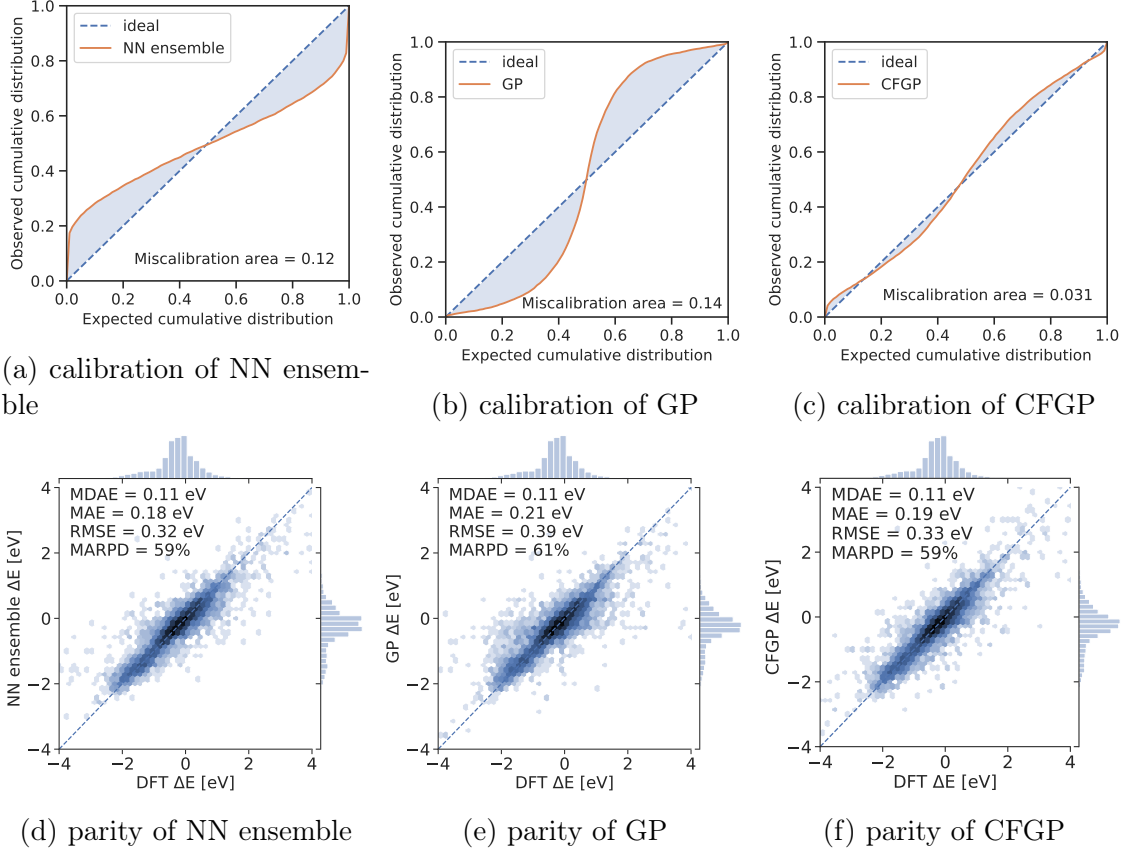


Figure 5-4: Calibration curves and parity plots of an overconfident NN ensemble, an underconfident GP, and better-calibrated CFGP. The vertical uncertainty bands in the parity plots indicate ± 2 standard deviations in the uncertainty predictions of each model. For clarity, we sampled only 20 points of the 8,289 test points to put in the parity plots. It follows that relatively overconfident models would have more points with uncertainty bands that do not cross the diagonal parity line; relatively underconfident models would have more points that cross the diagonal parity line; and a well-calibrated model would have *ca.* 19 out of 20 points cross the parity line.

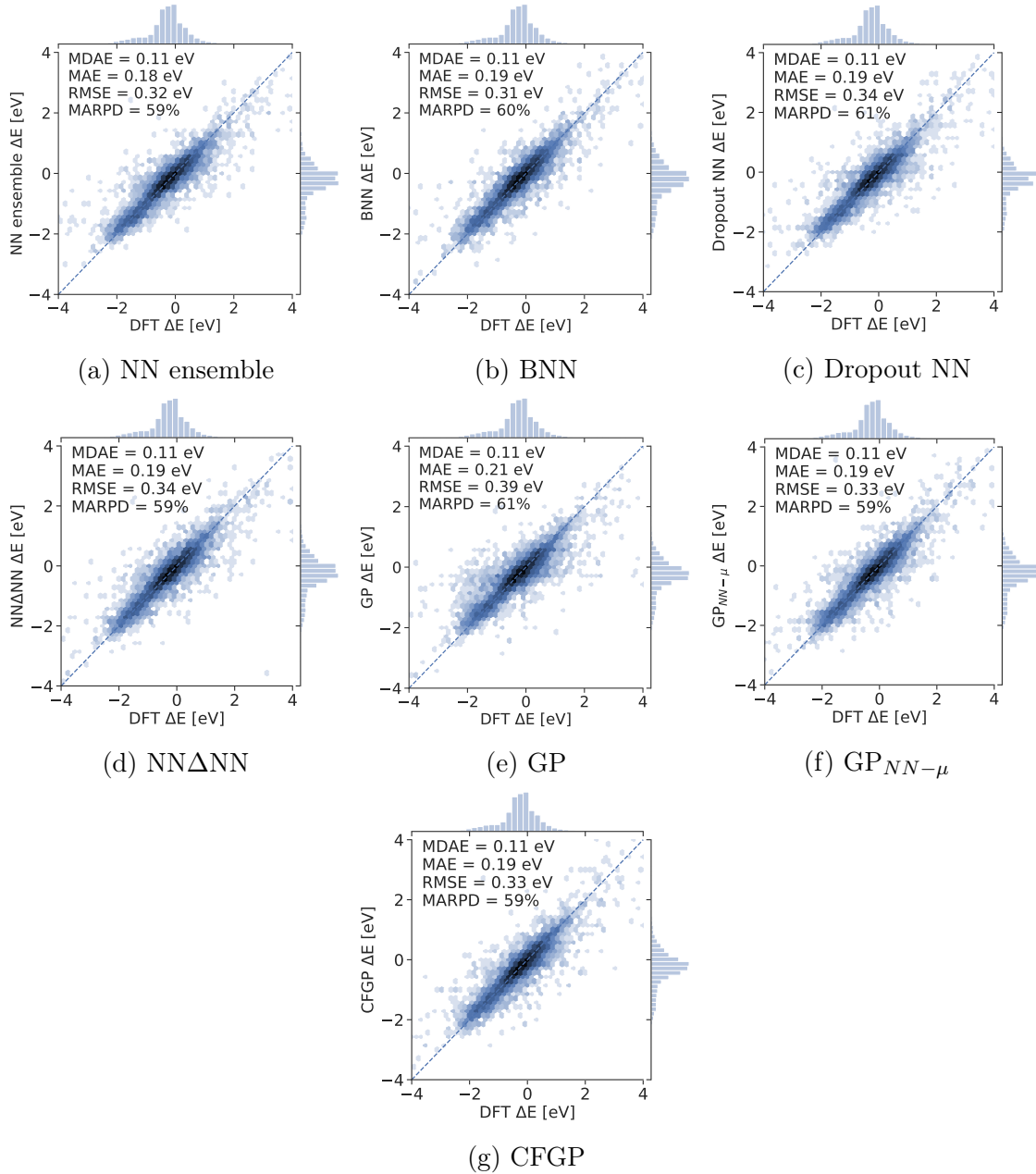


Figure 5-5: Parity plots for all UQ methods used in this study. Shading plots were used in lieu of scatter plots because the large number of test points (8,289) obfuscated patterns. Darker shading indicates a higher density of points. Logarithmically scaled shading was used to accentuate outliers. The dashed, diagonal lines indicate parity.

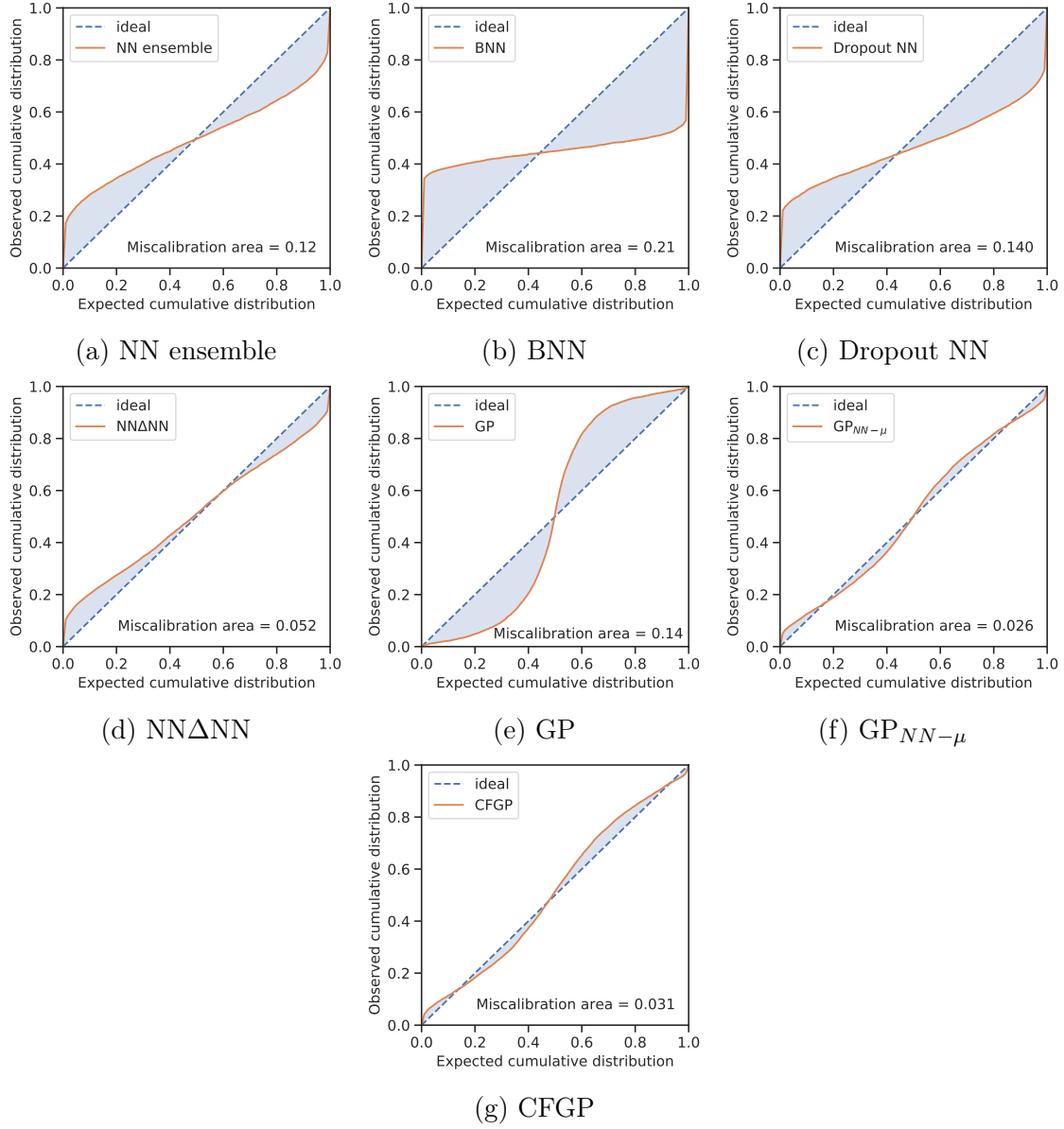


Figure 5-6: Calibration curves for all UQ methods used in this study. Dashed, blue lines indicate perfect calibration while solid orange lines indicate the experimental calibration of the test set. The blue, shaded area between these lines is defined as the miscalibration area.

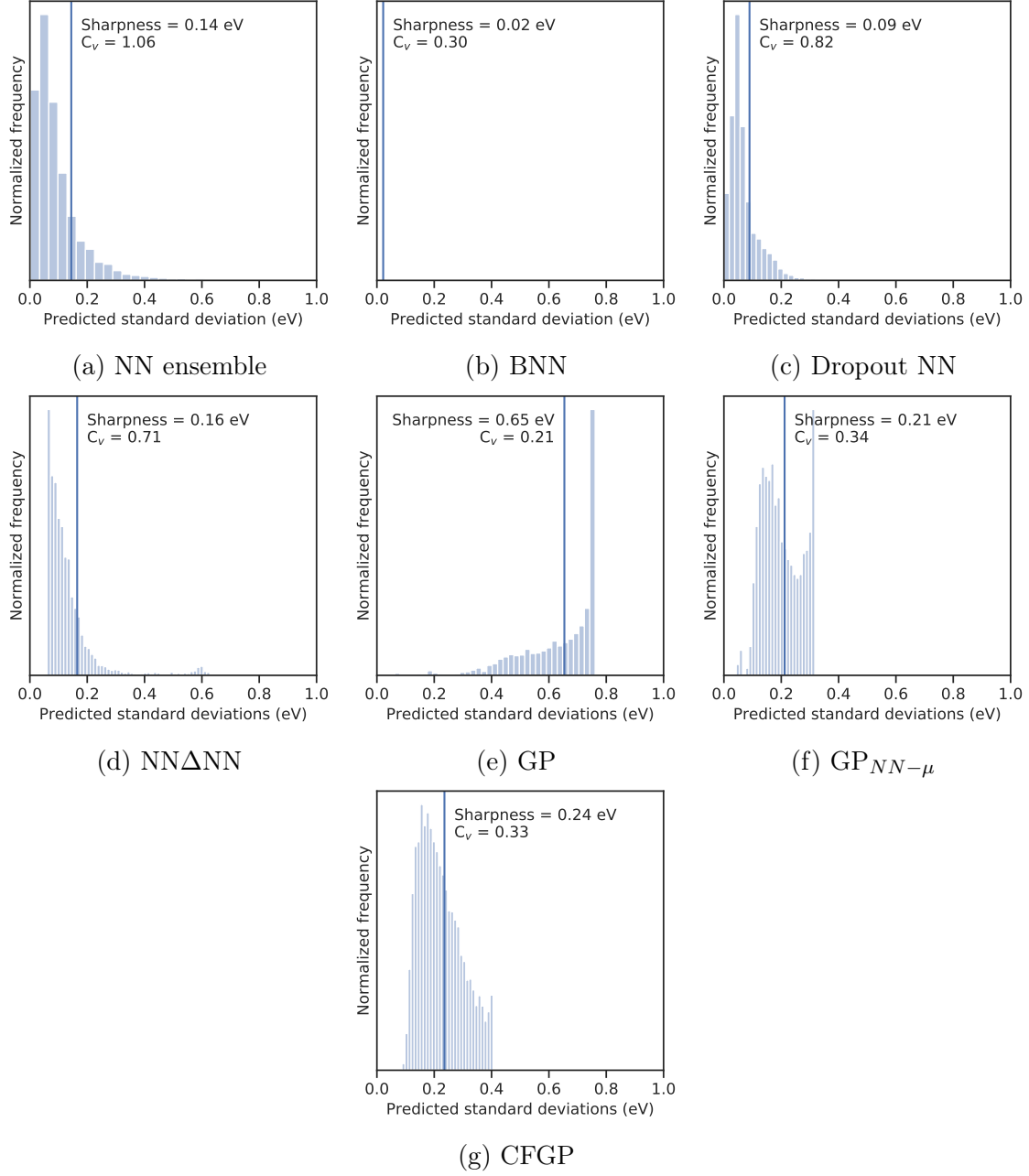


Figure 5-7: Distribution plots of the ML-predicted standard deviations for each method. Sharpness values are indicated by vertical lines.

Table 5.1: Performance metrics for all methods used in this study, which include: Median Absolute Error (MDAE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Relative Percent Difference (MARPD), R^2 correlation coefficient (R^2), miscalibration area (MisCal), calibration error (CalErr), sharpness (Sha), coefficient of variation (C_v), and negative log-likelihood (NLL). The units of MDAE, MAE, RMSE, and sharpness are all in eV. The units of MARPD are in %. The miscalibration area, calibration error, C_v , and NLL are unitless.

Method	MDAE	MAE	RMSE	MARPD	R^2	MisCal	CalErr	Sha	C_v	NLL $\cdot 10^3$
NN	0.11	0.19	0.34	61	0.80	N/A	N/A	N/A	N/A	N/A
NN ensemble	0.11	0.18	0.32	59	0.82	0.12	1.70	0.14	1.06	192.08
BNN	0.11	0.19	0.31	59	0.83	0.20	5.32	0.03	0.30	669.61
Dropout NN	0.11	0.19	0.34	61	0.79	0.14	2.52	0.09	0.82	$7.38\cdot 10^{14}$
NN Δ NN	0.11	0.19	0.34	59	0.80	0.05	0.39	0.16	0.71	18.61
GP	0.11	0.21	0.39	61	0.73	0.14	2.35	0.65	0.21	6.41
GP $_{NN-\mu}$	0.11	0.19	0.33	59	0.81	0.03	0.08	0.21	0.34	6.09
CFGP	0.11	0.19	0.33	59	0.80	0.03	0.13	0.24	0.33	2.80

alone, and then they end with some other in-series method to estimate uncertainty. Interestingly, this in-series method of learning predictions and then learning uncertainties is similar in spirit to how gradient boosted models “learn in stages” using an ensemble of models.

Regarding sharpness: The NN ensemble, BNN, and Dropout NN models yield the most sharp uncertainties, although they do so at the cost of calibration. Among the three more calibrated models, the NN Δ NN yields the lowest sharpness of 0.16 eV while the GP $_{NN-\mu}$ and CFGP yield sharpnesses of 0.21 and 0.24 eV, respectively. Note how GP-based UQ methods tend to yield less sharp uncertainties than methods based purely on NNs. This suggests that GPs may yield more conservative UQs.

Regarding NLL: The CFGP method yields the best (i.e., lowest) NLL value of *ca.* 2,800 while both the GP and GP $_{NN-\mu}$ models yield relatively moderate NLL values of *ca.* 6,000. Note how the under-confident GP model has a worse miscalibration area, calibration error, and sharpness than the NN Δ NN but a better NLL value. Simultaneously, the three most over-confident and sharp models (NN ensemble, BNN, and Dropout NN) yield the worst NLL results. This shows that better NLL values correlate with relatively conservative estimates of UQ, but not with relatively liberal estimates. In other words: If we use NLL as our main performance metric, then we

will favor under-confident UQ estimates in lieu of over-confident estimates.

Given the performance metrics for accuracy, calibration, sharpness, and NLL, we expect the CFGP or $\text{GP}_{NN-\mu}$ methods to yield the best performing UQ models for our dataset. When choosing UQ methods for different applications, other factors should be considered. For example: Although the $\text{GP}_{NN-\mu}$ method performed relatively well, it relied on hand-crafted features. If future researchers wish to use the $\text{GP}_{NN-\mu}$ method to predict other properties from atomic structures, they may have to define their own set of features. This process of feature engineering is non-trivial and varies from application to application. In some cases, it may be easier to use a UQ method that does not require any additional features beyond the NN input, such as $\text{NN}\Delta\text{NN}$ or CFGP. This is why we declare CFGP as the method of choice for our study here; it has a relatively competitive accuracy, calibration, and sharpness while requiring less information than $\text{GP}_{NN-\mu}$.

Another factor to consider is the overhead cost of implementation. For example: The NN ensemble method is arguably the simplest NN-based UQ method used here and may be the easiest method to implement. Conversely, NN ensembles also have a higher computational training cost than some of the other methods used here, such as $\text{NN}\Delta\text{NN}$ or CFGP. This high training cost is exacerbated if the ensemble is meant to be used in an active framework where the model needs to be trained continuously. As another example: The BNN method yielded perhaps the worst results of all the methods studied here. It could be argued that further optimization of the BNN could have resulted in higher performance. But creation and training of BNNs is still an active area of research with less literature and support than GPs or non-Bayesian NNs. This lack of support led to us spending nearly twice as long creating a BNN compared to the other methods. It follows that further optimization of the BNN would be non-trivial and may not be worth the overhead investment.

5.5 Conclusions

We examined a procedure for comparing different methods for uncertainty quantification (UQ). This procedure considers the accuracy of each method, the honesty of their uncertainty estimates (i.e., their calibration), and the size of their uncertainty estimates (i.e., their sharpness). To assess accuracy, we outlined a common set of error metrics such as MAE or RMSE, among others. To assess calibration, we showed how to create, interpret, and quantify calibration curves. To assess sharpness, we showed how to calculate and plot sharpness. To assess all three aspects simultaneously, we suggest using the negative log-likelihood (NLL) as a performance metric. The ensemble of all these metrics and figures can be used to judge the relative performance of various UQ methods in a holistic fashion.

As a case study, we tested six different methods for predicting Density Functional Theory (DFT) calculated adsorption energies with UQ. The best performing method was a Convolution-Fed Gaussian Process (CFGP), which used a pre-trained convolutional output from a NN as features for a subsequent GP that made probabilistic predictions. Our studies also showed that the GP-based methods we tested tended to yield higher and more conservative uncertainty estimates than the methods that used only NNs and NN derivatives. We also found that in-series methods tended to yield more calibrated models—i.e., methods that use one model to make value predictions and then a subsequent model to make uncertainty estimates were more calibrated than models that attempted to make value and uncertainty predictions simultaneously. These results are limited to our dataset. Results may vary for studies with different applications, different models, or different hyperparameters. But the underpinning procedure we used to compare these models is still broadly applicable.

Note that it would be possible to recalibrate[93] each of the models in this study to improve their uncertainty estimates. We purposefully omitted recalibration in this study to (1) simplify the illustration of the UQ assessment procedure; (2) assess the innate performance of each of these UQ methods without confounding with recalibration methods; and (3) reduce overhead investment. Future work should consider

recalibration if the feasible UQ methods provide insufficiently calibrated uncertainty predictions

Future work may also consider inductively biased UQs. For example: If we used the Bayesian Error Estimation Functional with van der Waals correlation (BEEF-vdW),[179] then our DFT calculated adsorption energies would have been distributions rather than single point estimates. Such distributions could be propagated to certain UQ surrogate models, e.g., as a variable-variance kernel in a GP-type method. As another example of inductively biased UQs: A model may be able to make low-uncertainty predictions on a DFT-optimized structure and then also make high-uncertainty predictions on a similar but DFT-unoptimized structure. UQs do not need to be derived strictly from data. They may also be derived from previous knowledge.

Chapter 6

Computational catalyst discovery: Active classification through myopic multiscale sampling

This work has been submitted recently to The Journal of Chemical Physics as: Kevin Tran, Willie Neiswanger, Kirby Broderick, Eric Xing, Jeffrey Schneider, and Zachary W. Ulissi. Computational materials discovery: Active classification through myopic multiscale sampling.

My contribution in this work involved inventing the multiscale modeling framework and co-inventing the multiscale sampling strategy with Willie, Jeff, and Zack. I also created and ran all the code used to perform the studies. I wrote the abstract, results, conclusions, and parts of the introduction & methods sections.

6.1 Abstract

The recent boom in computational chemistry has enabled several projects aimed at discovering useful materials or catalysts. We acknowledge and address two recurring issues in the field of computational catalyst discovery. First, calculating macro-scale catalyst properties is not straight-forward when using ensembles of atomic-scale calculations (e.g., density functional theory). We attempt to address this issue by creating

a multi-scale model that estimates bulk catalyst activity using adsorption energy predictions from both density functional theory and machine learning models. The second issue is that many catalyst discovery efforts seek to optimize catalyst properties, but optimization is an inherently exploitative objective that is in tension with the explorative nature of early-stage discovery projects. In other words: why invest so much time finding a “best” catalyst when it is likely to fail for some other, unforeseen problem? We address this issue by relaxing the catalyst discovery goal into a classification problem: “What is the set of catalysts that is worth testing experimentally?” Here we present a catalyst discovery method called myopic multiscale sampling, which combines multiscale modeling with automated selection of density functional theory calculations. It is an active classification strategy that seeks to classify catalysts as “worth investigating” or “not worth investigating” experimentally. Our results show a ~ 7 –16 times speedup in catalyst classification relative to random sampling. These results were based on offline simulations of our algorithm on two different datasets: a larger, synthesized dataset and a smaller, real dataset.

6.2 Introduction

Recent advances in computing hardware and software have led to substantial growth in the field of computational materials science. In particular, databases of high-throughput calculations[72, 33, 152, 136, 14, 124] have increased the amount of information available to researchers. These databases facilitate the development of models that supplement human understanding of physical trends in materials.[156, 155, 117] These models can then be used in experimental discovery efforts by identifying promising subsets of the search space, resulting in increased experimental efficiency.[30, 70, 25, 54, 171, 191]

However, many materials design efforts use material properties and calculation archetypes that are too problem-specific to be tabulated in generalized databases. When such efforts coincide with design spaces too large to search in a feasible amount of time, we need a way to search through the design space efficiently. Sequential

learning, sometimes referred to as optimal design of experiments or active learning, can fill this role. Sequential learning is the process of using the currently available data to decide which new data would be most valuable for achieving a particular goal.[107, 157, 81] In practice, this usually involves fitting a surrogate model to the available data and then pairing the model with an *acquisition function* that calculates the values of a new, potential data points. Then we *query* the most valuable data points, add them to the data set, and repeat this process. These sequential learning methods have been estimated to accelerate materials discovery efforts by up to a factor of 20.[149]

Sequential learning has numerous sub-types of methods that can and have been used for different goals. One such sub-type is active learning. With many active learning algorithms, the goal is to replace a relatively slow data-querying process with a faster-running surrogate model.[163] Since the surrogate model may be used to query any point, the acquisition functions focus on ensuring that the entire search space is explored. Another sub-type of sequential learning is active optimization.[45] With this sub-type, the goal is to maximize or minimize some objective function. Thus the acquisition functions generally focus on parts of the search space where maxima or minima are more likely to occur. One of the most common types of active optimization is Bayesian optimization.[45] Yet another sub-type of sequential learning is online or on-the-fly learning.[65] The goal for these methods is to accelerate the predictions of streams of data. In the field of computational material science, this is often applied to predicting trajectories for DFT or molecular dynamics calculations.[80, 174]

In computational materials discovery, we often have the following task: we have a set of available materials $\mathcal{X} = \{x_i\}_{i=1}^n$, where each material x_i has an associated quantity y_i , denoting its value for some application. Examples of common properties for y_i include—but are not limited to—formation energies of materials, catalyst activity, tensile strength, or conductivity. The value y_i is unknown and must be calculated, which can be costly in time, money, or other resources. Further, theoretical calculations of material properties may be inconsistent with experimental results. As per a common aphorism among statisticians: “All models are wrong, but some are

useful.”

Due to these potential model errors and due to the exploratory nature of materials discovery, we propose reframing the materials discovery question. Instead of trying to discover materials with optimal y_i values, what if we instead classify materials as having promising or unpromising y_i values? In other words, what if we frame materials discovery efforts as classification problems rather than optimization problems? Then these classes could then be used to design physical experiments. Mathematically, this is akin to assuming that material i has a binary value $y_i \in \{0, 1\}$, where 0 denotes “*not of interest*”, and 1 denotes “*of interest*”.

The goal is then to determine the values y_i for each $x_i \in \mathcal{X}$ as cheaply as possible. One can view this as the task of most-efficiently learning a classifier that, for each x_i , correctly predicts its value y_i . In this way, materials discovery problems can be framed as problems of *active classification*. Active classification is the task of choosing an ordering of $x_i \in \mathcal{X}$, over which we will iterate and sequentially measure their values y_i , in order to most efficiently (using the fewest measurements) learn a classifier that predicts the correct label for all materials $x_i \in \mathcal{X}$. [112, 188]

Another aspect of computational materials discovery is the ability to turn calculations into recommendations—e.g., how can we convert DFT results into actionable experiments? This conversion is relatively straight-forward when properties are directly calculable, which is the case for properties such as the enthalpy of formation.[44] If we perform a single DFT calculation that suggests a single material may be stable, then we can suggest that single material for experimentation. But for many applications, the properties of interest may not be calculable directly. For example, let us say we are interested in finding active catalysts. One way we can do that is by using DFT to calculate the adsorption energy between the catalyst and particular reaction intermediates, and then couple the resulting adsorption energy with a Sabatier relationship.[160] But *in situ*, a catalyst comprises numerous adsorption sites and surfaces. Thus the true activity of a catalyst may be governed by an ensemble of adsorption energies, and therefore may need multiple DFT calculations. How do we address the fact that we need multiple DFT queries to resolve the properties of a

single material?

Here we attempt to address both outlined issues: (1) we need an ensemble of DFT queries to calculate a single experimental property of interest, and (2) we need a sequential learning method designed for high-throughput discovery/classification. We overcome both issues by creating the MMS method (Figure 6-1). MMS addresses the first aforementioned issue by using a multiscale modeling framework for estimating the activity of a catalyst using an ensemble of both DFT and ML predicted adsorption energies. MMS then addresses the second issue by combining this multiscale modeling framework with a number of sequential learning methods, including active classification. Note that MMS, as we describe it in this paper, is tailored to discovering active catalysts. Although this method may not be directly transferable to other applications, we hope that others may be able to adapt the principles of the method to their own applications.

6.3 Methods

6.3.1 Multiscale Modeling

In this paper, we use the discovery of active catalysts as a case study. Catalyst activity is often correlated with the adsorption energy of particular reaction intermediates, as per the volcano relationships stemming from the Sabatier principle.[134, 160] These adsorption energies can be calculated using DFT. Each DFT-calculated adsorption energy is specific to a particular binding site of a particular surface of a particular catalyst. Thus the relationship between DFT-calculated adsorption energies and a catalyst’s activity is not simple.

For example: In cases of lower adsorbate coverage on the catalyst surface, adsorbates tend to adsorb to stronger-binding sites before weaker-binding sites. In cases of higher adsorbate coverage, adsorption energies are difficult to calculate, so it is not uncommon to assume low adsorbate coverage.[134, 133, 108, 6-26,97-113,126] It follows that the activity of a surface could be estimated by using the Sabatier-calculated

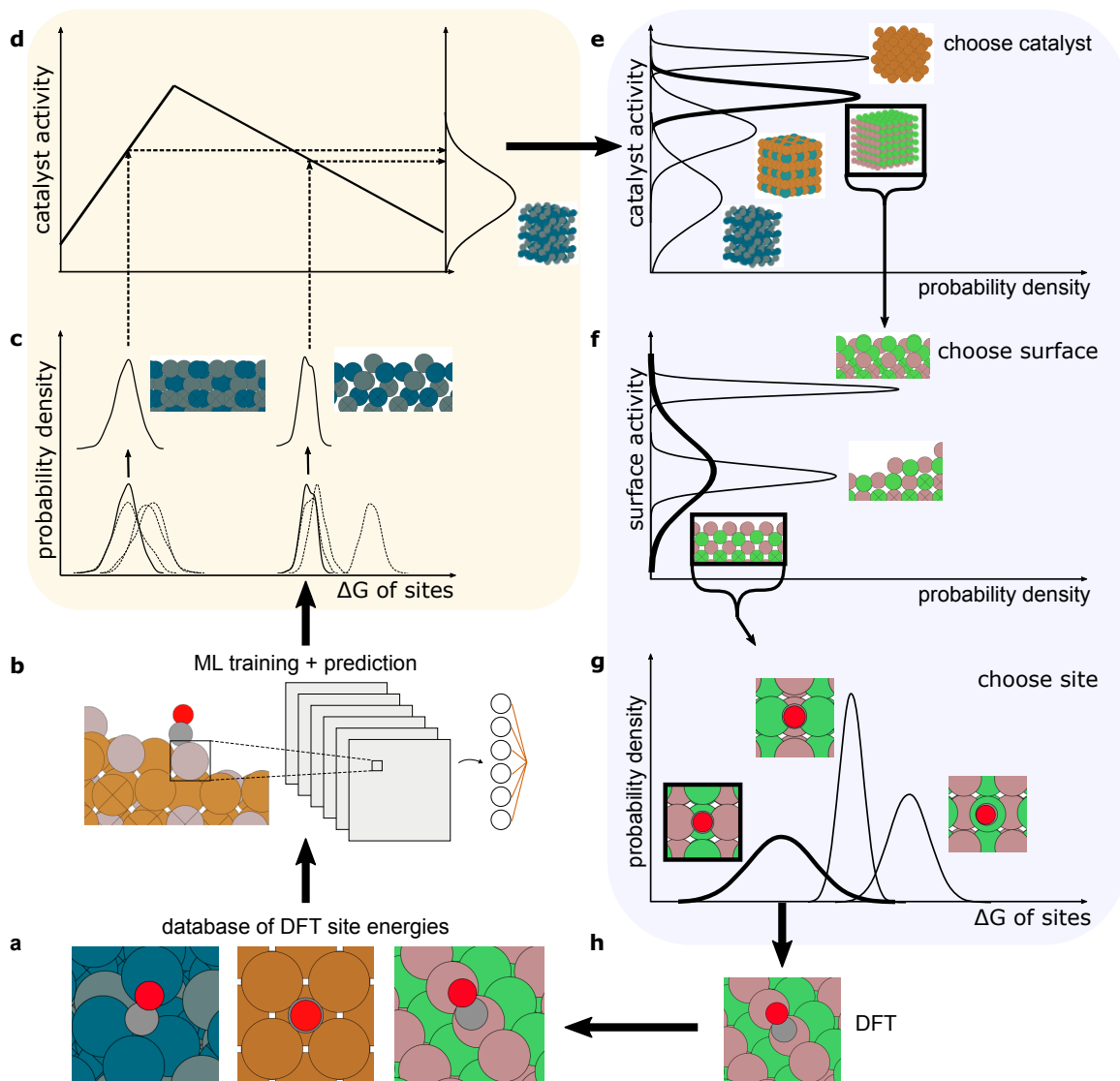


Figure 6-1: Illustration of Myopic Multiscale Sampling (MMS). Given a database of DFT-calculated adsorption energies (**a**), we train a ML model to predict adsorption energies (**b**). Then we use those adsorption energies to estimate activities of catalyst surfaces (**c**), which we then use to estimate the activities of the bulk catalysts (**d**). Then we choose which catalyst to sample next (**e**); then we choose which surface on the catalyst to sample (**f**); then we choose which site on the surface to sample (**g**); then we perform DFT of that site to add to the database (**h**). This procedure is repeated continuously with the goal of classifying all catalysts as either “relatively active” or “relatively inactive”.

activity of the strongest binding site on a surface.

Given the activities of the surfaces of a catalyst, the next step is to estimate the activity of the entire catalyst. One way to do this would be to perform a weighted average of the surface activities, where higher weights are given to surfaces that are more stable. For simplicity's sake, we instead propose a uniform average and recognize that future work may involve investigating more sophisticated averaging methods.

Concretely, suppose we have n catalyst candidates $\{x_i\}_{i=1}^n$, where each candidate x_i has m surfaces $\{u_{i,j}\}_{j=1}^m$, and surface $u_{i,j}$ has ℓ sites $\{s_{i,j,k}\}_{k=1}^\ell$. For a given site $s_{i,j,k}$, denote its adsorption energy by $\Delta G(s_{i,j,k})$, and for a given surface $u_{i,j}$, denote its catalytic activity by $\alpha(u_{i,j})$. Likewise, for a given catalyst material candidate x_i , denote the average catalytic activity for the candidate by $\alpha(x_i) = \frac{1}{m} \sum_{j=1}^m \alpha(u_{i,j})$. Suppose we have a predictive uncertainty estimate for the adsorption energy $\Delta G(s_{i,j,k})$ of a site, represented by a Normal distribution with mean $\mu_{i,j,k}$ and variance $\sigma_{i,j,k}^2$. We can then perform simulation-based uncertainty quantification of catalyst activity by using the multiscale modeling process we described above to propagate uncertainties from sites' adsorption energies. Specifically, for each material candidate x_i , we generate H samples of its catalytic activity, $\{\tilde{\alpha}_i^h\}_{h=1}^H$, by simulating from the following generative process:

$$\text{For } j = 1, \dots, m, k = 1, \dots, \ell : \quad (6.1)$$

$$\{\tilde{\Delta}G_{i,j,k}^h\}_{h=1}^H \stackrel{iid}{\sim} \mathcal{N}(\mu_{i,j,k}, \sigma_{i,j,k}^2)$$

$$\text{For } h = 1, \dots, H, j = 1, \dots, m :$$

$$\tilde{\alpha}_{i,j}^h = \begin{cases} \exp(M_1 \tilde{\Delta}G_{i,j,1:\ell}^h + B_1) & \text{if } \tilde{\Delta}G_{i,j,1:\ell}^h \geq t^* \\ \exp(M_2 \tilde{\Delta}G_{i,j,1:\ell}^h + B_2) & \text{otherwise} \end{cases}$$

$$\text{For } h = 1, \dots, H :$$

$$\tilde{\alpha}_i^h = \frac{1}{n} \sum_{j=1}^m \tilde{\alpha}_{i,j}^h$$

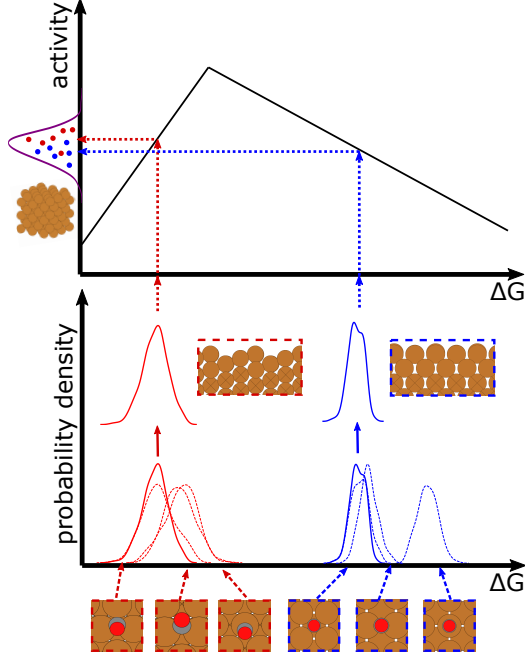


Figure 6-2: Multiscale modeling strategy for estimating the activity of a catalyst. For each adsorption site, we obtain a machine-learned estimate of its adsorption energy along with uncertainty. Then we aggregate the energy distributions for all sites within each surface through a minimum operator across sites. Next we transform the energy distributions for all surfaces into activities using a Sabatier relationship. Finally we average all the surface activities to obtain an estimate of overall catalyst activity.

where t^* is the optimal absorption energy for a given volcano relationship and M_1 , M_2 , B_1 , & B_2 are the linear coefficients associated with the two sides of the log-scaled volcano relationship of a given chemistry. Figure 6-2 illustrates how we use our multiscale modeling method to estimate catalyst activity from DFT-calculated adsorption energies, including uncertainty quantification.

Each catalyst material candidate $x \in \mathcal{X}$ has some true catalytic activity level $\alpha(x)$. Our goal will be to determine the top p -% of catalyst material candidates in terms of their activity levels, which we denote $\mathcal{X}_p = \{x \in \mathcal{X} : r(\alpha(x)) \geq \lfloor \frac{pn}{100} \rfloor\}$, where $r : \mathbb{R}_+ \rightarrow \{1, \dots, n\}$ is a function mapping the activity level $\alpha(x)$ to an index denoting it's rank (from highest to lowest activity). Given a specified p , if a candidate material is in this set, i.e. $x_i \in \mathcal{X}_p$, then we say that its associated binary value $y_i = 1$, and say $y_i = 0$ otherwise. In simpler terms: We want to find the top p -% most active catalysts. For this paper, we choose $p = 10\%$ arbitrarily. Any catalyst that falls

within the top 10% in terms of activity will be labeled as active, and anything below the top 10% will be labeled as inactive.

We can therefore frame our goal as determining the associated binary value y_i for each catalyst material candidate $x_i \in \mathcal{X} = \{x_i\}_{i=1}^n$. Suppose we have formed point estimates for each of the binary values, written $\{\hat{y}_i\}_{i=1}^n$. To assess the quality of this set of estimates with respect to the set of true candidate values, we focus on the F_1 score—a popular metric for classification accuracy, defined as

$$\begin{aligned} F_1 &= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \\ &= \frac{2 \sum_{i=1}^n y_i \hat{y}_i}{2 \sum_{i=1}^n y_i \hat{y}_i + \sum_{i=1}^n (1 - y_i) \hat{y}_i + \sum_{i=1}^n y_i (1 - \hat{y}_i)}. \end{aligned} \tag{6.2}$$

Given a set of ground-truth values $\{y_i\}_{i=1}^n$, we are able to compute the F_1 score for a chosen set of value estimates $\{\hat{y}_i\}_{i=1}^n$.

However, in practice, we will typically not have access to these ground-truth values, and thus cannot compute this score in an online procedure. For use in online experiments, we will take advantage of a metric that yields an estimate of the change in F_1 score. This metric is computable using only our model of the activity of each catalyst, without requiring access to ground-truth values $\{y_i\}_{i=1}^n$, and can be used to assess and compare the convergence of our methods. Furthermore, it can be used to provide an early stopping method for our active procedures. We will show experimentally in Section 6.4 that this metric shows a strong correlation to the F_1 score.

6.3.2 Sampling Strategy

The goal of MMS is to discover catalysts that are likely to be experimentally active. Optimization of catalytic activity is not the main priority, because we assume that unforeseen experimental issues are likely to obsolete most candidate catalysts. Instead, a greater focus is given on identification of a large number of candidates rather than finding “the most active” candidate. That is why the core sequential learning

algorithm we use in MMS is active classification.[112, 188] To be specific, we use Level Set Estimation (LSE) to identify catalysts for DFT sampling. After identifying catalysts for DFT sampling, we then need to choose which surface of the catalyst to sample; here we use active learning. Once a surface is chosen, we then attempt to find the strongest binding site on that surface by using active optimization of the adsorption energies. Thus we combine three different sequential learning strategies across three different length scales to decide which site-based DFT calculation will help us classify active vs. inactive catalysts (Figure 6-3).

We first describe the initial step of our sampling strategy, which consists of selecting a catalyst material candidate from our candidate set $\mathcal{X} = \{x_i\}_{i=1}^n$. Note that our high-level goal is binary classification, in that we want to efficiently produce accurate estimates $\{\hat{y}_i\}_{i=1}^n$ of the binary value for each material candidate. Based on our definition of $y_i = \mathbb{1}[x_i \in \mathcal{X}_p]$, this problem can be equivalently viewed as the task of LSE, in which we aim to efficiently produce an accurate estimate of the superlevel set $\mathcal{X}_p = \{x \in \mathcal{X} : r(\alpha(x)) \geq \lfloor \frac{pn}{100} \rfloor\}$. There has been a body of work on developing acquisition functions for choosing candidates to query in the task of LSE.[52, 79] In particular, we focus on the *probability of incorrect classification* acquisition function,[21] defined for an $x_i \in \mathcal{X}$ as

$$\varphi(x_i) = \min(p, 1 - p), \text{ where} \tag{6.3}$$

$$\begin{aligned} p &= \Pr \left(r(\alpha(x)) \geq \left\lfloor \frac{pn}{100} \right\rfloor \right) \\ &\approx \underbrace{\frac{1}{H} \sum_{h=1}^H \mathbb{1} \left[r(\tilde{\alpha}_i^h) \geq \left\lfloor \frac{pn}{100} \right\rfloor \right]}_{\text{Empirical probability } \alpha(x) \text{ in top } p\%} \end{aligned}$$

Thus to select a subsequent catalyst candidate, we compute $\varphi(x_i)$ for each $x_i \in \mathcal{X}$ and return the maximizer $x^* = \arg \max_{x_i \in \mathcal{X}} \varphi(x_i)$. In simpler terms: We choose the catalyst that we are most likely to classify incorrectly. Note how this implies that we *not* query catalysts that we are confident are active, which is different from active optimization methods. This provides a more exploratory method rather than

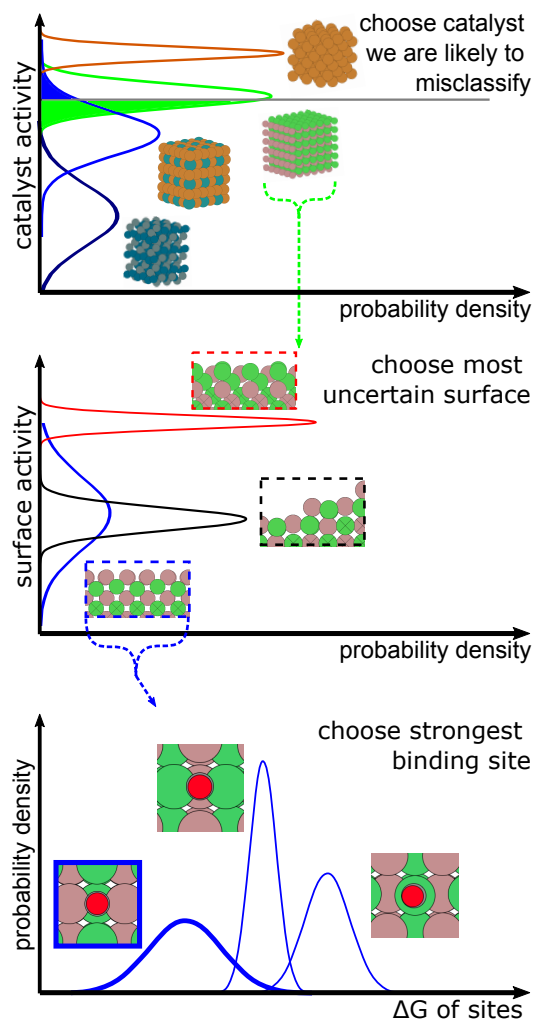


Figure 6-3: Myopic Multiscale Sampling (MMS) overview. At the highest level, we choose a catalyst to query using level-set estimation—to be specific, we use the probability of incorrect classification as our acquisition function. At the medial level, we choose a surface of the catalyst using uncertainty sampling. At the lowest level, we choose a site on the surface using Bayesian optimization to find the lowest energy site.

an exploitative one, which is appropriate in early-stage computational discoveries and screenings.

The selection of a catalyst candidate x_i depends on its estimated catalytic activity, which we model as an average of the catalytic activities across the surfaces of the candidate, i.e. $\alpha(x_i) = \frac{1}{m} \sum_{j=1}^m \alpha(u_{i,j})$. Though we select a candidate based on its ability to help improve our estimate of the superlevel set $\mathcal{X}_p$, once selected, we then wish to most efficiently improve our estimate of this candidate’s catalytic activity. Our goal at this stage is therefore to most efficiently learn the catalytic activities for each surface of that candidate. This can be viewed as an active regression task, where we aim to sample a surface that will most reduce the uncertainty of our surface activity estimates. To select a surface, we use an *uncertainty sampling for regression* acquisition function from the active learning literature[162], defined as

$$\begin{aligned} \varphi(u_{i,j}) &= \text{Var} \left[\text{Pr} \left(\alpha(u_{i,j}) \right) \right] \\ &\approx \frac{1}{H-1} \sum_{h=1}^H \left(\tilde{\alpha}_{i,j}^h - \frac{1}{H} \sum_{h'=1}^H \tilde{\alpha}_{i,j}^{h'} \right)^2, \end{aligned} \tag{6.4}$$

which selects a surface u_i^* of material candidate x_i that has the greatest variance. In simpler terms: We choose the surface of a catalyst that has the most uncertainty, because we suspect that this choice is most likely to reduce our uncertainty estimate of catalyst activity.

The catalytic activity of a given surface $\alpha(u_{i,j})$ is function of the adsorption energies of the sites on this surface, according to the relationship $\alpha(u_{i,j}) = \exp(-|M\tilde{\Delta}G_{i,j,1:\ell} + B|)$ from Equation (6.1), where $\tilde{\Delta}G_{i,j,1:\ell}$ is the set of adsorption energies over all sites on the surface. Therefore, given a selected surface $u_{i,j}$, we wish to determine efficiently the site on this surface with minimum adsorption energy. This can be viewed as an optimization task. We therefore use the *expected improvement* acquisition function from Bayesian optimization[126], defined as

$$\begin{aligned}
\varphi(s_{i,j,k}) &= \mathbb{E} \left[(\Delta G(s_{i,j,k}) \leq \Delta G^*) \mathbb{1} [\Delta G(s_{i,j,k}) - \Delta G^*] \right] \\
&\approx \Phi \left(\frac{\Delta G^* - \tilde{\mu}_{i,j,k}}{\tilde{\sigma}_{i,j,k}} \right) \phi \left(\frac{\Delta G^* - \tilde{\mu}_{i,j,k}}{\tilde{\sigma}_{i,j,k}} \right) \\
&\quad \times (\Delta G^* - \tilde{\mu}_{i,j,k}),
\end{aligned} \tag{6.5}$$

where $\tilde{\mu} = \frac{1}{H} \sum_{h=1}^H \Delta G_{i,j,k}^h$ is the expected adsorption energy, $\tilde{\sigma} = \sqrt{\frac{1}{H-1} \sum_{h=1}^H (\Delta G_{i,j,k}^h - \tilde{\mu})^2}$ is its standard deviation, Φ is the cumulative density function (CDF) of a standard normal distribution, ϕ is the PDF of a standard normal distribution, and ΔG^* is the minimum observed adsorption energy. This selects a site $s_{i,j}^*$ which is expected to most reduce the site adsorption energy relative to the current minimum observed energy, and allows for efficient estimation of the minimum energy site on surface $u_{i,j}$. In simpler terms: We choose the site on a surface that is most likely to help us identify the strongest/lowest binding site on the surface.

6.3.3 Active Learning Stopping Criteria

Assessing convergence of an active algorithm is useful for enabling early stopping, which can save resources. Measures of convergence can also provide diagnostics in online use settings. To quantify convergence, we use the $\Delta \hat{F}$ [10]. Intuitively speaking, this rule says to stop an active learning procedure when $\Delta \hat{F}$ drops below a predefined threshold ϵ when for k consecutive windows, i.e.,

Stop if $\Delta \hat{F} < \epsilon$ over k windows

Continue otherwise.

In our setting, $\Delta \hat{F}$ is defined to be

$$\Delta \hat{F} = 1 - \frac{2a}{2a + b + c}, \tag{6.6}$$

where a is the number of bulks for which the model at iterations i and $i + 1$ both yield a positive label, b is the number of bulks for which the model at iteration i yields a positive label while at iteration $i + 1$ yields a negative label, and c is the number of bulks for which the model at iteration i yields a negative label while at iteration $i + 1$ yields a positive label. Each of a , b , and c are computed over the previous k iterations. This measure provides an estimate of the change in accuracy at each iteration, and it allows us to control how conservatively (or aggressively) we stop early via an interpretable parameter ϵ . We show results of this measure alongside our $F1$ score in Section 6.4. Note that Altschuler & Bloodgood[10] recommend using a *stop set* of unlabeled points over which to calculate $\Delta\hat{F}$. Here we use the entire search space of catalysts in lieu of a stop set, because it was non-trivial for us to define a stop set that was representative of the search space.

6.3.4 Management of Data Queries

Implementation of MMS also involves definition of several hyper-parameters. For example, most surrogate models require training data before making predictions to feed the sampling method. This means that we needed to seed MMS with initial training data. We chose to create the initial training data by randomly sampling 1,000 adsorption energies from the search space. We used random sampling for simplicity, and we sampled 1,000 adsorption energies because that was the minimum amount of data on which CFGP could train on and maintain numerical stability.

Another consideration for MMS is the *batch size* and how to handle queries in tandem. Normal sequential learning assumes that we can make one query at a time. But in applications such as ours, it may be possible to make multiple queries in parallel—i.e., we can perform multiple DFT calculations at a time. There are several methods for handling queries in parallel; we chose to use a type of look-ahead sampling.[40] With look-ahead sampling, we began by choosing the first point to sample using the standard acquisition strategy. Then, while that point was still “being queried”, we assumed that the first point was queried successfully and set the “observed” value equal to our predicted value. In other words, we pretend that we

sampled the first data point and that our prediction of it was perfect. This allowed us to then recalculate our acquisition values to choose a second point. This process of “looking ahead” one point at a time was continued until a predetermined number of points were selected for querying—i.e., the batch size. Here we chose a batch size of 200 points, because that was roughly the number of DFT calculations that we could perform in a day during our previous high-throughput DFT studies.[171] Note that we did not re-train the surrogate models within each batch of 200 points; we only re-calculated acquisition values between each sample within each batch. We skipped re-training of surrogate models within each batch to reduce the amount of model training time required to perform this study. Although this may have reduced the effectiveness of the look-ahead method, we found the increased algorithm speed to be worthwhile.

6.3.5 Estimating Performance through Simulation

We aim to experimentally assess the performance of MMS and compare it with a variety of baseline methods without incurring the high cost of repeated DFT calculations. To do this, we simulate each procedure using a database of pre-determined adsorption energies. Specifically, suppose we have chosen a set of n catalyst material candidates $\{x_i\}_{i=1}^n$ of interest. For each candidate x_i , we already have all the adsorption energies $\Delta G(s_{i,j,k})$ for the full set of sites across the full set of surfaces on x_i . We can then run our procedures in a relatively fast manner, where we can quickly query the database at each iteration of a given method rather than running DFT. Similar offline-data discovery procedures have been pursued by previous work in optimization and active learning, where expensive evaluations have been collected offline and used for rapid online evaluation[26, 187, 180].

One notable baseline method is *random search*, which at each iteration samples sites to carry out DFT calculations uniformly at random from the full set of sites over all catalyst material candidates. We provide simulation results using random search as a benchmark to compare MMS against.

Surrogate Models Used

Our objective in this paper is to assess the performance of MMS. The performance of MMS is likely to depend on the surrogate model used to predict adsorption energies from atomic structures. We assume that surrogate models with high predictive accuracy and calibrated uncertainty estimates[93] will outperform models with low accuracy and uncalibrated uncertainty estimates, but we are unsure of the magnitude of this difference. We therefore propose to pair at least two different models with MMS: a “perfect” model and an “ignorant” model.

We define the “perfect” model, hereby referred to as the “prime” model, as a model that returns the true adsorption energy of whatever data point is queried. This perfect prediction ensures a high model accuracy. When asked for a standard deviation in the prediction, the prime model will return a sample from a χ^2 distribution whose mean is 0.1 eV. This uncertainty ensures a sharp and calibrated[93, 170] measure of uncertainty. We do not use standard deviation of zero because (1) it causes numerical issues during multiscale modeling and (2) any model in practice should not be returning standard deviations of zero.

We define the “ignorant” model, hereby referred to as the “null” model, as a model that returns the optimal adsorption energy no matter what is queried. This constant prediction ensures a relatively low model accuracy. When asked for a standard deviation in the prediction, the null model will return 1 eV. This uncertainty ensures a relatively dull and uncalibrated measure of uncertainty.

Lastly, we also choose to use a third, most practical model: CFGP.[170] CFGP is a Gaussian process regressor whose features are the output of the final convolutional layer in a trained graph convolutional neural network. This model is our best current estimate of both an accurate and calibrated model that could be used in practice. Thus we have three models: null, CFGP, and prime, which are intended to give quantitative estimates of the minimal, medial, and maximal performance of MMS, respectively.

Search Spaces Used

Previous studies have shown that different materials discovery problems have varying difficulties.[81] Searching for a needle in a hay stack is generally more difficult than searching for a leaf on a branch. Thus any simulation we do depends on the search space we use. To obtain a range of potential MMS performances, we perform simulations using two different data sets. Both data sets comprise thousands of atomic structures that represent CO adsorbing onto various catalyst surfaces, as well as corresponding adsorption energies. We then use Sabatier relationships from literature to transform the adsorption energies into estimates of activity.[105]

We defined our first search space by synthesizing it randomly. We did so by retrieving a database of enumerated adsorption sites from GASpy[171, 169]. These sites composed all the unique sites on all surfaces with Miller indices between -2 and 2 across over 10,000 different bulk crystal structures. We then randomly selected 200 of the bulk crystals along with all of the resulting surfaces and sites, yielding over 390,000 adsorption sites. Then for each bulk crystal, we randomly sampled its “bulk mean adsorption energy” from a unit normal distribution. Then for each surface within each crystal, we randomly sampled its “surface mean adsorption energy” from a normal distribution whose mean was centered at the corresponding bulk mean and whose standard deviation was set to 0.3 eV. Then for each site within each surface, we randomly sampled its adsorption energy from a normal distribution whose mean was centered at the corresponding surface mean and whose standard deviation was set to 0.1 eV. Thus the adsorption energies were correlated within each bulk, and they were also correlated within each surface.

We defined our second search space by retrieving our database of *ca.* 19,000 DFT-calculated CO adsorption energies calculated by GASpy, hereafter referred to as the GASpy dataset. The sites in this database were chosen using previous iterations of our sequential learning methods,[171] and they therefore have bias in the locations at which they were sampled. Specifically, the sites in this database were chosen based on the likelihood that their adsorption energies were close to the optimal value of

-0.67 eV.[105, 171]

There are several advantages of using the synthesized data set over the real GASpy data set, and vice versa. The synthesized data set contains pseudo-random adsorption energies that are difficult for CFGP to predict, thereby hindering its performance unfairly. Therefore, we should not and did not use CFGP with the synthesized data set; we used it with the GASpy data set only. On the other hand, the number of surfaces per bulk and the number of sites per surface in the GASpy data set was relatively sparse compared to the synthesized data set. This can result in catalysts that require relatively few site queries to sample fully, which reduces the number of queries necessary to classify a catalyst. This reduction in the number of required queries per catalyst could artificially improve the observed performance of MMS.

6.4 Results

At the beginning of the simulations, the multiscale models made their catalyst class predictions (i.e., active or inactive) using the adsorption energy predictions and uncertainties of the models. As the simulations progressed and adsorption energies were queried, the models’ predictions of each queried energy were replaced with the “true” value of the query and the corresponding uncertainty was collapsed to 0 eV. This was done to mimic a realistic use case where we would not use model predictions when we had the “real” DFT data instead. It follows that, as the simulations progressed and nearly all points were queried, most models performed similarly because they all had comparable amounts of “true” data to use in the multiscale model.

6.4.1 Performance on Synthesized Data

This behavior is seen in Figure 6-4a, which shows how the $F1$ changes at each point in the simulation of the synthesized data set. Here we see that the simulations using the prime model began with an $F1$ score of *ca.* 0.6 that increased to 1 over time. On the other hand, simulations using the null model began with an $F1$ score closer to 0 or 0.2 before gradually increasing to 1. This shows that more accurate surrogate

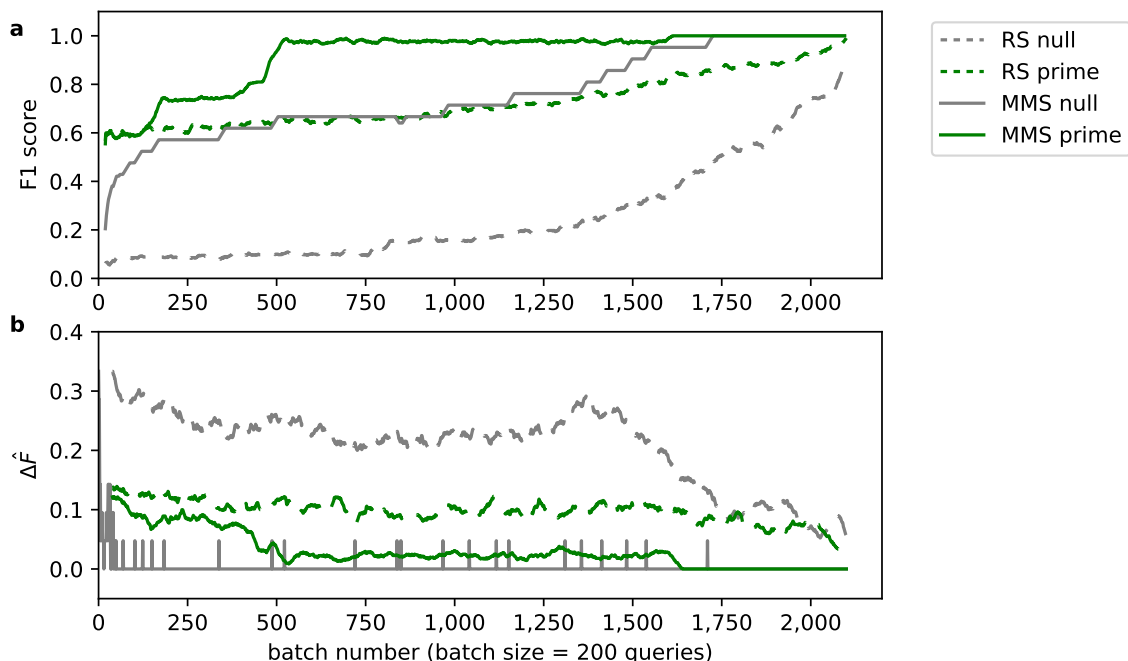


Figure 6-4: Performance and convergence results for the simulations on the synthesized dataset. **a.** $F1$ score of the multiscale model during simulation of the synthesized data. For clarity of visualization, we plotted the rolling average of the $F1$ score using a window of 20 batches. **b.** $\Delta\hat{F}$ of the multiscale model during simulation of the synthesized data. For clarity of visualization, we plotted the rolling average of $\Delta\hat{F}$ using a window of 40 batches (excluding the MMS null line, where no averaging was done). RS represents “random search” while MMS represents Myopic Multiscale Sampling.

models for adsorption energies led to more accurate multiscale models, even initially. Note also that the rate at which the $F1$ score improved was better when using MMS than when using random sampling, especially when using the null model. These data may suggest that the rate of improvement is governed by the acquisition strategy while the initial performance is governed by the model.

Figure 6-4b shows how the $\Delta\hat{F}$ changes at each point in the simulation of the synthesized data set. The simulations using random search generally yielded higher $\Delta\hat{F}$ values. This indicates slower convergence, which is consistent with the slower $F1$ increase seen in the random search curves Figure 6-4a. Note also how the $\Delta\hat{F}$ values for the MMS-prime simulation decreased at around 500 batches, which is the number of batches it took the $F1$ score to reach ca. 1. Lastly, we note that the $\Delta\hat{F}$

values for the MMS-null simulation were often zero. This is because the null model was a “stiff” learner that did not result in any multiscale modeling changes unless a low-coverage adsorption site was found. This shows that slow-learning models may result in relatively low $\Delta\hat{F}$ values, which may necessitate higher κ values to offset this behavior. In other words: Worse models may need longer horizons before stopping the discovery to mitigate the chances of missing important information.

These simulations provided us with an estimate of the improvement in active classification that we may get from using MMS. With the synthesized data set, we saw that the MMS-with-null case achieved an $F1$ score of ~ 0.6 after *ca.* 250 batches (or 50,000 queries). This was over seven times faster than the random-sample-with-null case, which achieved an $F1$ score of ~ 0.6 after *ca.* 1,800 batches (or 360,000 queries). When using the prime model, MMS was able to achieve an $F1$ score of ~ 0.75 in 200 batches, while the random search achieved this same performance in *ca.* 1,200 batches, or six times slower.

6.4.2 Performance on DFT Data

Figure 6-5 shows the $F1$ score and the $\Delta\hat{F}$ of the multiscale model at each point in the simulation of the GASpy data set. Interestingly, the system performance when using CFGP was similar to the performance when using the null model, both of which were overshadowed by the relatively good performance when using the prime model. This suggests that there is a large room for improvement for the CFGP model. Note also how the MMS strategy outperforms random sampling for this data set as well.

These simulations provided us with a second estimate of the improvement in active classification that we may get from using MMS. With the GASpy data set, we saw that the MMS-with-null case achieved an $F1$ score of ~ 0.8 after *ca.* 6 batches (or 1,200 queries). This was over sixteen times faster than the random-sample-with-null case, which achieved an $F1$ score of ~ 0.6 after *ca.* 80 batches (or 16,000 queries). When using the prime model, both MMS and random search were able to achieve an $F1$ score of ~ 0.8 after only a single batch.

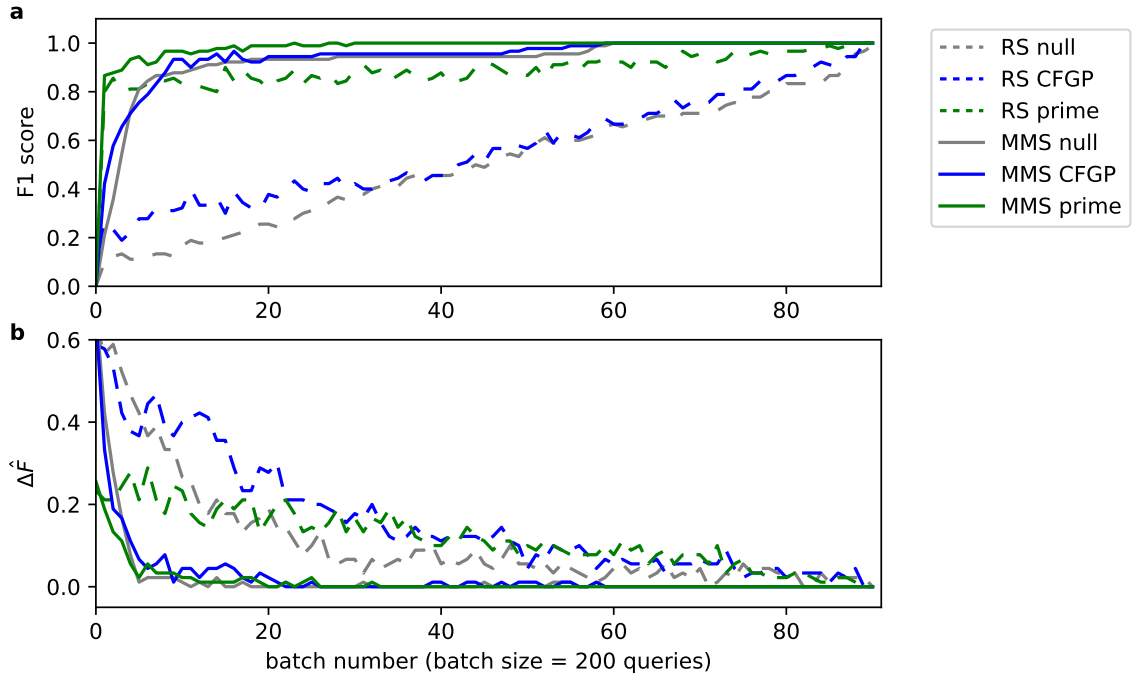


Figure 6-5: Performance and convergence results for the simulations on the GASpy dataset. **a.** $F1$ score of the multiscale model during simulation of the GASpy dataset. **b.** $\Delta \hat{F}$ of the multiscale model during simulation of the synthesized data. RS represents “random search” while MMS represents Myopic Multiscale Sampling.

6.4.3 Recommended diagnostics

We note that the $F1$ scores illustrated in Figures 6-4a and 6-5a cannot be calculated without knowing all the true classes, which is not possible to know during a real discovery process. We need metrics to monitor the behavior of both our discovery algorithm. We recommend monitoring the $\Delta\hat{F}$ as well as the accuracy, calibration, and sharpness (i.e., the magnitude of the predicted uncertainties) of the surrogate model over time. Figure 6-6 shows an example of such diagnostic metrics over the course our simulation that used MMS and CFGP on the GASpy dataset.

$\Delta\hat{F}$ estimates the amount of overall improvement in the discovery process. Sustained low values of $\Delta\hat{F}$ are a necessary but not sufficient indicator of convergence. To improve our confidence in the predictive strength of $\Delta\hat{F}$, we can test one of its underlying assumptions: that the multiscale model becomes progressively more accurate as it receives more data. This assumption is true when we replace surrogate model predictions with incoming DFT results, but it is not necessarily true for unqueried points. We can estimate the accuracy on unqueried points by calculating the residuals between the surrogate model and the incoming DFT results (Figure 6-6b). As each “batch” of queries is received, we compare the queried, true adsorption energies with the energies predicted by the surrogate model just before retraining—i.e., the predictions used to choose that batch. Any improvements in accuracy on these points show that the overall, multiscale model is improving over time and that the $\Delta\hat{F}$ metric is an honest indicator of convergence. Figure 6-6b shows that model accuracy improves within the first *ca.* 10 batches (or 2,000 adsorption energy queries), but plateaus afterwards. This indicates that, after 10 batches, improvements in overall classification accuracy came from receipt of additional DFT data rather than improvements in surrogate model predictions.

Prediction accuracy of adsorption energies is not the only indicator of improved model performance. If a surrogate model’s accuracy does not change but its uncertainty predictions decrease/improve, then our confidence in the overall material classification may still improve. Of course, improvements in uncertainty must not be

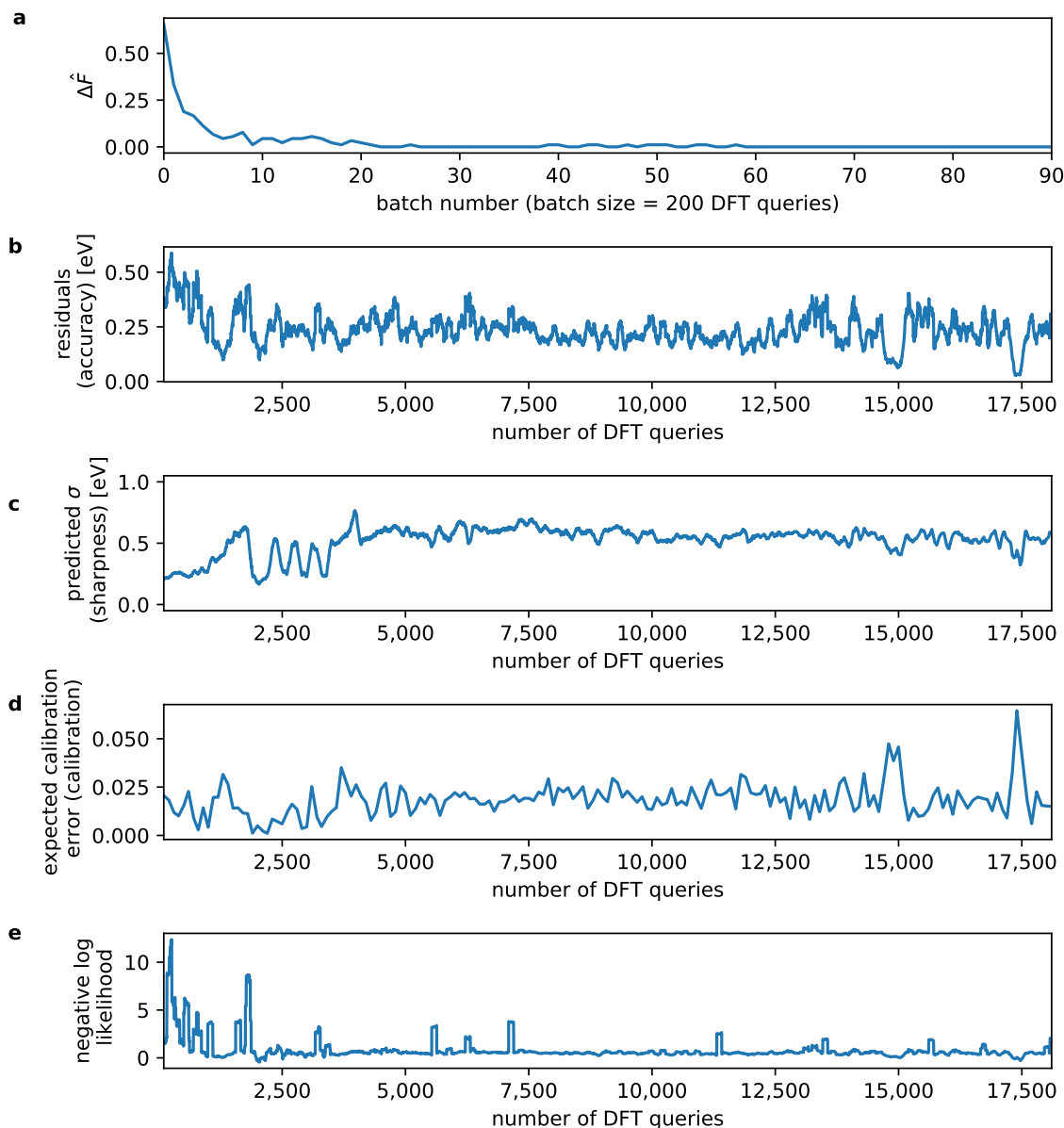


Figure 6-6: Example of diagnostic plots that we recommend monitoring during an active discovery campaign: **a.** predicted change in $F1$ score ($\Delta\hat{F}$); **b.** residuals between the real data and the surrogate model’s predictions; **c.** expected calibration error[170] of the surrogate model; **d.** the predicted uncertainties of surrogate model in the form of the predicted standard deviation (σ); and **e.** the negative-log-likelihood of the surrogate model.[170] These results were simulated by using the Myopic Multiscale Sampling (MMS) method with the Convolution-Fed Gaussian Process (CFGF) model on the GASpy dataset. For clarity of visualization, we plotted rolling averages of all values in this figure using a window of 100 queries (excluding the $\Delta\hat{F}$ values, where no averaging was done)

obtained at the expense of worse calibration. In other words, reductions in predicted uncertainties may also indicate improved model performance and better confidence in $\Delta\hat{F}$, but only if the expected calibration error[170] does not increase. In our illustrative example, Figure 6-6c shows the predicted uncertainty while Figure 6-6d shows the calibration. Unfortunately, the uncertainty predictions do not decrease over the course of the discovery process. Note that all uncertainty and calibration estimates for each batch should be calculated using the surrogate model predictions used to choose that batch, just as was done for the residuals.

Lastly, we also recommend monitoring the negative-log-likelihood[170] of the surrogate model for each incoming batch. This metric incorporates model accuracy, calibration, and sharpness into a single metric. Lower values of negative-log-likelihood indicate better model performance. Figure 6-6e shows that this metric improves until *ca.* 2,000 queries, after which it stagnates. This is consistent with the improvement in accuracy until 2,000 queries and subsequent stagnation of all performance metrics thereafter.

6.5 Conclusions

Here we created a multi-scale modeling method for combining atomic-scale DFT results with surrogate/ML models to create actionable plans for experimentalists—i.e., a classification of catalysts as “worthy of experimental study” or “not worthy”. We then coupled this modeling method with a Myopic Multiscale Sampling (MMS) strategy to perform automated catalyst discovery via active classification. We tested this strategy on two hypothetical datasets using three different surrogate models, giving us an estimate on the range of performance we might see in the future. In some cases, the results show up to a 16-fold reduction in the number of DFT queries compared to random sampling. The degree of speed-up depends on the quality of the ML model used, the homogeneity of the search space, and the hyperparameters used to define convergence of the active classification. Speed-up estimates on more realistic use cases show a more conservative 7-fold reduction in number of DFT queries.

Lastly, we provide a set of recommended diagnostic metrics to use during active classification (Figure 6-6): $\Delta\hat{F}$ and the ML model’s residuals, uncertainty estimates, and calibration.

Our results elucidated a number of qualitative behaviors of active classification. First, we observed that higher-quality ML models yielded better initial performance of the classification process. Conversely, we observed that higher-quality sampling strategies yielded better rates of improvement over time. We also observed that our latest ML model (CFGP) yielded performance closer to a naive, ignorant model than to a perfect, omniscient model. This suggests that there is a relatively large amount of potential improvement left in the ML modeling space. Next, we observed that better sampling strategies (as quantified by $F1$ score) led to lower rates of change in classes (as quantified by $\Delta\hat{F}$), suggesting that $\Delta\hat{F}$ may be an indicator of sampling strategy performance. Conversely, we observed that slow-learning ML models may also reduce $\Delta\hat{F}$. This phenomena could be counteracted by using more conservative convergence criteria. All these details were observed in specific and synthetic use cases though. The behaviors seen here may not be observed in situations where search spaces and/or ML models differ.

We encourage readers to focus on the main goals of this work: (1) converting atomic-scale simulations and ML models into actionable decisions for experimentalists, and (2) relaxing the active discovery process from an optimization/regression problem to a classification problem. The ability to convert computational results into experimental recommendations helps us serve the research community better. Simultaneously, relaxing the discovery process to a classification problem helps us prioritize exploration rather than exploitation, which is more appropriate for early-stage discovery projects.

We also recognize several future directions that may stem from this research. Future work might include incorporation of DFT-calculated surface stability by performing weighted averaging of surface activities when calculating bulk activities. Future work may also include cost-weighted sampling such that less computationally intensive calculations are chosen more frequently than more intensive ones, which may

improve discovery rates in real-time. Perhaps most importantly, future work should incorporate some ability to feed experimental data and information to computational sampling strategies—e.g., multi-fidelity modeling.

Chapter 7

Conclusions

In Chapter 2, we begin by outlining several aspects of catalysis informatics and how they can be used. Specifically, we review the advantages and disadvantages of databases, workflow managers, surrogate models, and any combination thereof. The underpinning theme with these tools is that they require an initial overhead investment with varying rates of return, i.e., the project will take longer to start up. This means that these tools complement larger-scale projects that continue for longer times, because those projects are more likely to see a positive return on investment. We then outline an example of such an investment: GASpy, the software we built to perform high-throughput DFT calculations.

Next we paired GASpy with a heuristic active discovery algorithm to calculate *ca.* 40,000 CO and H adsorption energies, as outlined in Chapter 3. These adsorption energies were then combined with volcano scaling relationships to screen for candidate catalysts that may be active for either CO₂ reduction or H₂ evolution. Several of the candidates we discovered were already studied in literature and were found to be active, thereby validating our use of the GASpy framework. Unfortunately, most of the catalysts validated this way were unsuitable for further study for other reasons (e.g., insufficient selectivity or stability). This warranted follow-up studies for experimentalists to test the other candidates we found.

The Sargent group at The University of Toronto were kind enough to study one of our CO₂ reduction candidates in particular: CuAl. The results of this collaboration

are outlined in Chapter 4. We found that CuAl was indeed active in reducing CO₂, but we also found that it was able to produce ethylene with a selectivity of up to 80%—higher than we had ever seen before. These findings further validated our methods for catalyst discovery.

Given the heuristic nature of our discovery methods, we sought to develop more sophisticated algorithms for active discovery. But such work required that we first have surrogate models for predicting DFT-calculated adsorption energies and that these models also have corresponding uncertainty estimates for their predictions. Chapter 5 outlines how we and our collaborators in the Machine Learning Department attempted to create such a model. This chapter also outlines the methods we used to assess the quality of uncertainty predictions. We show how uncertainty predictions should be sharp (i.e., small), but not at the expense of being dishonest (i.e., they must be calibrated). Then we show that, of the various models we developed, a Convolution-Fed Gaussian Process (CFGP) yielded a relatively good balance between model accuracy, sharpness, and calibration.

Armed with CFGP, we continued to work with the Machine Learning department to begin development of an active discovery algorithm better tailored for our application, as outlined in Chapter 6. Here we created a multiscale method that estimated macroscopic catalyst activity by combining DFT calculations of adsorbate binding energies with CFGP predictions. The method bridged the gap between atomic-scale simulations and lab-scale decision making. Next we coupled the multiscale model with a Myopic Multiscale Sampling (MMS) method that we created. MMS comprised several active selection strategies for choosing calculations that helped us differentiate higher-quality catalyst candidates from lower-quality ones. Our key insight here was relaxing the discovery process from an optimization problem to a classification problem. This relaxation allowed us to implicitly favor exploration over exploitation, which was more appropriate for our early-stage catalyst discovery work.

This thesis illustrates a series of projects all aimed to accelerate the rate of catalyst discovery for solar fuel production. Although we show direct evidence of success both here and in other studies,[108] we believe that further work should be done to quantify

the degree of success. We believe that it would be valuable to quantify the impact that the catalysis informatics field is having on the field of experimental catalyst discovery. Such quantification would offer an explicit feedback mechanism between experimental and computational studies. It may also help us maintain focus on our primary goal: mitigating climate change through commercialization of sustainable energy storage technologies.

Bibliography

- [1] The NOMAD Laboratory: A European Centre of Excellence.
- [2] *Fifth Assessment Synthesis Report*. Intergovernmental Panel on Climate Change, 2014.
- [3] Annual Energy Outlook 2017 with projections to 2050. Technical report, U.S. Energy Information Administration, 2017.
- [4] World Energy Outlook 2017. Technical report, International Energy Agency, 2017.
- [5] Annual Energy Outlook 2017 with projections to 2050. Technical report, U.S. Energy Information Administration, 2020.
- [6] *World Energy Outlook 2020*. World Energy Outlook. OECD, 2020.
- [7] Frank Abild-Pedersen and Martin P. Andersson. CO adsorption energies on metals with correction for high coordination adsorption sites - A density functional study. *Surface Science*, 601(7):1747–1753, 2007.
- [8] Frank Abild-Pedersen, Jeffrey Greeley, Felix Studt, Jan Rossmeisl, Ture R. Munter, Poul G. Moses, Egill Skúlason, Thomas Bligaard, and Jens K. Nørskov. Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces. *Physical Review Letters*, 99(1):4–7, 2007.
- [9] Kirstin Alberi, Marco Buongiorno Nardelli, Andriy Zakutayev, Lubos Mitas, Stefano Curtarolo, Anubhav Jain, Marco Fornari, Nicola Marzari, Ichiro Takeuchi, Martin L Green, Mercouri Kanatzidis, Mike F Toney, Sergiy Butenko, Bryce Meredig, Stephan Lany, Ursula Kattner, Albert Davydov, Eric S Toberer, Vladan Stevanovic, Aron Walsh, Nam-Gyu Park, Alán Aspuru-Guzik, Daniel P Tabor, Jenny Nelson, James Murphy, Anant Setlur, John Gregoire, Hong Li, Ruijuan Xiao, Alfred Ludwig, Lane W Martin, Andrew M Rappe, Su-Huai Wei, and John Perkins. The 2019 materials by design roadmap. *J. Phys. D: Appl. Phys.*, 52(1):1—48, 2019.
- [10] Michael Altschuler and Michael Bloodgood. Stopping Active Learning Based on Predicted Change of F Measure for Text Classification. *Proceedings - 13th IEEE International Conference on Semantic Computing, ICSC 2019*, pages 47–54, 2019.

- [11] Nongnuch Artrith and Alexander Urban. An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO₂. *Computational Materials Science*, 114:135–150, 2016.
- [12] Seoin Back, Junwoong Yoon, Nianhan Tian, Wen Zhong, Kevin Tran, and Zachary W. Ulissi. Convolutional Neural Network of Atomic Surface Structures To Predict Binding Energies for High-Throughput Screening of Catalysts. *The Journal of Physical Chemistry Letters*, 10(15):4401–4408, 2019.
- [13] Jörg Behler. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Physical Chemistry Chemical Physics*, 13(40):17930, 2011.
- [14] Helen M. Berman, Tammy Battistuz, T. N. Bhat, Wolfgang F. Bluhm, Philip E. Bourne, Kyle Burkhardt, Zukang Feng, Gary L. Gilliland, Lisa Iype, Shri Jain, Phoebe Fagan, Jessica Marvin, David Padilla, Veerasamy Ravichandran, Bohdan Schneider, Narmada Thanki, Helge Weissig, John D. Westbrook, and Christine Zardecki. The protein data bank. *Nucleic Acids Research*, 28:235–242, 2000.
- [15] Erik Bernhardsson, Elias Freider, and Arash Rouhani. Luigi, a Python package that builds complex pipelines of batch jobs, 2012.
- [16] Eli Bingham, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D. Goodman. Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, 20:1–6, 2019.
- [17] Thomas Bligaard, Jens K Nørskov, S Dahl, J Matthiesen, C Christensen, and J. Sehested. The Brønsted-Evans-Polanyi relation and the volcano curve in heterogeneous catalysis. *Journal of Catalysis*, 224:206–217, 2004.
- [18] P E Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50(24):17953–17979, dec 1994.
- [19] Jacob R Boes and John R Kitchin. Modeling Segregation on AuPd(111) Surfaces with Density Functional Theory and Monte Carlo Simulations. *Journal of Physical Chemistry C*, 121(6):3479–3487, 2017.
- [20] Oleg Borodin, Marco Olguin, Carrie E. Spear, Kenneth W. Leiter, and Jaroslaw Knap. Towards high throughput screening of electrochemical stability of battery electrolytes. *Nanotechnology*, 26(35), 2015.
- [21] Brent Bryan, Robert C Nichol, Christopher R Genovese, Jeff Schneider, Christopher J Miller, and Larry Wasserman. Active learning for identifying function threshold boundaries. *Advances in neural information processing systems*, 18:163–170, 2005.

- [22] Federico Calle-Vallejo, David Loffreda, Marc T M Koper, and Philippe Sautet. Introducing structural sensitivity into adsorption-energy scaling relations by means of coordination numbers. *Nature Chemistry*, 7(5):403–410, 2015.
- [23] Federico Calle-Vallejo, José I. Martínez, Juan M. García-Lastra, Philippe Sautet, and David Loffreda. Fast prediction of adsorption properties for platinum nanocatalysts with generalized coordination numbers. *Angewandte Chemie - International Edition*, 53(32):8316–8319, 2014.
- [24] Zachary P. Cano, Dustin Banham, Siyu Ye, Andreas Hintennach, Jun Lu, Michael Fowler, and Zhongwei Chen. Batteries and fuel cells for emerging electric vehicle markets. *Nature Energy*, 3(4):279–289, 2018.
- [25] Sudip Chakraborty, Wei Xie, Nripan Mathews, Matthew Sherburne, Rajeev Ahuja, Mark Asta, and Subodh G. Mhaisalkar. Rational Design: A High-Throughput Computational Screening and Experimental Validation Methodology for Lead-Free and Emergent Hybrid Perovskites. *ACS Energy Letters*, 2(4):837–845, 2017.
- [26] Ian Char, Youngseog Chung, Willie Neiswanger, Kirthevasan Kandasamy, Andrew Oakleigh Nelson, Mark D. Boyer, Egemen Kolenen, and Jeff Schneider. Offline contextual Bayesian optimization for nuclear fusion. In *33rd Conference on Neural Information Processing Systems (NeurIPS)*, pages 1–12, 2020.
- [27] Pavel V. Cherepanov, Muthupandian Ashokkumar, and Daria V. Andreeva. Ultrasound assisted formation of Al-Ni electrocatalyst for hydrogen evolution. *Ultrasonics Sonochemistry*, 23:142–147, 2015.
- [28] Steven Chu, Yi Cui, and Nian Liu. The path towards sustainable energy. *Nature Materials*, 16(1):16–22, 2016.
- [29] Wei Chu, Martin Zinkevich, Lihong Li, Achint Thomas, and Belle Tseng. Unbiased online active learning in data streams. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 195–203, 2011.
- [30] Yongchul G. Chung, Jeffrey Camp, Maciej Haranczyk, Benjamin J. Sikora, Wojciech Bury, Vaiva Krungleviciute, Taner Yildirim, Omar K. Farha, David S. Sholl, and Randall Q. Snurr. Computation-ready, experimental metal-organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. *Chemistry of Materials*, 26(21):6185–6192, 2014.
- [31] Stephen Comello and Stefan Reichelstein. The emergence of cost effective battery storage. *Nature Communications*, 10(1):1–9, 2019.
- [32] Alison Cozad, Nikolaos V. Sahinidis, and David C. Miller. Learning surrogate models for simulation-based optimization. *AIChE Journal*, 60(6):2211–2227, 2014.

- [33] Stefano Curtarolo, Gus L.W. Hart, Marco Buongiorno Nardelli, Natalio Mingo, Stefano Sanvito, and Ohad Levy. The high-throughput highway to computational materials design. *Nature Materials*, 12(3):191–201, 2013.
- [34] Stefano Curtarolo, Wahyu Setyawan, Shidong Wang, Junkai Xue, Kesong Yang, Richard H. Taylor, Lance J. Nelson, Gus L W Hart, Stefano Sanvito, Marco Buongiorno-Nardelli, Natalio Mingo, and Ohad Levy. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science*, 58:227–235, 2012.
- [35] Stuart J. Davie, Nicodemo Di Pasquale, and Paul L.A. Popelier. Kriging atomic properties with a variable number of inputs. *Journal of Chemical Physics*, 145(10):1–11, 2016.
- [36] Lawrence E. Davis, Noel C. MacDonald, Paul W. Palmberg, and Gerald Riach. *Handbook of Auger Electron Spectroscopy*. Physical Electronics Industries, 2 edition, 1976.
- [37] Alexander Philip Dawid and Monica Musio. Theory and applications of proper scoring rules. *Metron*, 72(2):169–183, 2014.
- [38] John P. De Long and Oskar Burger. Socio-economic instability and the scaling of energy use with population size. *PLoS ONE*, 10(6):1–12, 2015.
- [39] Phil De Luna, Rafael Quintero-Bermudez, Cao-Thang Dinh, Michael B. Ross, Oleksandr S. Bushuyev, Petar Todorović, Tom Regier, Shana O. Kelley, Peidong Yang, and Edward H. Sargent. Catalyst electro-redeposition controls morphology and oxidation state for selective carbon dioxide reduction. *Nature Catalysis*, 1(February), 2018.
- [40] Thomas Desautels, Andreas Krause, and Joel Burdick. Parallelizing exploration-exploitation tradeoffs with Gaussian process bandit optimization. *Journal of Machine Learning Research*, 15:4053—4103, 2014.
- [41] Cao Thang Dinh, Thomas Burdyny, Golam Kibria, Ali Seifitokaldani, Christine M. Gabardo, F. Pelayo García De Arquer, Amirreza Kiani, Jonathan P. Edwards, Phil De Luna, Oleksandr S. Bushuyev, Chengqin Zou, Rafael Quintero-Bermudez, Yuanjie Pang, David Sinton, and Edward H. Sargent. CO₂ electroreduction to ethylene via hydroxide-mediated copper catalysis at an abrupt interface. *Science*, 360(6390):783–787, 2018.
- [42] John Eason and Selen Cremaschi. Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers and Chemical Engineering*, 68:220–232, 2014.
- [43] Ottmar Edenhofer, Ramon Pichs Madruga, and Youba Sokona. *Renewable Energy Sources and Climate Change Mitigation (Special Report of the Intergovernmental Panel on Climate Change)*. Cambridge University Press, 1 edition, 2012.

- [44] Raul A. Flores, Christopher Paolucci, Kirsten T. Winther, Ankit Jain, Jose Antonio Garrido Torres, Muratahan Aykol, Joseph Montoya, Jens K. Nørskov, Michal Bajdich, and Thomas Bligaard. Active Learning Accelerated Discovery of Stable Iridium Oxide Polymorphs for the Oxygen Evolution Reaction. *Chemistry of Materials*, 32(13):5854–5863, 2020.
- [45] Peter I Frazier. A Tutorial on Bayesian Optimization, 2018.
- [46] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian Approximation : Representing Model Uncertainty in Deep Learning. In *33rd International Conference on Machine Learning*, New York, NY, USA, 2016.
- [47] Shan Gao, Yue Lin, Xingchen Jiao, Yongfu Sun, Qiquan Luo, Wenhua Zhang, Dianqi Li, Jinlong Yang, and Yi Xie. Partially oxidized atomic cobalt layers for carbon dioxide electroreduction to liquid fuel. *Nature*, 529(7584):68–71, 2016.
- [48] Jacob R. Gardner, Geoff Pleiss, David Bindel, Kilian Q. Weinberger, and Andrew Gordon Wilson. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. In *32nd Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [49] Roman Garnett, Yamuna Krishnamurthy, Xuehan Xiong, Jeff Schneider, and Richard Mann. Bayesian optimal active search and surveying. *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, 2:1239–1246, 2012.
- [50] Tilman Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.
- [51] Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, David Duvenaud, Dougal Maclaurin, Martin A. Blood-Forsythe, Hyun Sik Chae, Markus Einzinger, Dong Gwang Ha, Tony Wu, Georgios Markopoulos, Soonok Jeon, Hosuk Kang, Hiroshi Miyazaki, Masaki Numata, Sunghan Kim, Wenliang Huang, Seong Ik Hong, Marc Baldo, Ryan P. Adams, and Alán Aspuru-Guzik. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials*, 15(10):1120–1127, 2016.
- [52] Alkis Gotovos, Nathalie Casati, Gregory Hitz, and Andreas Krause. Active learning for level set estimation. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 1344–1350. AAAI Press, 2013.
- [53] Jeff Greeley, Thomas F. Jaramillo, Jacob Bonde, Ib Chorkendorff, and Jens K. Nørskov. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature Materials*, 5(11):909–913, 2006.

- [54] M. L. Green, C. L. Choi, J. R. Hattrick-Simpers, A. M. Joshi, I. Takeuchi, S. C. Barron, E. Campo, T. Chiang, S. Empedocles, J. M. Gregoire, A. G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt, J. Van Duren, and A. Zakutayev. Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies. *Applied Physics Reviews*, 4(1), 2017.
- [55] Geun Ho Gu, Juhwan Noh, Inkyung Kim, and Yousung Jung. Machine learning for renewable energy materials. *Journal of Materials Chemistry A*, 7(29):17096–17117, 2019.
- [56] Konstantin Gubaev, Evgeny V. Podryabinkin, and Alexander V. Shapeev. Machine learning of molecular properties: Locality and active learning. *Journal of Chemical Physics*, 148(24):1–9, 2018.
- [57] Bjørk Hammer, Lars Boye Hansen, and Jens Nørskov. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Physical Review B*, 59(11):7413–7421, 1999.
- [58] Bjørk Hammer and Jens K. Nørskov. Electronic factors determining the reactivity of metal surfaces. *Surface Science*, 343(3):211–220, 1995.
- [59] Zhong-Hua Han and Ke-Shi Zhang. Surrogate-Based Optimization. In Olympia Roeva, editor, *Real-World Applications of Genetic Algorithms*, chapter 17. In-Tech, 2012.
- [60] Chris M Handley and Paul L a Popelier. Potential energy surfaces fitted by artificial neural networks. *Journal of Physical Chemistry A*, 114(10):3371–83, 2010.
- [61] H. A. Hansen, C. Shi, A. C. Lausche, A. A. Peterson, and J. K. Nørskov. Bifunctional alloys for the electroreduction of CO₂ and CO. *Physical Chemistry Chemical Physics*, 18(13):9194–9201, 2016.
- [62] Graeme Henkelman, Blas P Uberuaga, Hannes Jónsson, and Graeme Henkelman. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, 113(22):9901–9904, 2000.
- [63] Joanne Hill, Arun Mannodi-Kanakkithodi, Ramamurthy Ramprasad, and Bryce Meredig. Materials Data Infrastructure and Materials Informatics. In Dongwon Shin and James Sall, editors, *Computational Materials System Design*, pages 193–225. Springer International Publishing, 2018.
- [64] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, Eric D Hermes, Paul C Jennings, Peter Bjerre Jensen, James Kermode, John R Kitchin, Esben Leonhard Kolsbjerg, Joseph Kubal, Kristen Kaasbjerg, Steen Lysgaard, Jón Bergmann Maronsson, Tristan Maxson,

- Thomas Olsen, Lars Pastewka, Andrew Peterson, Carsten Rostgaard, Jakob Schiøtz, Ole Schütt, Mikkel Strange, Kristian S Thygesen, Tejs Vegge, Lasse Villhelmsen, Michael Walter, Zhenhua Zeng, and Karsten Wedel Jacobsen. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.
- [65] Steven C. H. Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. Online Learning: A Comprehensive Survey, 2018.
- [66] Yoshio Hori, Ichiro Takahashi, Osamu Koga, and Nagahiro Hoshi. Selective formation of C2 compounds from electrochemical reduction of CO2 at a series of copper single crystal electrodes. *Journal of Physical Chemistry B*, 106(1):15–17, 2002.
- [67] Benjamin Z. Houlton, Maya Almaraz, Viney Aneja, Amy T. Austin, Edith Bai, Kenneth G. Cassman, Jana E. Compton, Eric A. Davidson, Jan Willem Erisman, James N. Galloway, Baojing Gu, Guolin Yao, Luiz A. Martinelli, Kate Scow, William H. Schlesinger, Thomas P. Tomich, Chao Wang, and Xin Zhang. A World of Cobenefits: Solving the Global Nitrogen Challenge. *Earth’s Future*, 7(8):865–872, 2019.
- [68] Jens S. Hummelshøj, Frank Abild-Pedersen, Felix Studt, Thomas Bligaard, and Jens K. Nørskov. CatApp: A web application for surface chemistry and heterogeneous catalysis. *Angewandte Chemie - International Edition*, 51(1):272–274, 2012.
- [69] Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and practice*. otexts.com, 1 edition, 2014.
- [70] Efthymios I. Ioannidis, Terry Z.H. Gani, and Heather J. Kulik. molSimplify: A toolkit for automating discovery in inorganic chemistry. *Journal of Computational Chemistry*, pages 2106–2117, 2016.
- [71] Anubhav Jain, Shyue Ping Ong, Wei Chen, Bharat Medasani, Xiaohui Qu, Michael Kocher, Miriam Brafman, Guido Petretto, Gian-Marco Rignanese, Geoffroy Hautier, Daniel Gunter, and Kristin A. Persson. FireWorks: a dynamic workflow system designed for high- throughput applications. *Concurrency Computation: Practice and Experience*, 27:5037–5059, 2015.
- [72] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 011002(1):1–11, 2013.
- [73] Anubhav Jain, Kristin A Persson, and Gerbrand Ceder. Research Update : The materials genome initiative : Data sharing and the impact of collaborative ab initio databases. *APL Materials*, 053102(4):1–14, 2016.

- [74] Jon Paul Janet, Chenru Duan, Tzuhsiung Yang, Aditya Nandy, and Heather J. Kulik. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chemical Science*, 10(34):7913–7922, 2019.
- [75] Philippe Jeanty, Christian Scherer, Erhard Magori, Kerstin Wiesner-Fleischer, Olaf Hinrichsen, and Maximilian Fleischer. Upscaling and continuous operation of electrochemical CO₂ to CO conversion in aqueous solutions on silver gas diffusion electrodes. *Journal of CO₂ Utilization*, 24(March):454–462, 2018.
- [76] Ryosuke Jinnouchi, Jonathan Lahnsteiner, Ferenc Karsai, Georg Kresse, and Menno Bokdam. Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference. *Physical Review Letters*, 122(22):225701, 2019.
- [77] Ilana Y. Kanal, Steven G. Owens, Jonathon S. Bechtel, and Geoffrey R. Hutchison. Efficient computational screening of organic polymer photovoltaics. *Journal of Physical Chemistry Letters*, 4(10):1613–1623, 2013.
- [78] Kirthivasan Kandasamy, Willie Neiswanger, Reed Zhang, Akshay Krishnamurthy, Jeff Schneider, and Barnabas Poczos. Myopic Bayesian Design of Experiments via Posterior Sampling and Probabilistic Programming, 2018.
- [79] Kirthivasan Kandasamy, Willie Neiswanger, Reed Zhang, Akshay Krishnamurthy, Jeff Schneider, and Barnabas Poczos. Myopic posterior sampling for adaptive goal oriented design of experiments. In *International Conference on Machine Learning*, pages 3222–3232, 2019.
- [80] Alireza Khorshidi and Andrew Peterson. Amp: A modular approach to machine learning in atomistic simulations. *Computer Physics Communications*, 207:1–15, 2016.
- [81] Yoolhee Kim, Edward Kim, Erin Antono, Bryce Meredig, and Julia Ling. Machine-learned metrics for predicting the likelihood of success in materials discovery. *npj Computational Materials*, 2019.
- [82] Scott Kirklin, James E. Saal, Bryce Meredig, Alex Thompson, Jeff W. Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies. *npj Computational Materials*, 1(September), 2015.
- [83] John R. Kitchin. vasp: New ASE compliant Python interface to VASP. Technical report.
- [84] John R. Kitchin. Examples of effective data sharing in scientific publishing. *ACS Catalysis*, 5(6):3894–3899, 2015.
- [85] Brian Kolb, Levi C. Lentz, and Alexie M. Kolpak. Discovering charge density functionals and structure-property relationships with PROPhet: A general

- framework for coupling machine learning and first-principles methods. *Scientific Reports*, 7(1):1–9, 2017.
- [86] R. Kortlever, I. Peters, C. Balemans, R. Kas, Y. Kwon, G. Mul, and M. T. M. Koper. Palladium-gold catalyst for the electrochemical reduction of CO₂ to C₂–C₅ hydrocarbons. *Chemical Communications*, 52(67):10229–10232, 2016.
 - [87] Ruud Kortlever, Jing Shen, Klaas Jan P. Schouten, Federico Calle-Vallejo, and Marc T.M. Koper. Catalysts and Reaction Pathways for the Electrochemical Reduction of Carbon Dioxide. *Journal of Physical Chemistry Letters*, 6(20):4073–4082, 2015.
 - [88] G Kresse and D Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59(3):1758–1775, jan 1999.
 - [89] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, 1996.
 - [90] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16):11169–11186, 1996.
 - [91] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47(1):558–561, 1993.
 - [92] Georg Kresse and Jürgen Hafner. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B*, 49(20):14251–14269, 1994.
 - [93] Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate Uncertainties for Deep Learning Using Calibrated Regression. In *35th International Conference on Machine Learning*, Stockholm, Sweden, 2018.
 - [94] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *31st Conference on Neural Information Processing Systems (NIPS)*, 2017.
 - [95] David D. Landis, Jens S. Hummelshøj, Svetlozar Nestorov, Jeffrey Greeley, Marcin Dulak, Thomas Bligaard, Jens K. Nørskov, and Karsten W. Jacobsen. The Computational Materials Repository. *Computing in Science & Engineering*, 14(6):51–57, 2012.
 - [96] Dan Levi, Liran Gispán, Niv Giladi, and Ethan Fetaya. Evaluating and Calibrating Uncertainty Prediction in Regression Tasks, 2020.
 - [97] Peter G. Levi and Jonathan M. Cullen. Mapping Global Flows of Chemicals: From Fossil Fuel Feedstocks to Chemical Products. *Environmental Science and Technology*, 52(4):1725–1734, 2018.

- [98] Nathan S Lewis and Daniel G Nocera. Powering the planet: Chemical challenges in solar energy utilization. *Proceedings of the National Academy of Sciences*, 104(50):15729–15735, 2007.
- [99] Christina W. Li, Jim Ciston, and Matthew W. Kanan. Electroreduction of carbon monoxide to liquid fuel on oxide-derived nanocrystalline copper. *Nature*, 508(7497):504–507, 2014.
- [100] Yuguang C. Li, Dekai Zhou, Zhifei Yan, Ricardo H. Gonçalves, Danielle A. Salvatore, Curtis P. Berlinguette, and Thomas E. Mallouk. Electrolysis of CO₂ to Syngas in Bipolar Membrane-Based Electrochemical Cells. *ACS Energy Letters*, 1(6):1149–1153, 2016.
- [101] Zheng Li, Siwen Wang, Wei Shan Chin, Luke E. Achenie, and Hongliang Xin. High-throughput screening of bimetallic catalysts enabled by machine learning. *Journal of Materials Chemistry A*, 5:24131–24138, 2017.
- [102] Hanbin Liao, Chao Wei, Jingxian Wang, Adrian Fisher, Thirumany Sritharan, Zhenxing Feng, and Zhichuan J. Xu. A Multisite Strategy for Enhancing the Hydrogen Evolution Reaction on a Nano-Pd Surface in Alkaline Media. *Advanced Energy Materials*, 7(21):1–7, 2017.
- [103] Song Lin, Christian S. Diercks, Yue Biao Zhang, Nikolay Kornienko, Eva M. Nichols, Yingbo Zhao, Aubrey R. Paris, Dohyung Kim, Peidong Yang, Omar M. Yaghi, and Christopher J. Chang. Covalent organic frameworks comprising cobalt porphyrins for catalytic CO₂ reduction in water. *Science*, 349(6253):1208–1213, 2015.
- [104] Min Liu, Yuanjie Pang, Bo Zhang, Phil De Luna, Oleksandr Voznyy, Jixian Xu, Xueli Zheng, Cao Thang Dinh, Fengjia Fan, Changhong Cao, F. Pelayo García De Arquer, Tina Saberi Safaei, Adam Mepham, Anna Klinkova, Eugenia Kumacheva, Tobin Filleter, David Sinton, Shana O. Kelley, and Edward H. Sargent. Enhanced electrocatalytic CO₂ reduction via field-induced reagent concentration. *Nature*, 537(7620):382–386, 2016.
- [105] Xinyan Liu, Jianping Xiao, Hongjie Peng, Xin Hong, Karen Chan, and Jens K. Nørskov. Understanding trends in electrochemical carbon dioxide reduction rates. *Nature Communications*, 8, 2017.
- [106] Turab Lookman, Francis J. Alexander, and Krishna Rajan, editors. *Information Science for Materials Discovery and Design*. Springer International Publishing, 1 edition, 2016.
- [107] Turab Lookman, Prasanna V. Balachandran, Dezhen Xue, and Ruihao Yuan. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Computational Materials*, 5(1), 2019.

- [108] Eric M Lopato, Emily A Eikey, Zoe C Simon, Seoin Back, Kevin Tran, Jacqueline Lewis, Jakub F Kowalewski, Sadegh Yazdi, John R Kitchin, Zachary W Ulissi, Jill E Millstone, and Stefan Bernhard. Parallelized Screening of Characterized and DFT-Modelled Bimetallic Colloidal Co-Catalysts for Photocatalytic Hydrogen Evolution. *ACS Catalysis*, 10:4244—4252, 2020.
- [109] Qi Lu, Jonathan Rosen, Yang Zhou, Gregory S. Hutchings, Yannick C. Kimmel, Jingguang G. Chen, and Feng Jiao. A selective and efficient electrocatalyst for carbon dioxide reduction. *Nature Communications*, 5:1–6, 2014.
- [110] Mentel Łukasz. Mendeleeev, 2014.
- [111] Yanwei Lum, Binbin Yue, Peter Lobaccaro, Alexis T. Bell, and Joel W. Ager. Optimizing C-C Coupling on Oxide-Derived Copper Catalysts for Electrochemical CO₂ Reduction. *Journal of Physical Chemistry C*, 121(26):14191–14203, 2017.
- [112] Yifei Ma, Dougal J. Sutherland, Roman Garnett, and Jeff Schneider. Active pointillistic pattern search. In *18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 38, pages 672–680, 2015.
- [113] Laurens Van Der Maaten. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research*, 15:1–21, 2014.
- [114] David J C Mackay. *Sustainable Energy—without the hot air*, volume 2. UIT Cambridge Ltd., 2009.
- [115] Robert C. Martin. *Clean Code - a Handbook of Agile Software Craftsmanship*. Prentice Hall, 1 edition, 2009.
- [116] Kiran Mathew, Joseph H. Montoya, Alireza Faghaninia, Shyam Dwarakanath, Muratahan Aykol, Hanmei Tang, Iek heng Chu, Tess Smidt, Brandon Bocklund, Matthew Horton, John Dagdelen, Brandon Wood, Zi Kui Liu, Jeffrey Neaton, Shyue Ping Ong, Kristin Persson, and Anubhav Jain. Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Computational Materials Science*, 139:140–152, 2017.
- [117] Andrew J Medford, M .Ross Kunz, Sarah M Ewing, Tammie Borders, and Rebecca Fushimi. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catalysis*, 8:7403–7429, 2018.
- [118] Andrew J. Medford, Chuan Shi, Max J. Hoffmann, Adam C. Lausche, Sean R. Fitzgibbon, Thomas Bligaard, and Jens K. Nørskov. CatMAP: A Software Package for Descriptor-Based Microkinetic Mapping of Catalytic Trends. *Catalysis Letters*, 145(3):794–807, 2015.
- [119] Andrew J. Medford, Aleksandra Vojvodic, Jens S. Hummelshøj, Johannes Voss, Frank Abild-Pedersen, Felix Studt, Thomas Bligaard, Anders Nilsson,

- and Jens K. Nørskov. From the Sabatier principle to a predictive theory of transition-metal heterogeneous catalysis. *Journal of Catalysis*, 328:36–42, 2015.
- [120] J. Melorose, R. Perroy, and S. Careas. World population prospects. *United Nations*, 1(6042):587–92, 2015.
- [121] Bryce Meredig, Ankit Agrawal, Scott Kirklin, James E. Saal, Jeff W. Doak, Alex Thompson, Ke-Shi Zhang, Alok Choudhary, and Christopher Wolverton. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Physical Review B - Condensed Matter and Materials Physics*, 89(9):1–7, 2014.
- [122] Bryce Meredig, Erin Antono, Carena Church, Maxwell Hutchinson, Julia Ling, Sean Paradiso, Ben Blaiszik, Ian Foster, Brenna Gibbons, Jason Hattrick-Simpers, Apurva Mehta, and Logan Ward. Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Molecular Systems Design and Engineering*, 3(5):819–825, 2018.
- [123] Bryce Meredig and Chris Wolverton. Dissolving the periodic table in cubic zirconia: Data mining to discover chemical trends. *Chemistry of Materials*, 26(6):1985–1991, 2014.
- [124] Tim Meyer, Marco D’Abramo, Adam Hospital, Manuel Rueda, Carles Ferrer-Costa, Alberto Pérez, Oliver Carrillo, Jordi Camps, Carles Fenollosa, Dmitry Repchevsky, Josep Lluís Gelpí, and Modesto Orozco. MoDEL (Molecular Dynamics Extended Library): A Database of Atomistic Molecular Dynamics Trajectories. *Structure*, 18(11):1399–1409, 2010.
- [125] Hemma Mistry, Ana Sofia Varela, Cecile S. Bonifacio, Ioannis Zegkinoglou, Ilya Sinev, Yong Wook Choi, Kim Kisslinger, Eric A. Stach, Judith C. Yang, Peter Strasser, and Beatriz Roldan Cuenya. Highly selective plasma-activated copper catalysts for carbon dioxide reduction to ethylene. *Nature Communications*, 7, 2016.
- [126] Jonas Moćkus. On Bayesian methods for seeking the extremum. In *Optimization techniques IFIP technical conference*, pages 400–404. Springer, 1975.
- [127] Hendrik J Monkhorst and James D Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188–5192, jun 1976.
- [128] Joseph H. Montoya and Kristin A. Persson. A high-throughput framework for determining adsorption energies on solid surfaces. *npj Computational Materials*, 3(1):14, 2017.
- [129] Joseph H. Montoya, Linsey C. Seitz, Pongkarn Chakthranont, Aleksandra Vojvodic, Thomas F. Jaramillo, and Jens K. Nørskov. Materials for solar fuels and chemicals. *Nature Materials*, 16(1):70–81, 2016.

- [130] Masayuki Morimoto, Yoshiyuki Takatsuji, Ryota Yamasaki, Hikaru Hashimoto, Ikumi Nakata, Tatsuya Sakakura, and Tetsuya Haruyama. Electrodeposited Cu-Sn Alloy for Electrochemical CO₂ Reduction to CO/HCOO⁻. *Electrocatalysis*, pages 1–10, 2017.
- [131] Félix Musil, Michael J. Willatt, Mikhail A. Langovoy, and Michele Ceriotti. Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. *Journal of Chemical Theory and Computation*, 15(2):906–915, 2019.
- [132] Takahito Nakajima and Keisuke Sawada. Discovery of Pb-free perovskite solar cells via high-throughput simulation on the K computer. *Journal of Physical Chemistry Letters*, 8(19):4826–4831, 2017.
- [133] J. K. Nørskov, T. Bligaard, A. Logadottir, J. R. Kitchin, J. G. Chen, S. Pandalov, and U. Stimming. Trends in the Exchange Current for Hydrogen Evolution. *Journal of The Electrochemical Society*, 152(3):J23, 2005.
- [134] Jens K. Nørskov, Felix Studt, Frank Abild-Pedersen, and Thomas Bligaard. *Fundamental Concepts in Heterogeneous Catalysis*. John Wiley & Sons, Inc., 2015.
- [135] Randal S Olson, Ryan J Urbanowicz, Peter C Andrews, Nicole A Lavender, La Creis Kidd, and Jason H Moore. Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. In *European Conference on the Applications of Evolutionary Computation*, pages 123–137, Porto, Portugal, 2016. Springer International Publishing.
- [136] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.
- [137] Robert G Parr and Yang Weitao. *Density-Functional Theory of Atoms and Molecules*. International Series of Monographs on Chemistry. Oxford University Press, 1994.
- [138] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2012.
- [139] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 1996.

- [140] Richard Perez and Marc Perez. A fundamental look at supply side energy reserves for the planet. *The International Agency SHC Programme Solar Update*, 50:2—3, 2009.
- [141] Kristin A. Persson, Bryn Waldwick, Predrag Lazic, and Gerbrand Ceder. Prediction of solid-aqueous equilibria: Scheme to combine first-principles calculations of solids with experimental aqueous states. *Physical Review B - Condensed Matter and Materials Physics*, 85(23):1–12, 2012.
- [142] Andrew A. Peterson. Acceleration of saddle-point searches with machine learning. *Journal of Chemical Physics*, 145(7), 2016.
- [143] Andrew A. Peterson, Frank Abild-Pedersen, Felix Studt, Jan Rossmeisl, and Jens K. Nørskov. How copper catalyzes the electroreduction of carbon dioxide into hydrocarbon fuels. *Energy & Environmental Science*, 3(9):1311, 2010.
- [144] Andrew A Peterson, Rune Christensen, and Alireza Khorshidi. Addressing uncertainty in atomistic machine learning. *Phys. Chem. Chem. Phys. Phys. Chem. Chem. Phys*, 19(19):10978–10985, 2017.
- [145] Giovanni Pizzi, Andrea Cepellotti, Riccardo Sabatini, Nicola Marzari, and Boris Kozinsky. AiiDA: automated interactive infrastructure and database for computational science. *Computational Materials Science*, 111:218–230, 2015.
- [146] Albert Polman, Mark Knight, Erik C. Garnett, Bruno Ehrler, and Wim C. Sinke. Photovoltaic materials: Present efficiencies and future challenges. *Science*, 352(6283), 2016.
- [147] Rampi Ramprasad, Rohit Batra, Ghanshyam Pilania, Arun Mannodi-Kanakkithodi, and Chiho Kim. Machine learning in materials informatics: Recent applications and prospects. *npj Computational Materials*, 3(1), 2017.
- [148] Johan Rockström, Will Steffen, Kevin Noone, Åsa Persson, F. Stuart Chapin, Eric F. Lambin, Timothy M. Lenton, Marten Scheffer, Carl Folke, Hans Joachim Schellnhuber, Björn Nykvist, Cynthia A. de Wit, Terry Hughes, Sander van der Leeuw, Henning Rodhe, Sverker Sörlin, Peter K. Snyder, Robert Costanza, Uno Svedin, Malin Falkenmark, Louise Karlberg, Robert W. Corell, Victoria J. Fabry, James Hansen, Brian Walker, Diana Liverman, Katherine Richardson, Paul Crutzen, and Jonathan A. Foley. A safe operating space for humanity. *Nature*, 461(7263):472–475, 2009.
- [149] Brian Rohr, Helge S. Stein, Dan Guevarra, Yu Wang, Joel A. Haber, Muratahan Aykol, Santosh K. Suram, and John M. Gregoire. Benchmarking the acceleration of materials discovery by sequential learning. *Chemical Science*, 11(10):2696–2706, 2020.
- [150] David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha

- Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling Climate Change with Machine Learning, 2019.
- [151] Daniel H. Rothman. Thresholds of catastrophe in the Earth system. *Science Advances*, 3(9):1–13, 2017.
- [152] James E. Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and C. Wolverton. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). *Jom*, 65(11):1501–1509, 2013.
- [153] Gabriele Scalia, Colin A. Grambow, Barbara Pernici, Yi-Pei Li, and William H. Green. Evaluating Scalable Uncertainty Estimation Methods for DNN-Based Molecular Property Prediction, 2019.
- [154] Matthias Scheffler and Claudia Draxl. Computer Center of the Max-Planck Society, Garching, The NoMaD Repository, 2014.
- [155] Gabriel Ravanhani Schleder, Antonio Claudio Michejevs Padilha, Carlos Mera Acosta, Marcio Costa, and Adalberto Fazzio. From DFT to Machine Learning: recent approaches to Materials Science – a review. *Journal of Physics: Materials*, 2:1–46, 2019.
- [156] Philomena Schlexer Lamoureux, Kirsten T. Winther, Jose Antonio Garrido Torres, Verena Streibel, Meng Zhao, Michal Bajdich, Frank Abild-Pedersen, and Thomas Bligaard. Machine Learning for Computational Heterogeneous Catalysis. *ChemCatChem*, 11(16):3581–3601, 2019.
- [157] Jonathan Schmidt, Mário R.G. Marques, Silvana Botti, and Miguel A.L. Marques. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials*, 5(1), 2019.
- [158] Marcel Schreier, Florent Héroguel, Ludmilla Steier, Shahzada Ahmad, Jeremy S. Luterbacher, Matthew T. Mayer, Jingshan Luo, and Michael Grätzel. Solar conversion of CO₂ to CO using earth-abundant electrocatalysts prepared by atomic layer modification of CuO. *Nature Energy*, 2(7):1–9, 2017.
- [159] Anthony Scopatz and Kathryn D. Huff. *Effective Computation in Physics*. O’Reilly Media, 2015.
- [160] Zhi Wei Seh, Jakob Kibsgaard, Colin F. Dickens, Ib Chorkendorff, Jens K. Nørskov, and Thomas F. Jaramillo. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science*, 355(6321):1–12, 2017.
- [161] Atsuto Seko, Atsushi Togo, Hiroyuki Hayashi, Koji Tsuda, Laurent Chaput, and Isao Tanaka. Prediction of Low-Thermal-Conductivity Compounds with

- First-Principles Anharmonic Lattice-Dynamics Calculations and Bayesian Optimization. *Physical Review Letters*, 115(20):1–5, 2015.
- [162] Burr Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.
 - [163] Burr Settles. *Active Learning*. Morgan & Claypool, 2012.
 - [164] Chuan Shi, Heine A. Hansen, Adam C. Lausche, and Jens K. Nørskov. Trends in electrochemical CO₂ reduction activity for open and close-packed metal surfaces. *Physical Chemistry Chemical Physics*, 16(10):4720, 2014.
 - [165] Felix Studt, Frank Abild-Pedersen, Joel B. Varley, and Jens K. Nørskov. CO and CO₂ hydrogenation to methanol calculated using the BEEF-vdW functional. *Catalysis Letters*, 143(1):71–73, 2013.
 - [166] Steve K. Thompson. Stratified Sampling. In Walter A. Shewhart and Samuel S. Wilks, editors, *Sampling*, chapter 11, pages 139–156. John Wiley and Sons Inc., 3 edition, 2012.
 - [167] Daniel A. Torelli, Sonja A. Francis, J. Chance Crompton, Alnald Javier, Jonathan R. Thompson, Bruce S. Brunschwig, Manuel P. Soriaga, and Nathan S. Lewis. Nickel-Gallium-Catalyzed Electrochemical Reduction of CO₂ to Highly Reduced Products at Low Overpotentials. *ACS Catalysis*, 6(3):2100–2104, 2016.
 - [168] José A. Garrido Torres, Paul C. Jennings, Martin H. Hansen, Jacob R. Boes, and Thomas Bligaard. Low-Scaling Algorithm for Nudged Elastic Band Calculations Using a Surrogate Machine Learning Model. *Physical Review Letters*, 156001:1–6, 2018.
 - [169] Kevin Tran, Palizhati Aini, Seoin Back, and Zachary W Ulissi. Dynamic Workflows for Routine Materials Discovery in Surface Science. *Journal of Chemical Information and Modeling*, 58(12):2392–2400, 2018.
 - [170] Kevin Tran, Willie Neiswanger, Junwoong Yoon, Qingyang Zhang, Eric Xing, and Zachary W Ulissi. Methods for comparing uncertainty quantifications for material property predictions. *Machine Learning: Science and Technology*, 1(025006), 2020.
 - [171] Kevin Tran and Zachary W. Ulissi. Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution. *Nature Catalysis*, 1:696–703, 2018.
 - [172] Zachary W Ulissi, Andrew J Medford, Thomas Bligaard, and Jens K Nørskov. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nature Communications*, 8, 2017.

- [173] Zachary W Ulissi, Michael T Tang, Jianping Xiao, Xinyan Liu, Daniel A. Torelli, Mohammadreza Karamad, Kyle Cummins, Christopher Hahn, Nathan S. Lewis, Thomas F. Jaramillo, Karen Chan, and Jens K. Nørskov. Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO₂ Reduction. *ACS Catalysis*, 7(10):6600–6608, 2017.
- [174] Jonathan Vandermause, Steven B. Torrisi, Simon Batzner, Yu Xie, Lixin Sun, Alexie M. Kolpak, and Boris Kozinsky. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Computational Materials*, 6(1):1–11, 2020.
- [175] Lei Wang, Stephanie A. Nitopi, Erlend Bertheussen, Marat Orazov, Carlos G. Morales-Guio, Xinyan Liu, Drew C. Higgins, Karen Chan, Jens K. Nørskov, Christopher Hahn, and Thomas F. Jaramillo. Electrochemical Carbon Monoxide Reduction on Polycrystalline Copper: Effects of Potential, Pressure, and pH on Selectivity toward Multicarbon and Oxygenated Products. *ACS Catalysis*, 8(8):7445–7454, 2018.
- [176] Logan Ward, Ankit Agrawal, Alok Choudhary, and Christopher Wolverton. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2(July):1–7, 2016.
- [177] Logan Ward, Ruoqian Liu, Amar Krishna, Vinay I. Hegde, Ankit Agrawal, Alok Choudhary, and Chris Wolverton. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Physical Review B*, 96(2):1–12, 2017.
- [178] Manfred K. Warmuth, Jun Liao, Gunnar Rätsch, Michael Mathieson, Santosh Putta, and Christian Lemmen. Active learning with support vector machines in the drug discovery process. *Journal of Chemical Information and Computer Sciences*, 43(2):667–673, 2003.
- [179] Jess Wellendorff, Keld T. Lundgaard, Andreas Møgelhøj, Vivien Petzold, David D. Landis, Jens K. Nørskov, Thomas Bligaard, and Karsten W. Jacobsen. Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Physical Review B - Condensed Matter and Materials Physics*, 85(23):32–34, 2012.
- [180] Colin White, Willie Neiswanger, and Yash Savani. Bananas: Bayesian optimization with neural architectures for neural architecture search, 2020.
- [181] Greg Wilson, D. A. Aruliah, C. Titus Brown, Neil P. Chue Hong, Matt Davis, Richard T. Guy, Steven H.D. Haddock, Kathryn D. Huff, Ian M. Mitchell, Mark D. Plumbley, Ben Waugh, Ethan P. White, and Paul Wilson. Best Practices for Scientific Computing. *PLoS Biology*, 12(1), 2014.
- [182] Hai Xiao, Tao Cheng, and William A. Goddard. Atomistic Mechanisms Underlying Selectivities in C₁ and C₂ Products from Electrochemical Reduction

- of CO on Cu(111). *Journal of the American Chemical Society*, 139(1):130–136, 2017.
- [183] Hai Xiao, William A. Goddard, Tao Cheng, and Yuanyue Liu. Cu metal embedded in oxidized matrix catalyst to promote CO₂ activation and CO dimerization for electrochemical reduction of CO₂. *Proceedings of the National Academy of Sciences*, 2017.
- [184] Tian Xie and Jeffrey C. Grossman. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters*, 120(14):145301, 2018.
- [185] Miho Yamauchi, Ryu Abe, Tatsuya Tsukuda, Kenichi Kato, and Masaki Takata. Highly selective ammonia synthesis from nitrate with photocatalytically generated hydrogen on CuPd/TiO₂. *Journal of the American Chemical Society*, 133(5):1150–1152, 2011.
- [186] H. Yano, T. Tanaka, M. Nakayama, and K. Ogura. Selective electrochemical reduction of CO₂ to ethylene at a three-phase interface on copper(I) halide-confined Cu-mesh electrodes in acidic solutions of potassium halides. *Journal of Electroanalytical Chemistry*, 565(2):287–293, 2004.
- [187] Chris Ying, Aaron Klein, Esteban Real, Eric Christiansen, Kevin Murphy, and Frank Hutter. NAS-BENCH-101: Towards reproducible neural architecture search. In *36th International Conference on Machine Learning*, 2019.
- [188] Andrea Zanette, Junzi Zhang, and Mykel J. Kochenderfer. Robust super-level set estimation using gaussian processes. *Lecture Notes in Computer Science*, 11052 LNAI:276–291, 2019.
- [189] Zhenhua Zeng, Kee Chul Chang, Joseph Kubal, Nenad M. Markovic, and Jeffrey Greeley. Stabilization of ultrathin (hydroxy)oxide films on transition metal substrates for electrochemical energy conversion. *Nature Energy*, 2(May):1–9, 2017.
- [190] Ying Zhang and Chen Ling. A strategy to apply machine learning to small datasets in materials science. *npj Computational Materials*, 25(December 2017):28–33, 2018.
- [191] Miao Zhong, Kevin Tran, Yimeng Min, Chuanhao Wang, Ziyun Wang, Cao-Thang Dinh, Phil De Luna, Zongqian Yu, Armin Sedighian Rasouli, Peter Brodersen, Song Sun, Oleksandr Voznyy, Chih-Shan Tan, Mikhail Askerka, Fanglin Che, Min Liu, Ali Seifitokaldani, Yuanjie Pang, Shen-Chuan Lo, Alexander Ip, Zachary Ulissi, and Edward H. Sargent. Accelerated discovery of CO₂ electrocatalysts using active machine learning. *Nature*, 581(7807):178–183, 2020.

Appendix A

Supplementary Information for Chapter 3

This work originally appeared as the Supplementary Information for: Kevin Tran and Zachary W. Ulissi. Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution. Nature Catalysis, 1:696–703, 2018.

A.1 Supplementary Figures

1																	18
H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Lr	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Nh	Fl	Mc	Lv	Ts	Og

Figure A-1: **Elements considered in this screening.** Shading of an element indicates its inclusion in this screening study.

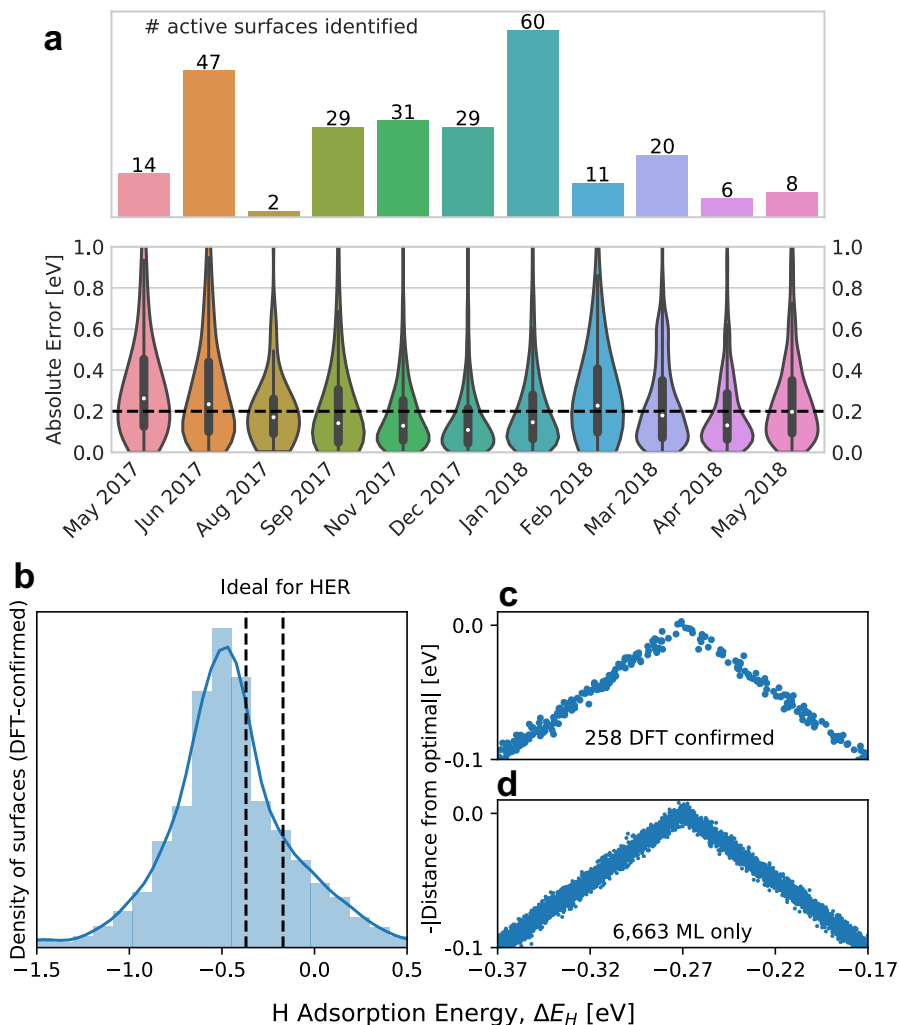


Figure A-2: **Identification of surfaces with near-optimal ΔE_H values for HER.** **a**, Distribution of the number of near-optimal surfaces identified. **b**, The normalized distribution of the low-coverage H adsorption energies of all of the surfaces enumerated by this study. Dashed lines indicate the ± 0.1 eV range around the optimal ΔE_H value of -0.27 eV. **c**, Surfaces whose low-coverage H adsorption energies have been calculated and verified with DFT. **d**, Surfaces whose low-coverage H adsorption energies have been calculated only by the machine learning models.

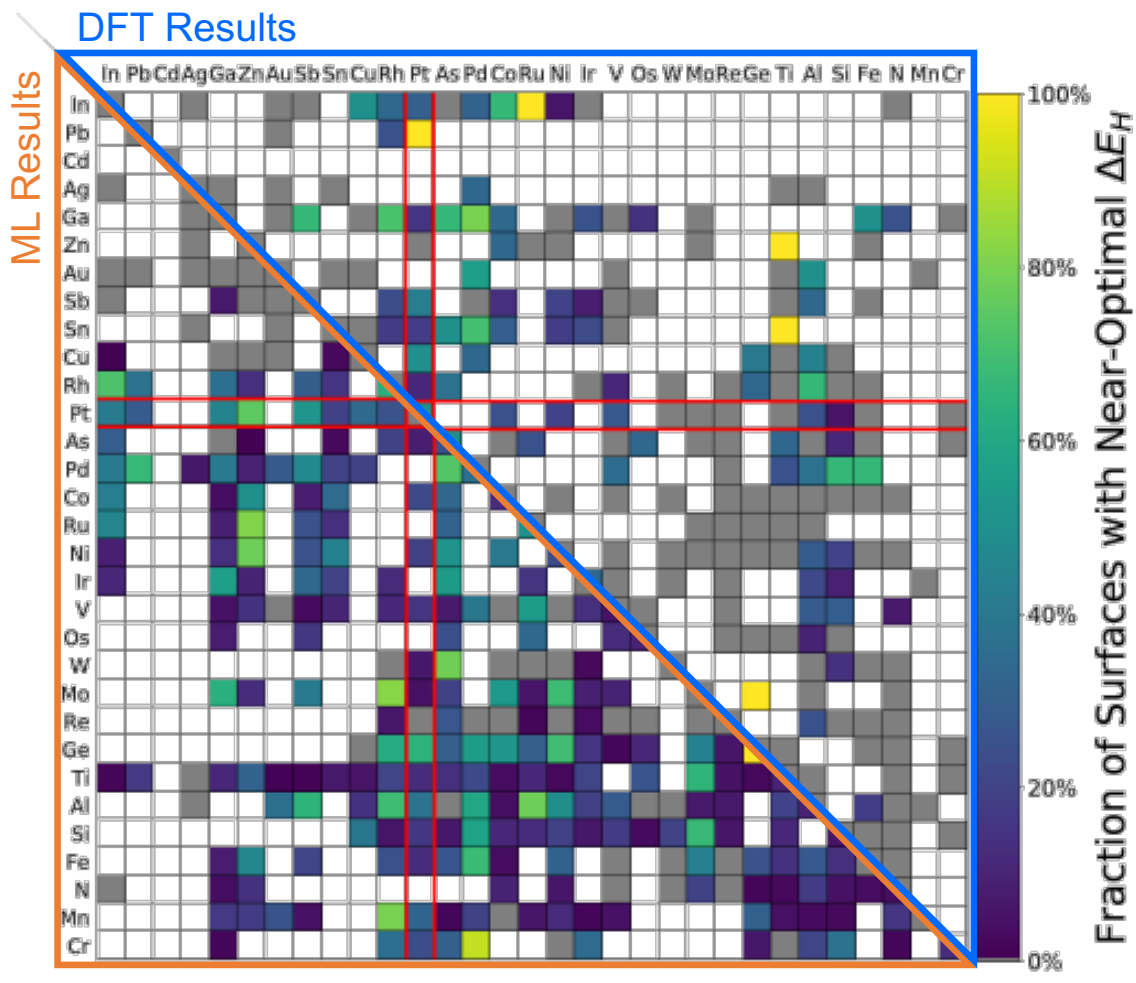


Figure A-3: **H₂ evolution efficiency map for bimetallics.** Visualization of two component intermetallics whose surfaces have low-coverage H adsorption energy (ΔE_H) values inside the range of $[-0.37, -0.17]$ eV. White shading indicates an absence of any enumerated surfaces; grey shading indicates that all ΔE_H values are outside the range of $[-0.37, -0.17]$ eV; and colored shading indicates possible efficiency. The ΔE_{CO} values used to create the upper half of this figure were calculated by DFT, and the values used to create the bottom half were calculated by the surrogate model.

A.2 Supplementary Notes

A.2.1 Supplementary Note 1

We performed a coarse study of the TPOT[135] hyperparameters and how they effect model performance. Details can be found on the Github website mentioned in the main text’s Code Availability section. This coarse study yielded the following

settings for using TPOT: 1 generation, a population size of 16, and an offspring size of 16. A generation size of 1 was used because the coarse study showed that the error did not change as a function of the number of generations. The study also showed that population and offspring sizes of 16 were sufficient and that changes in population/offspring were overshadowed by stochastic variations inherent in TPOT’s genetic algorithm. Note that we used a median absolute deviation scoring function to allow the model to fit the majority of the data well at the expense of more severe outliers, and we used TPOT’s default setting of 5-fold cross validation.

We also performed a cursory hyperparameter optimization test to verify whether or not TPOT’s hyperparameter tuning worked well, because we were concerned that the resolution of hyperparameter changes that TPOT examined were too coarse to perform an optimization properly. We performed this test by comparing TPOT’s hyperparameter tuning with SigOpt’s tuning for one example pipeline that TPOT has chosen—Random Forest Regression. Using a 90/10 train/test split for predicting ΔE_{CO} , TPOT’s choice of hyperparameters yielded a test RMSE, MAE, and MAD of 0.32, 0.18, and 0.10 eV, respectively. SigOpt yielded a test RMSE, MAE, and MAD of 0.33, 0.19, and 0.10 eV, respectively. These results suggest that TPOT’s hyperparameter tuning is sufficient for our use-case.

A.2.2 Supplementary Note 2

Static surrogate models are typically judged via train/test errors and parity plots. These were omitted from the main manuscript in favor of evaluations that focus on dynamic, iterative workflows—i.e., evaluation on a rolling forecast origin.[69] We still show the classical evaluations here for the sake of completeness.

We took the 19,644 DFT-calculated ΔE_{CO} values that we have and performed 16 different stratified 90/10 train/test splits,[166] and then we used TPOT create 16 different models from these splits. We repeated this for our 23,141 ΔE_{H} calculations, resulting in 32 total models. Of the 32 models, 19 of them were K-nearest-neighbor models, 3 were gradient boosting regressors, and 10 were random forest or extra trees regressors. The subsequent distributions of root-mean-squared-error (RMSE),

mean absolute error (MAE), and median absolute deviation (MAD) are shown in Supplementary Figures A-4a and b for ΔE_{CO} and ΔE_{H} , respectively.

We then used TPOT to train final models on 100% of our DFT data—i.e., 19,644 for ΔE_{CO} and 23,141 for ΔE_{H} . When predicting ΔE_{CO} , TPOT identified an extra trees regressor with a 5-fold cross-validation (CV) MAD of 0.09 eV. This regressor was created with a maximum feature splitting proportion of 0.95, a minimum of 5 samples per leaf, a minimum of 6 samples per split, 100 trees, and is fit to minimize RMSE. When predicting ΔE_{H} , TPOT identified a random forest regressor with a 5-fold CV MAD of 0.07 eV. This regressor was created with a maximum feature splitting proportion of 0.65, a minimum of 11 samples per leaf, a minimum of 18 samples per split, 100 trees, and is fit to minimize RMSE. We then converted the final models’ adsorption energy predictions to performance metrics using scaling relations.[53, 105] The subsequent parity plots of catalyst performance are shown in Supplementary Figures A-4c and d for CO_2RR and HER, respectively. Darker shading indicates a higher density of points. Lastly: We created parity plots for this model’s predictions of ΔE_{CO} and ΔE_{H} in Supplementary Figures A-4e and f, respectively.

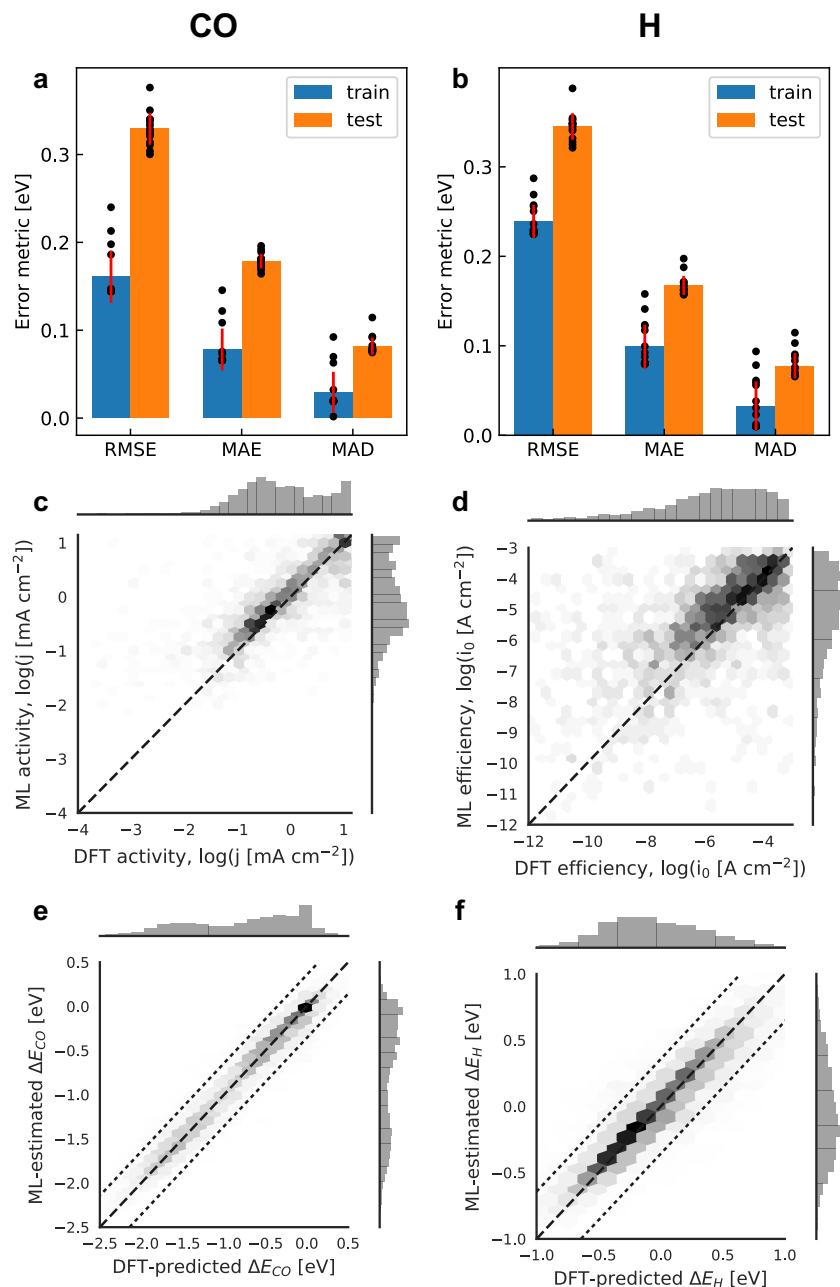


Figure A-4: **Surrogate modeling performance of final model.** Performance metrics for 16 models created from 90/10 train/test splits on all DFT data for predicting (a) ΔE_{CO} ($n=19,644$) and (b) ΔE_{H} ($n=23,141$). Black points represent results for each of the 16 models, bar heights indicate average values, and red lines indicate average values ± 1 standard deviation. Parity plots when predicting catalyst performance for (c) CO_2RR and (d) HER . Parity plots when predicting (e) ΔE_{CO} and (f) ΔE_{H} . Darker shading indicates a higher density of points; dashed lines indicate parity; and dotted lines indicate 95% prediction intervals—i.e., $\text{parity} \pm 2 \times \text{MAE}$. The parity plots in (c-f) were generated using a model that was trained using a 100/0 train/test split.

It is common with surrogate modeling to calculate the modeling error as a function of training set size—i.e., a learning curve. Learning curves give the modeler a quantitative estimate for how much training data was actually necessary to converge on a final model. We do not have a final model because we are not yet finished exploring (or even defining) our search space, and so any learning curve we make may not illustrate properly the amount of data needed to converge. We can still show retrospective learning curves as if we treated our current DFT data set as if it was our entire search space—Supplementary Figure A-5. These figures were created by fitting 20 different models using varying sizes of training data that were randomly selected from our database of DFT results. Test errors were calculated from the remaining data—e.g., given a training size of 2,000 points from a pool of 19,000, the test size would be 17,000 points.

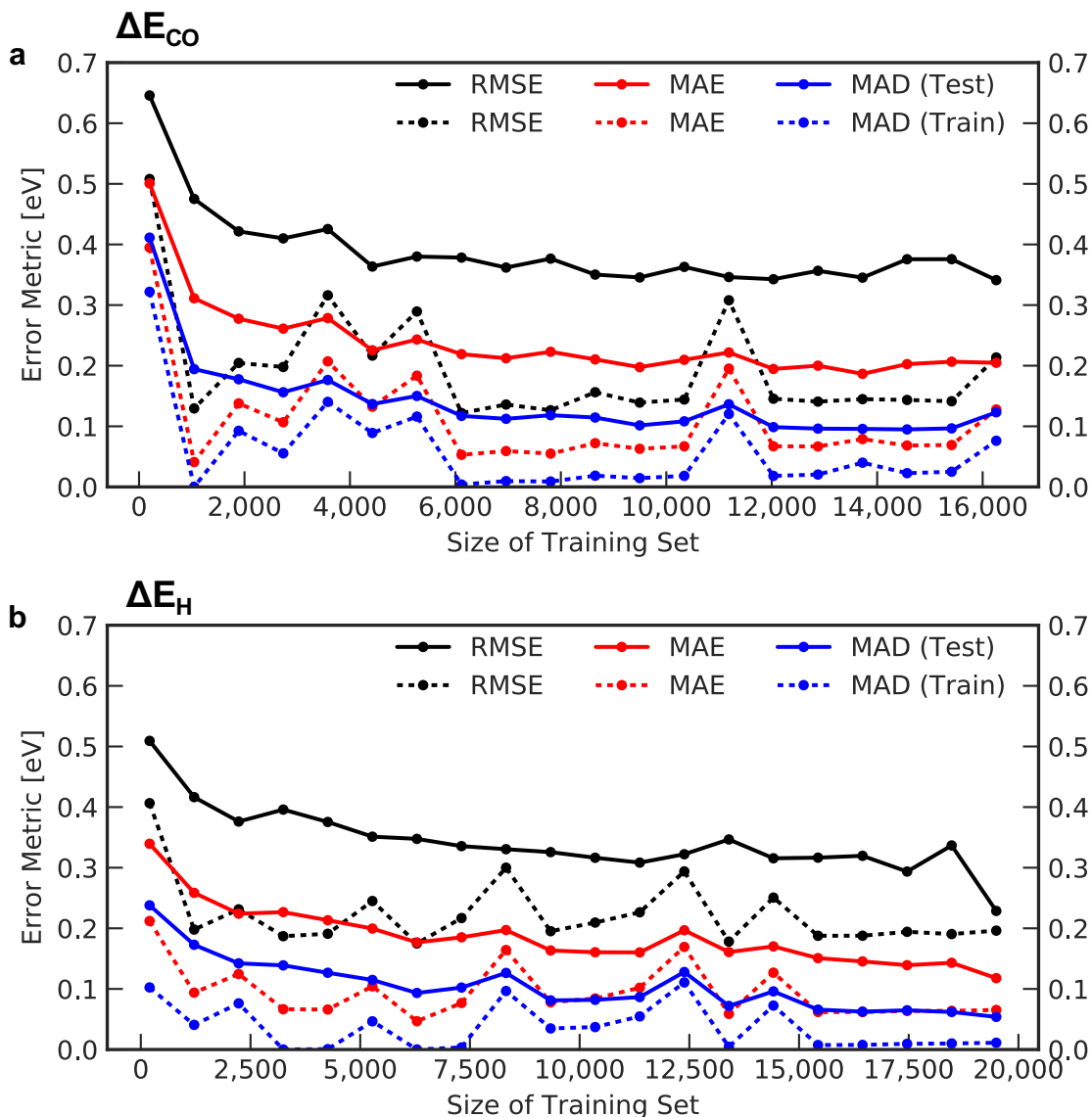


Figure A-5: **Learning curves.** Error metrics vs. size of training set when calculating adsorption energies for (a) CO and (b) H. RMSE is the root mean squared error; MAE is the mean absolute error; and MAD is the median absolute deviation.

A.2.3 Supplementary Note 3

To verify the integrity of our 42,785 DFT calculations, we compared a subset of our results for pure metal calculations on simple facets with RPBE calculations found in literature,[7] including sources[164] that used the BEEF-vdW functional (Supplementary Figure A-6). Some of the differences between our results and literature

results may be caused by our frameworks' neglect in accounting for spin magnetism for magnetic elements like Fe and Ni, in contrast to other literature methods.

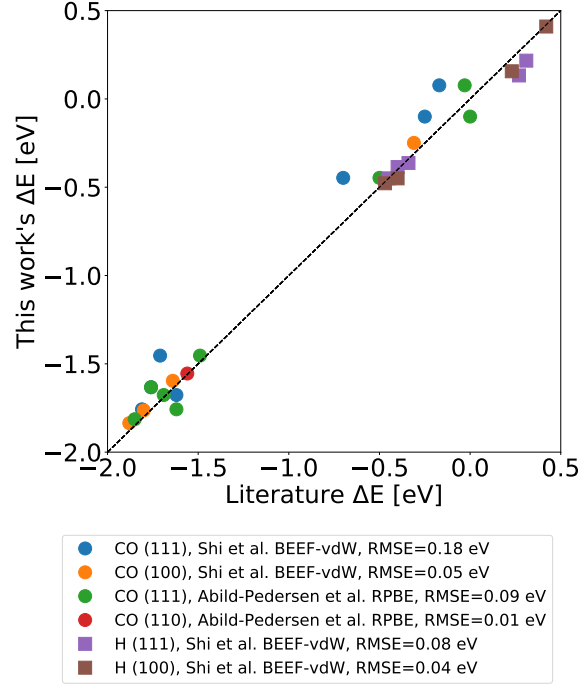


Figure A-6: **Benchmarking of DFT results.** The adsorption energies (ΔE) of various adsorbate/surface combinations calculated by our framework vs. the energies calculated by various literature sources.[7, 164] The diagonal line indicates parity.

A.3 Supplementary Methods

A.3.1 Calculation of free energies

Supplementary Equation A.1 can be used to calculate the change in chemical potential (ΔG_X) during adsorption of arbitrary adsorbate X , [134] which is the descriptor that this study uses to predict catalyst performance [53, 105]. μ_{X*} is the chemical potential of the adsorbate bound on a surface, μ_* is the chemical potential of the surface, and μ_X is the chemical potential of the adsorbate in gas phase.

$$\Delta G_X = \mu_{X*} - \mu_* - \mu_X \quad (\text{A.1})$$

Chemical potentials can be calculated[134, 105] using Supplementary Equation A.2, where E is the electronic energy as calculated by density functional theory (DFT), ZPE is the zero-point energy, C_p is the heat capacity, T is temperature, S is entropy, $\Delta\mu_{solv}$ is the change in chemical potential from solvent [de]stabilization, and $\Delta\mu_{corr}$ is any experimental correction required to account for differences between experimental chemical potentials and DFT-based chemical potentials.

$$\mu = E + ZPE + \int C_p dT - TS + \Delta\mu_{solv} + \Delta\mu_{corr} \quad (\text{A.2})$$

Literature[165] reports that ZPE_{CO} is 0.130 eV, $\int C_{p,CO} dT$ is 0.091 eV at 298 K, S_{CO} is 0.002092 eV K⁻¹, and that $\Delta\mu_{exp,CO}$ is 0.02 eV for gas phase CO. Thus the chemical potential of CO in the gas phase is:

$$\begin{aligned} \mu_{CO,g} &= E_{CO,g} + ZPE_{CO,g} + \int C_{p,CO} dT - TS_{CO,g} + \Delta\mu_{solv,g} + \Delta\mu_{corr,g} \\ \mu_{CO,g} &= E_{CO,g} + 0.130 \text{ eV} + 0.091 \text{ eV} - (298 \cdot 0.002092) \text{ eV} + 0 \text{ eV} + 0.02 \text{ eV} \\ \mu_{CO,g} &= E_{CO,g} - 0.362 \text{ eV} \end{aligned} \quad (\text{A.3})$$

The same method can be used to calculate the chemical potential of CO in the adsorbed-state (μ_{CO*}) given literature values[165] for ZPE , C_p , and S as well as a solvation correction term[105] (μ_{solv}) and an empirical correction correction term[7] (μ_{corr})—see Supplementary Equation A.4. Note that we assumed a CO stretch frequency of 2,000 cm⁻¹, and deviations from this frequency may cause errors.

$$\begin{aligned} \mu_{CO*} &= E_{CO*} + ZPE_{CO*} + \int C_p dT - TS_{CO*} + \Delta\mu_{solv} + \Delta\mu_{corr} \\ \mu_{CO*} &= E_{CO*} + 0.192 \text{ eV} + 0.085 \text{ eV} - (298 \cdot 0.000452) \text{ eV} - 0.2 \text{ eV} + 0.2 \text{ eV} \\ \mu_{CO*} &= E_{CO*} + 0.142 \text{ eV} \end{aligned} \quad (\text{A.4})$$

Assuming that μ_* is equal to E_* and by combining Supplementary Equations A.1, A.3, and A.4, we can calculate the change in chemical potential of adsorbing CO:

$$\begin{aligned}
\Delta G_{CO} &= \mu_{CO*} - \mu_* - \mu_{CO} \\
\Delta G_{CO} &= [E_{CO*} + 0.142 \text{ eV}] - E_* - [E_{CO,g} - 0.362 \text{ eV}] \\
\Delta G_{CO} &= E_{CO*} - E_* - E_{CO,g} + 0.50 \text{ eV} \\
\Delta G_{CO} &= \Delta E_{CO} + 0.50 \text{ eV}
\end{aligned} \tag{A.5}$$

where the change in electronic energy, ΔE_{CO} , is defined using Supplementary Equation A.6 and is calculated with DFT.[134]

$$\Delta E_X = E_{X*} - E_* - E_{X,g} \tag{A.6}$$

We combined Supplementary Equation A.5 with the optimal ΔG_{CO} found in literature,[105] -0.17 eV, to calculate the optimal ΔE_{CO} : -0.67 eV. Note that the optimal ΔG_{CO} of -0.17 eV was based on microkinetic modeling of the reaction pathway from CO₂ to methane on single metal surfaces where the rate-limiting step was hydrogenation of CO, which is why the binding energy of CO is used as a descriptor of activity. By using this descriptor, we assume that CO is still involved in the rate-limiting step of CO₂ reduction. This is acceptable because our framework is used only as a screening method. If we want to find a catalyst for a reaction where CO is not a key intermediate in a rate-limiting step, then we can use our framework to model the adsorption of whichever intermediate is appropriate. Note also that the source of the CO₂RR scaling relationship used the BEEF functional[179] while our work used the RPBE functional,[57] which may impart errors in our estimates of activity.

In contrast to the methods we used to find the optimal ΔE_{CO} for CO₂ reduction, the literature sources[133, 53] we used for the HER already performed the chemical potential calculations, allowing us to use their energy calculation methods directly:

$$\Delta G_H = \Delta E_H + 0.24 \text{ eV} \tag{A.7}$$

where the optimal ΔG_H value is -0.03 eV, yielding an optimal ΔE_H value of -0.27 eV.

A.3.2 Previous surrogate modeling methods

We used various methods for fingerprinting adsorption sites and various regression methods during the development of our workflow. The primary method used during development involved both a different fingerprint and a different regression method than the one described in the main text. This alternative fingerprint was composed of a vector of coordination numbers for each element and an array of neighbors' coordination numbers as determined by Voronoi tessellation performed by pymatgen.[136] Supplementary Figure A-7a illustrates a simplified example of the fingerprint. The coordination vector contains one item for each of the 31 elements considered in the screening, and each item in this vector is the sum of the number of atoms of that element that are coordinated with the adsorbate. The neighbors' coordination vector is a flattened array that contains 31×31 items for each elemental pairing, and each item in this vector is the sum of the number of atoms of one element that are coordinated with the all adsorbate neighbors that belong to another element. Algorithm 1 outlines explicitly how we calculated the neighbors' coordination number.

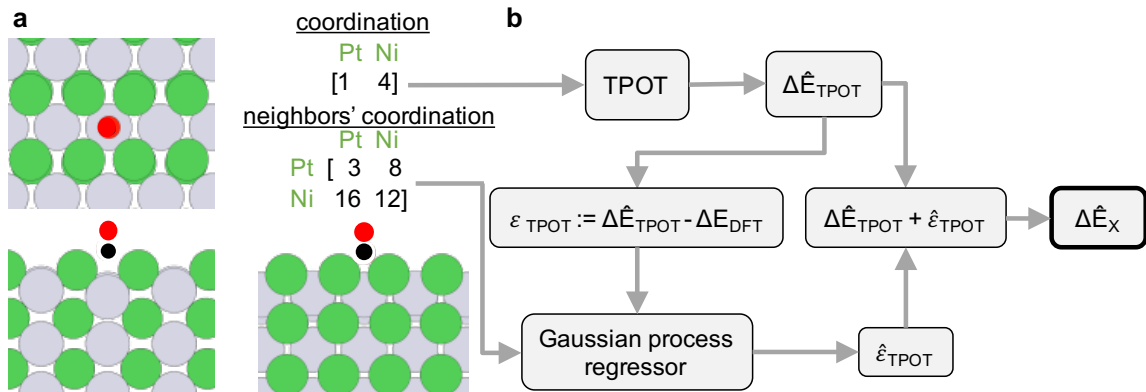


Figure A-7: **Alternative method for modeling adsorption energy.** **a**, Alternative fingerprinting method. Each item in the “coordination” vector represents the coordination number for a particular element, e.g., Pt or Ni. Each item in the “neighbors’ coordination” array represents this same coordination vector, but for each of the adsorbate’s neighbors. Note that this example is for illustrative purposes only. The vectors and arrays actually used contained enough items to represent 31 different elements. **b**, Old method for performing regression. Coordination was used by TPOT to estimate adsorption energy, then the residuals of that model were combined with the neighbors’ coordinations and used by a Gaussian process regressor to estimate residuals of the first model. Summation of the results of the first and second models yielded the final estimate.

Algorithm 1 Calculating array of neighbors’ coordination numbers for an adsorption site

- 1: $ncn := \text{zeros}(n, n)$ ▷ where $n=31$, the total number of elements we are investigating
 - 2: **for** all neighbors **do**
 - 3: $i :=$ index of the element of neighbor
 - 4: **for** all neighbor’s neighbors **do**
 - 5: $j :=$ index of the element of neighbor’s neighbor
 - 6: $ncn_{i,j} := ncn_{i,j} + 1$
-

The alternative regression method used these coordination and neighbors’ coordination fingerprints in a multi-staged regression approach to predict adsorption energies (Supplementary Figure A-7b). In the first stage, we used TPOT[135] to regress the DFT-calculated adsorption energies (ΔE_{DFT}) against the vector of coordination numbers. In the second layer, we used SKLearn’s[138] Gaussian processor

to regress the residuals of the first stage’s estimates ($\hat{\epsilon}_{TPO}$) against the neighbors’ coordination numbers. The dimensionality of the neighbors’ coordination numbers was reduced by using principal component analysis as implemented in SKLearn. We then added the estimates of the models from both stages to yield a final estimate for adsorption energy ($\Delta\hat{E}_{X*}$).

These alternative fingerprinting and regression methods were eventually forgone in favor of methods that yielded lower test errors. Test set root mean squared error, mean absolute error, and median absolute deviation values were approximately 0.91, 0.23, and 0.14 eV, respectively. A stratified, 90/10 train/test split was used to build the model that yielded these errors.

A.3.3 t-SNE methods

We then used the t-SNE algorithm[113] to visualize our data by reducing all of the DFT-modeled adsorption sites into a 2-dimensional representation. We did so by taking the fingerprints illustrated in Supplementary Figure A-7 and then scaling them such that each feature had a mean of zero and a variance of one, and then sending the fingerprints through SKLearn’s[138] principle component analyzer where the dimensionality was reduced until only 85% of the variance was accounted for, yielding a vector of 113 items instead of 31×32 items. This reduced vector was processed by the t-SNE algorithm[113] with a perplexity of 50 and a learning rate of (ϵ) of 750 and then stopped after 3,000 iterations, yielding a 2-dimensional reduction of the adsorption sites. Note that the fingerprint in Supplementary Figure A-7a was still used to create the t-SNE diagrams instead of the fingerprint in Figure 2, because Supplementary Figure A-7’s fingerprint codified adsorption sites only by the identities and the numbers of neighboring atoms. Thus the sites in the reduced space were described only by elemental identities and coordination numbers, which we assumed to be more representative of our controllable materials design space.

A.4 Supplementary Tables

When using adsorption energies as descriptors of catalyst performance, it is common to characterize an entire surface by calculating the adsorption energies of all the adsorption sites on that surface and then assuming that the lowest energy site is representative of the entire surface, because that site is the most thermodynamically stable and therefore more likely to bind the adsorbate. Thus we define an “adsorption energy of a surface” as the minimum adsorption energy of all sites on that surface. Using this method, we identified dozens of surfaces with near-optimal adsorption energies. These surfaces are listed in Table A.1 and Table A.2 for CO₂ reduction and H₂ evolution, respectively. Note that we chose to include surfaces whose energies were below the optimal energy targets by 0.2 eV and above the optimal targets by 0.1 eV, because we assumed that over-binding surfaces were more likely to perform well than under-binding surfaces.

Table A.1: **List of possibly active surfaces for CO₂ reduction.** All of the surfaces whose minimum CO adsorption energies were between -0.87 and -0.57 eV. MPID represents the Materials Project ID. Apparently duplicate entries for MPID/Miller index combinations are caused by different, nonidentical cuts along a Miller plane, which yields multiple surfaces for a single Miller index. All energies reported here were calculated by DFT, not with machine learning.

Formula	Miller index	ΔE_{CO} [eV]	MPID
Ag ₃ Pd	[2, 0, 1]	-0.84	mp-985296
Al ₂ Cu	[1, 1, 0]	-0.66	mp-985806
Al ₂ Cu	[1, 1, 1]	-0.75	mp-985806
Al ₂ Cu	[1, 1, 1]	-0.73	mp-985806
Al ₂ Cu	[2, 1, 0]	-0.75	mp-985806
Al ₂ Cu	[2, 1, 0]	-0.68	mp-985806
Al ₂ Cu	[2, 2, 1]	-0.73	mp-985806
Al ₂ Cu ₆	[0, 0, 1]	-0.79	mp-12802
Al ₂ Cu ₆	[0, 0, 1]	-0.64	mp-12802
Al ₂ Cu ₆	[1, 1, 0]	-0.80	mp-12802
Al ₂ Cu ₆	[1, 1, 2]	-0.84	mp-12802
Al ₂ Cu ₆	[1, 2, 0]	-0.65	mp-12802
Al ₂ Cu ₆	[1, 2, 1]	-0.69	mp-12802
Al ₂ Cu ₆	[1, 2, 2]	-0.78	mp-12802
Al ₂ Cu ₆	[2, 1, 0]	-0.70	mp-12802
Al ₂ Cu ₆	[2, 2, 1]	-0.76	mp-12802
Al ₃ Cu ₂	[1, 1, 1]	-0.83	mp-10886
Al ₃ Cu ₂	[1, 1, 1]	-0.82	mp-10886
Al ₃ Cu ₂	[2, 0, 1]	-0.82	mp-10886
Al ₅ Cu ₅	[1, 1, -2]	-0.83	mp-2500
Al ₆ Cu ₂	[1, 0, 1]	-0.58	mp-985825
Al ₆ Cu ₂	[2, 1, 0]	-0.83	mp-985825

Continued on next page...

Table A.1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
AlCu	[1, 1, 1]	-0.80	mp-1022721
AlCu3	[1, 1, 0]	-0.76	mp-1008555
AlCu3	[1, 1, 0]	-0.72	mp-1008555
AlCu3	[1, 1, 1]	-0.84	mp-12777
AlCu3	[2, 1, 0]	-0.78	mp-1008555
AlCu3	[2, 1, 0]	-0.75	mp-1008555
Al2N2	[2, 1, 2]	-0.66	mp-661
Al2Pd	[2, 2, 1]	-0.85	mp-16522
Al6Pd10	[1, 0, 0]	-0.78	mp-16523
AlPd	[1, 0, 0]	-0.85	mp-829
Al2Pt	[1, 0, 0]	-0.72	mp-1502
Al2Ti2	[0, 0, 1]	-0.60	mp-1953
Al4Ti2	[0, 0, 1]	-0.82	mp-11809
Al12W	[1, 0, 0]	-0.77	mp-11227
AsGa	[1, 0, 0]	-0.81	mp-2534
AsGa	[2, 1, 1]	-0.76	mp-2534
As4Si3	[1, 1, 1]	-0.87	mp-570744
As6Si6	[2, 2, -1]	-0.67	mp-1863
As4Zn4Cu4	[1, 0, 1]	-0.66	mp-676828
As4Zn4Cu4	[1, 2, 1]	-0.75	mp-676828
As4Zn4Cu4	[2, 1, 0]	-0.75	mp-676828
As4Zn4Cu4	[2, 1, 2]	-0.79	mp-676828
Au2Al	[0, 0, 1]	-0.73	mp-1018179
Au2Al	[1, 0, 2]	-0.74	mp-1018179
Au4Al4	[0, 0, 1]	-0.62	mp-1399
Au2AlCu	[1, 1, 1]	-0.73	mp-867306

Continued on next page...

Table A.1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
Au ₂ AlCu	[2, 1, 0]	-0.79	mp-867306
Au ₂ Cu ₂	[0, 0, 1]	-0.67	mp-522
Au ₂ Cu ₂	[1, 0, 0]	-0.69	mp-522
Au ₂ Cu ₂	[2, 1, 0]	-0.84	mp-522
Au ₃ Cu	[2, 1, 0]	-0.76	mp-2103
Au ₃ Cu	[2, 1, 0]	-0.61	mp-2103
Au ₃ Cu	[2, 1, 1]	-0.64	mp-2103
Au ₃ Cu	[2, 2, 1]	-0.71	mp-2103
AuCu ₃	[1, 1, 0]	-0.68	mp-2258
AuCu ₃	[1, 1, 0]	-0.63	mp-2258
AuCu ₃	[2, 1, 0]	-0.80	mp-2258
AuCu ₃	[2, 1, 0]	-0.79	mp-2258
AuCu ₃	[2, 1, 1]	-0.75	mp-2258
AuGa ₂	[1, 0, 0]	-0.57	mp-2776
AuPd ₂ In	[2, 2, 1]	-0.79	mp-863724
Au ₅ Sn	[1, 1, 0]	-0.64	mp-30418
Au ₄ Ti	[1, 1, 0]	-0.58	mp-12635
Co ₈ Al ₂₀	[0, 0, 1]	-0.69	mp-196
Cu	[1, 1, 0]	-0.63	mp-30
Cu	[2, 1, 0]	-0.78	mp-30
Cu	[2, 1, 1]	-0.69	mp-30
CuAlPt ₂	[1, 1, 0]	-0.77	mp-12550
Cu ₂₈ In ₁₂	[0, 0, 1]	-0.84	mp-21985
Fe ₂ Al ₁₂	[1, 1, 0]	-0.69	mp-570001
FeAl	[1, 0, 0]	-0.60	mp-2658
Ga ₂ Cu	[0, 0, 1]	-0.78	mp-11359

Continued on next page...

Table A.1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
Ga ₂ Cu	[1, 0, 0]	-0.64	mp-11359
Ga ₂ Cu	[1, 0, 1]	-0.67	mp-11359
Ga ₂ Cu	[1, 1, 1]	-0.66	mp-11359
Ga ₂ Cu	[1, 1, 2]	-0.74	mp-11359
Ga ₂ Cu	[2, 0, 1]	-0.72	mp-11359
Ga ₂ Cu	[2, 1, 0]	-0.82	mp-11359
Ga ₂ Cu ₆	[1, 0, 0]	-0.79	mp-865798
Ga ₂ Cu ₆	[1, 0, 0]	-0.69	mp-865798
Ga ₂ Cu ₆	[1, 0, 1]	-0.75	mp-865798
Ga ₂ Cu ₆	[1, 1, 0]	-0.80	mp-865798
Ga ₂ Cu ₆	[1, 1, 1]	-0.72	mp-865798
Ga ₂ Cu ₆	[2, 2, 1]	-0.84	mp-865798
Ga ₁₀ Pd ₂	[1, 1, 2]	-0.66	mp-30660
Ga ₆ Pd ₁₀	[1, 0, 0]	-0.79	mp-2408
Ga ₂ Pt	[1, 0, 0]	-0.64	mp-22095
Ga ₂ Pt	[1, 1, 0]	-0.65	mp-22095
Ga ₂ Pt	[2, 1, 0]	-0.70	mp-22095
Ga ₃ Pt ₂	[1, 1, 0]	-0.65	mp-21400
Ga ₃ Pt ₂	[2, 1, 1]	-0.68	mp-21400
Ga ₃ Pt ₂	[2, 2, 1]	-0.81	mp-21400
GaPt ₂ Cu	[1, 1, 0]	-0.74	mp-644280
Ge ₂	[2, 1, 1]	-0.75	mp-32
Ge ₄ Co ₄ Cu ₂	[1, 0, 0]	-0.71	mp-19955
Ge ₂ Cu ₆	[1, 0, 2]	-0.71	mp-19724
Ge ₂ Cu ₆	[1, 2, 1]	-0.75	mp-19724
Ge ₂ Cu ₆	[2, 1, 0]	-0.72	mp-19724

Continued on next page...

Table A.1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
Ge4Fe4Cu2	[1, 0, 0]	-0.70	mp-21141
H6Al2	[2, 1, 1]	-0.74	mp-23933
H2Si2	[1, 0, 0]	-0.77	mp-29803
H2Si2	[2, 0, 1]	-0.85	mp-29803
H2Ti	[0, 0, 1]	-0.68	mp-24726
H2Ti	[1, 0, 0]	-0.64	mp-24161
H2Ti	[1, 1, 0]	-0.63	mp-24726
Ir4In12	[0, 2, 1]	-0.86	mp-636498
IrRhGa2	[1, 0, 0]	-0.81	mp-865743
Mn2Al2	[0, 0, 1]	-0.74	mp-771
Mn3GaN	[1, 0, 0]	-0.85	mp-627439
N4	[1, 1, 0]	-0.84	mp-999498
NIn	[1, 0, 0]	-0.76	mp-20411
N2V2	[1, 1, 1]	-0.74	mp-1017532
N2V2	[2, 1, 0]	-0.77	mp-1017532
Ni4Al12	[0, 0, 1]	-0.64	mp-622209
NiAl	[1, 0, 0]	-0.69	mp-1487
NiFeAl2	[1, 0, 0]	-0.73	mp-867330
NiGa	[1, 1, 1]	-0.75	mp-1941
Ni4Si4	[0, 1, 0]	-0.70	mp-351
NiZnCu2	[1, 0, 0]	-0.68	mp-30593
NiZnCu2	[1, 1, 0]	-0.87	mp-30593
OsAl	[1, 0, 0]	-0.73	mp-875
Pd4Cu16	[0, 0, 1]	-0.71	mp-30594
Pd4Cu16	[2, 1, 0]	-0.82	mp-30594
PdCu	[1, 0, 0]	-0.66	mp-1018029

Continued on next page...

Table A.1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
PdCu3	[1, 0, 0]	-0.66	mp-580357
PdCu3	[1, 1, 0]	-0.76	mp-580357
PdCu3	[1, 1, 1]	-0.82	mp-580357
PdCu3	[2, 2, 1]	-0.84	mp-580357
Pd2CuIn	[1, 0, 0]	-0.79	mp-867308
PtCu	[0, 0, 1]	-0.72	mp-644311
PtCu3	[1, 0, 0]	-0.69	mp-12086
PtCu3	[1, 1, 0]	-0.79	mp-12086
PtCu3	[2, 2, 1]	-0.76	mp-12086
PtIn2	[2, 1, 0]	-0.74	mp-22682
Re2Al12	[0, 0, 1]	-0.66	mp-16528
Re2Al12	[1, 1, 0]	-0.70	mp-16528
ReAl12	[1, 0, 0]	-0.80	mp-1648
RuCoGa2	[1, 0, 0]	-0.69	mp-865779
RuIrGa2	[1, 0, 0]	-0.65	mp-866041
Sb2Al2	[1, 0, 2]	-0.82	mp-1018100
SbAl	[1, 1, 1]	-0.82	mp-2624
SbAl	[2, 1, 1]	-0.80	mp-2624
SbCuNi2	[1, 0, 0]	-0.66	mp-30069
Sb2Ga2	[2, 0, 1]	-0.76	mp-1018059
SbGa	[1, 1, 1]	-0.57	mp-1156
Sb2Pd2	[1, 1, 1]	-0.57	mp-1769
Sb2Pd2	[2, 2, 1]	-0.58	mp-1769
Sb8Pd4	[2, 1, 0]	-0.71	mp-1356
Sb2Pt2	[1, 0, 1]	-0.61	mp-2845
Sb2Pt2	[1, 1, 0]	-0.67	mp-2845

Continued on next page...

Table A.1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
Sb2Pt2	[2, -1, 2]	-0.61	mp-2845
Sb2Pt2	[2, 1, 0]	-0.68	mp-2845
Si2Cu6	[1, 0, 0]	-0.60	mp-867317
Si2Cu6	[1, 0, 2]	-0.73	mp-867317
Si2Cu6	[1, 0, 2]	-0.73	mp-867317
Si2Cu6	[1, 1, 0]	-0.78	mp-867317
Si2Cu6	[1, 1, 1]	-0.79	mp-867317
Si2Cu6	[2, 0, 1]	-0.83	mp-867317
Si2Cu6	[2, 1, 1]	-0.71	mp-867317
SiCu3	[0, 0, 1]	-0.65	mp-972828
SiCu3	[1, 0, 0]	-0.69	mp-972828
SiCu3	[2, 0, 1]	-0.82	mp-972828
SiCu3	[2, 1, 0]	-0.77	mp-972828
SiCu3	[2, 1, 1]	-0.81	mp-972828
SiCu3	[2, 1, 1]	-0.79	mp-972828
Si16Fe8	[0, 1, 0]	-0.86	mp-1714
Si4Pd4	[0, 1, 1]	-0.83	mp-389
Si4Pd4	[1, 0, 0]	-0.63	mp-389
Si4Pt4	[1, 0, 1]	-0.80	mp-696
Si4Pt4	[1, 2, 0]	-0.67	mp-696
Si4Rh4	[0, 1, 0]	-0.83	mp-818
Sn10Cu12	[0, 1, 0]	-0.64	mp-1233
Sn10Cu12	[1, 0, 0]	-0.68	mp-1233
Sn10Cu12	[1, 1, -2]	-0.70	mp-1233
Sn16Cu20	[0, 0, 1]	-0.65	mp-845
Sn16Cu20	[1, 0, 0]	-0.63	mp-845

Continued on next page...

Table A.1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
Sn16Cu20	[1, 0, 0]	-0.61	mp-845
Sn2Cu2	[2, 1, 1]	-0.61	mp-10598
Sn12Pd4	[1, 1, 2]	-0.76	mp-1371
Sn12Pd4	[1, 2, 1]	-0.77	mp-1371
Sn12Pd4	[2, 0, 1]	-0.74	mp-1371
Sn16Pd8	[1, 1, 0]	-0.66	mp-1573
Sn4Pd4	[1, 0, 0]	-0.58	mp-2369
Sn2Pt	[1, 1, 0]	-0.69	mp-19962
Sn2Pt2	[1, 0, 1]	-0.75	mp-19856
Sn2Pt2	[2, -1, 2]	-0.74	mp-19856
Sn2Pt2	[2, 1, 0]	-0.85	mp-19856
Zn4Au2Cu2	[0, 1, 2]	-0.62	mp-12759
Zn4Au2Cu2	[1, 1, 1]	-0.68	mp-12759
Zn4Au2Cu2	[1, 2, 1]	-0.63	mp-12759
Zn4Au2Cu2	[1, 2, 2]	-0.79	mp-12759
ZnAu2Cu	[1, 0, 0]	-0.78	mp-864623
ZnAu2Cu	[2, 1, 0]	-0.72	mp-864623
Zn16Cu10	[1, 1, 0]	-0.82	mp-1368
ZnCu	[1, 0, 0]	-0.61	mp-987
ZnCu	[2, 1, 0]	-0.79	mp-987
Zn2Pd2	[2, 0, 1]	-0.58	mp-1652

Table A.2: **List of possibly efficient surfaces for H₂ evolution.** All of the surfaces whose minimum H adsorption energies were between -0.47 and -0.17 eV. MPID represents the Materials Project ID. Apparently duplicate entries for MPID/Miller index combinations are caused by different, nonidentical cuts along a Miller plane, which yields multiple surfaces for a single Miller index. All energies reported here were calculated by DFT, not with machine learning.

Formula	Miller index	ΔE_H [eV]	MPID
Ag3Pd	[1, 0, 0]	-0.24	mp-985296
Al2Cu	[1, 1, 1]	-0.45	mp-985806
Al2Cu	[1, 1, 1]	-0.21	mp-985806
Al2Cu	[2, 1, 0]	-0.34	mp-985806
Al2Cu	[2, 1, 0]	-0.28	mp-985806
Al2Cu	[2, 1, 1]	-0.37	mp-985806
Al2Cu6	[1, 1, 0]	-0.18	mp-12802
Al2Cu6	[1, 1, 1]	-0.21	mp-12802
Al2Cu6	[1, 2, 2]	-0.21	mp-12802
Al3Cu2	[1, 1, 1]	-0.20	mp-10886
Al4Cu2	[1, 0, 1]	-0.22	mp-998
Al4Cu2	[1, 1, 2]	-0.23	mp-998
Al5Cu5	[1, 1, -1]	-0.17	mp-2500
Al5Cu5	[1, 1, -2]	-0.25	mp-2500
Al6Cu2	[1, 1, 1]	-0.22	mp-985825
Al6Cu2	[1, 1, 2]	-0.20	mp-985825
AlCu3	[1, 1, 1]	-0.30	mp-12777
Al2Pd	[2, 1, 0]	-0.37	mp-16522
Al4Pd4	[1, 0, 0]	-0.22	mp-7189
Al4Pd8	[0, 0, 1]	-0.24	mp-2824
Al2Pt	[2, 2, 1]	-0.37	mp-1502
Al3Pt2	[2, 1, 1]	-0.46	mp-10905

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Al4Pt12	[0, 0, 1]	-0.29	mp-607111
Al4Pt12	[1, 0, 1]	-0.40	mp-607111
Al4Pt4	[1, 0, 0]	-0.42	mp-10904
Al4Pt4	[2, 1, 1]	-0.38	mp-10904
Al4Pt4	[2, 1, 1]	-0.34	mp-10904
Al4Pt8	[0, 0, 1]	-0.42	mp-16526
Al4Pt8	[1, 0, 0]	-0.37	mp-16526
Al6Pt10	[1, 0, 0]	-0.34	mp-1501
AlPt3	[1, 0, 0]	-0.38	mp-188
AlPt3	[1, 1, 1]	-0.45	mp-188
Al3V	[1, 1, 0]	-0.46	mp-2554
Al3V	[1, 1, 1]	-0.35	mp-2554
Al3V	[1, 1, 2]	-0.33	mp-2554
Al12W	[1, 1, 1]	-0.47	mp-11227
As2	[1, 1, 1]	-0.21	mp-11
As2	[2, 1, 0]	-0.47	mp-11
As2	[2, 2, 1]	-0.23	mp-11
As4Co4	[1, 0, 0]	-0.43	mp-15679
As4Co4	[1, 1, 0]	-0.45	mp-583
As2Ga2	[2, 2, 1]	-0.19	mp-8883
AsGa	[2, 1, 0]	-0.21	mp-2534
AsIn	[2, 1, 0]	-0.44	mp-20305
As4Os2	[1, 1, 1]	-0.37	mp-2455
AsPd5In	[2, 1, 0]	-0.23	mp-1025293
AsPd5In	[2, 1, 2]	-0.46	mp-1025293
As4Rh4	[0, 0, 1]	-0.46	mp-22079

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
As4Rh4	[0, 0, 1]	-0.19	mp-22079
As4Rh4	[1, 0, 2]	-0.25	mp-22079
As4Rh4	[2, 2, 1]	-0.45	mp-22079
As4Rh4	[2, 2, 1]	-0.33	mp-22079
As8Rh4	[1, 0, -1]	-0.45	mp-15954
As2Rh10Ga4	[1, 0, 0]	-0.46	mp-18561
As6Si6	[1, 1, 0]	-0.20	mp-1863
AsSn	[1, 1, 1]	-0.23	mp-2182
As6Ti8	[1, 1, 0]	-0.33	mp-567082
As4Zn4Cu4	[1, 1, 1]	-0.36	mp-676828
As4Zn4Cu4	[1, 1, 1]	-0.24	mp-676828
As4Zn4Cu4	[1, 2, 1]	-0.31	mp-676828
As4Zn4Cu4	[2, 1, 0]	-0.30	mp-676828
AsZnPt5	[0, 0, 1]	-0.40	mp-1025356
AsZnPt5	[1, 1, 1]	-0.44	mp-1025356
Au4Al4	[0, 0, 1]	-0.29	mp-1399
Au4Al4	[0, 0, 1]	-0.21	mp-1399
Au4Al4	[2, 2, 1]	-0.30	mp-1399
AuAl2	[1, 1, 1]	-0.46	mp-2647
AuAl2	[2, 1, 0]	-0.36	mp-2647
AuAl2	[2, 2, 1]	-0.37	mp-2647
Au2AlCu	[2, 1, 0]	-0.28	mp-867306
Au3Pd	[2, 1, 1]	-0.19	mp-973834
AuPd3	[1, 0, 0]	-0.30	mp-999298
AuPd3	[1, 0, 0]	-0.18	mp-999298
AuPd3	[2, 1, 1]	-0.40	mp-999298

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
AuPd3	[2, 1, 1]	-0.36	mp-999298
AuPd3	[2, 2, 1]	-0.29	mp-999298
AuPdCu2	[2, 1, 0]	-0.19	mp-862256
Co4Al18	[0, 0, 1]	-0.41	mp-16488
CoFeAl2	[2, 1, 0]	-0.45	mp-862691
Co4Ga12	[1, 0, 1]	-0.21	mp-20559
Co4Ga12	[1, 1, 0]	-0.18	mp-20559
Co4In12	[1, 0, 0]	-0.33	mp-22236
Co4In12	[2, 1, 0]	-0.31	mp-22236
Co4N2	[1, 0, 1]	-0.41	mp-22631
Co2Pt2	[1, 0, 1]	-0.40	mp-949
CoPt3	[1, 0, 0]	-0.36	mp-922
CoPt3	[2, 1, 0]	-0.36	mp-922
CoPt3	[2, 2, 1]	-0.37	mp-922
Co2Sn4	[1, 1, 1]	-0.24	mp-20155
Co2Sn4	[1, 1, 1]	-0.18	mp-20155
Co2Sn4	[2, 1, 1]	-0.20	mp-20155
Co3Sn3	[2, 1, 1]	-0.40	mp-20536
CrCoPt2	[1, 1, 1]	-0.35	mp-570863
CrCoPt2	[1, 1, 2]	-0.36	mp-570863
CrCoPt2	[2, 1, 0]	-0.45	mp-570863
CrCoPt2	[2, 1, 0]	-0.39	mp-570863
CuAlPt2	[0, 0, 1]	-0.40	mp-12550
CuAlPt2	[1, 0, 0]	-0.27	mp-12550
CuAlPt2	[1, 1, 1]	-0.33	mp-12550
CuAlPt2	[1, 1, 2]	-0.29	mp-12550

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
CuAlPt2	[2, 1, 0]	-0.43	mp-12550
CuAlPt2	[2, 1, 2]	-0.44	mp-12550
Cu28In12	[0, 1, 0]	-0.29	mp-21985
Fe2Al12	[0, 0, 1]	-0.30	mp-570001
Fe2Al6	[1, 0, 0]	-0.43	mp-984873
Fe2Al6	[1, 1, 0]	-0.24	mp-984873
Fe2Al6	[2, 1, 1]	-0.46	mp-984873
Fe4Ga12	[1, 0, 1]	-0.33	mp-636368
Fe4Ga12	[1, 1, 0]	-0.25	mp-636368
Fe2Pd2	[1, 0, 1]	-0.44	mp-2831
Fe2Pd2	[1, 1, 0]	-0.30	mp-2831
Fe2Pd2	[2, 1, 2]	-0.35	mp-2831
Fe2Pd2	[2, 1, 2]	-0.33	mp-2831
FePd3	[1, 0, 0]	-0.30	mp-21845
FePt	[1, 0, 0]	-0.40	mp-2260
FePt	[1, 1, 1]	-0.41	mp-2260
FePt	[2, 0, 1]	-0.41	mp-2260
FePt	[2, 1, 1]	-0.47	mp-2260
Fe2WAl	[1, 0, 0]	-0.44	mp-862288
Fe2WAl	[2, 1, 0]	-0.44	mp-862288
Ga2N2	[1, 0, 1]	-0.37	mp-804
Ga10Pd2	[1, 1, 0]	-0.27	mp-30660
Ga3Pd7	[0, 0, 1]	-0.27	mp-31467
Ga3Pd7	[1, 0, 0]	-0.38	mp-31467
Ga3Pd7	[1, 0, 0]	-0.35	mp-31467
Ga3Pd7	[1, 1, -1]	-0.31	mp-31467

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Ga4Pd8	[0, 0, 1]	-0.27	mp-1869
Ga4Pd8	[0, 1, 0]	-0.24	mp-1869
Ga4Pd8	[1, 0, 1]	-0.24	mp-1869
Ga4Pd8	[1, 0, 1]	-0.23	mp-1869
Ga4Pd8	[1, 1, 0]	-0.34	mp-1869
Ga4Pd8	[1, 1, 0]	-0.31	mp-1869
Ga4Pd8	[2, 1, 0]	-0.32	mp-1869
Ga6Pd10	[0, 0, 1]	-0.34	mp-2408
Ga2Pd2Ti2	[1, 0, 0]	-0.26	mp-1025045
Ga2Pt6	[0, 0, 1]	-0.27	mp-862621
Ga2Pt6	[0, 0, 1]	-0.19	mp-862621
Ga2Pt6	[1, 0, 1]	-0.42	mp-862621
Ga2Pt6	[1, 0, 2]	-0.42	mp-862621
Ga3Pt5	[0, 0, 1]	-0.43	mp-30663
Ga3Pt5	[0, 1, 0]	-0.39	mp-30663
Ga3Pt5	[0, 1, 1]	-0.43	mp-30663
Ga3Pt5	[0, 1, 1]	-0.43	mp-30663
Ga3Pt5	[1, 1, 2]	-0.43	mp-30663
Ga4Pt12	[0, 0, 1]	-0.32	mp-623066
Ga4Pt12	[1, 0, 2]	-0.43	mp-623066
Ga8Pt16	[1, 0, 0]	-0.45	mp-2223
GaPt3	[1, 0, 0]	-0.46	mp-11407
GaPt3	[1, 1, 0]	-0.33	mp-11407
GaPt3	[2, 1, 1]	-0.42	mp-11407
GaPt2Cu	[1, 0, 0]	-0.44	mp-644280
GaPt2Cu	[1, 0, 2]	-0.42	mp-644280

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
GaPt2Cu	[1, 1, 0]	-0.45	mp-644280
GaPt2Cu	[1, 1, 0]	-0.35	mp-862791
GaPt2Cu	[1, 1, 1]	-0.39	mp-862791
GaPt2Cu	[1, 1, 2]	-0.35	mp-644280
GaPt2Cu	[2, 0, 1]	-0.46	mp-644280
GaPt2Cu	[2, 1, 0]	-0.41	mp-644280
GaPt2Cu	[2, 1, 1]	-0.28	mp-644280
GaPt2Cu	[2, 1, 2]	-0.37	mp-644280
GaPt2Cu	[2, 2, 1]	-0.44	mp-644280
Ge4Co4Cu2	[1, 0, 0]	-0.22	mp-19955
Ge2Cu6	[1, 1, 2]	-0.21	mp-19724
Ge2Cu6	[1, 2, 1]	-0.18	mp-19724
Ge2Mn2Ga2	[2, 0, 1]	-0.47	mp-1018802
Ge2Mo	[1, 1, 2]	-0.35	mp-10201
GePtTi	[2, 1, 0]	-0.45	mp-1008680
Ge4Rh4	[0, 0, 1]	-0.44	mp-22239
Ge4Rh4	[0, 2, 1]	-0.38	mp-22239
Ge4Rh4	[1, 0, 0]	-0.28	mp-20866
Ir	[1, 1, 1]	-0.37	mp-101
Ir2Al6	[0, 0, 1]	-0.44	mp-2294
Ir2Al6	[0, 0, 1]	-0.41	mp-2294
Ir2Al6	[1, 0, 0]	-0.46	mp-2294
Ir2Al6	[1, 0, 0]	-0.26	mp-2294
Ir2Al6	[2, -1, 2]	-0.33	mp-2294
IrFeAl2	[1, 1, 0]	-0.45	mp-866031
Ir4Ga18	[1, 2, 0]	-0.31	mp-31311

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Ir6Ga12	[0, 0, 1]	-0.43	mp-31253
Ir6Ga12	[1, 1, 0]	-0.37	mp-31253
IrGa	[1, 1, 1]	-0.46	mp-11388
Ir4In8	[0, 1, 0]	-0.38	mp-22812
IrRhAl2	[1, 1, 1]	-0.33	mp-862694
Ir6Sn14	[1, 0, 0]	-0.21	mp-22040
IrSn2	[1, 0, 0]	-0.31	mp-2083
IrSn2	[2, 1, 1]	-0.22	mp-2083
Ir6V2	[1, 1, 0]	-0.39	mp-865496
MnRh2In	[2, 1, 0]	-0.47	mp-864968
Mn2RhPt	[2, 1, 0]	-0.39	mp-865032
Mn2RhPt	[2, 1, 0]	-0.33	mp-865032
MnRhTi2	[1, 0, 0]	-0.39	mp-866218
MoN	[2, 2, 1]	-0.38	mp-13036
MoZn7	[1, 1, 1]	-0.40	mp-644500
N2V2	[2, 1, 0]	-0.18	mp-1017532
Ni	[1, 1, 0]	-0.41	mp-23
Ni2Al3	[1, 1, 1]	-0.38	mp-1057
Ni2Al3	[2, -1, 2]	-0.45	mp-1057
Ni2Al3	[2, 1, 0]	-0.30	mp-1057
Ni2Al3	[2, 2, 1]	-0.32	mp-1057
Ni4Al12	[0, 0, 1]	-0.35	mp-622209
Ni4Al12	[0, 1, 0]	-0.40	mp-622209
Ni5Al3	[0, 0, 1]	-0.41	mp-16514
Ni5Al3	[0, 2, 1]	-0.31	mp-16514
Ni5Al3	[1, 1, 2]	-0.46	mp-16514

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Ni5Al3	[1, 1, 2]	-0.45	mp-16514
Ni5Al3	[2, 0, 1]	-0.32	mp-16514
NiAl	[1, 0, 0]	-0.36	mp-1487
NiAl	[1, 1, 1]	-0.23	mp-1487
NiAl	[2, 1, 0]	-0.43	mp-1487
Ni2AuSn4	[1, 0, 1]	-0.23	mp-568925
NiCo2Ga	[1, 0, 0]	-0.32	mp-20551
NiCo2Ga	[2, 1, 1]	-0.41	mp-1018060
Ni2Fe2	[1, 0, 0]	-0.41	mp-2213
Ni2Fe2	[2, 1, 0]	-0.42	mp-2213
Ni2Fe2	[2, 1, 2]	-0.41	mp-2213
Ni3Fe	[1, 0, 0]	-0.40	mp-1007855
Ni3Fe	[1, 1, 0]	-0.42	mp-1418
NiFeAl2	[2, 1, 0]	-0.40	mp-867330
NiFePt2	[0, 0, 1]	-0.39	mp-13463
NiFePt2	[0, 0, 1]	-0.31	mp-13463
NiFePt2	[1, 0, 0]	-0.39	mp-13463
NiFePt2	[1, 0, 0]	-0.37	mp-13463
NiFePt2	[1, 1, 1]	-0.38	mp-13463
NiFePt2	[2, 1, 1]	-0.45	mp-13463
Ni13Ga9	[1, 0, 0]	-0.42	mp-21589
Ni2Ga3	[2, -1, 2]	-0.43	mp-11397
Ni2Ga3	[2, -1, 2]	-0.43	mp-11397
Ni4Ga2	[1, 1, 1]	-0.39	mp-570904
Ni5Ga3	[0, 0, 1]	-0.44	mp-11398
Ni4Ge4	[1, 0, 2]	-0.37	mp-1099

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Ni ₂ In ₃	[1, 1, 1]	-0.22	mp-21385
Ni ₆ Mo ₂	[0, 1, 0]	-0.46	mp-11506
Ni ₂ Pt ₂	[0, 0, 1]	-0.36	mp-945
Ni ₂ Pt ₂	[1, 0, 1]	-0.39	mp-945
Ni ₂ Pt ₂	[2, 1, 0]	-0.47	mp-945
Ni ₂ Pt ₂	[2, 1, 0]	-0.39	mp-945
Ni ₂ Pt ₂	[2, 2, 1]	-0.45	mp-945
Ni ₂ Pt ₂	[2, 2, 1]	-0.42	mp-945
Ni ₃ Pt	[1, 0, 0]	-0.43	mp-12798
Ni ₃ Pt	[1, 0, 0]	-0.33	mp-12798
NiRh ₂ Ga	[1, 1, 1]	-0.44	mp-866037
NiRh ₂ Ga	[2, 1, 0]	-0.43	mp-866037
NiRh ₂ Sn	[1, 1, 1]	-0.40	mp-11519
Ni ₂ Si ₂	[0, 0, 1]	-0.31	mp-999192
Ni ₂ Si ₂	[1, 2, 1]	-0.38	mp-999192
Ni ₂ Si ₂	[1, 2, 2]	-0.36	mp-999192
Ni ₂ Si ₂	[2, 1, 2]	-0.36	mp-999192
Ni ₄ Si ₄	[1, 0, 2]	-0.39	mp-351
Ni ₈ Si ₄	[1, 1, 1]	-0.45	mp-1118
Ni ₃ Sn ₄	[1, 1, -2]	-0.28	mp-20174
Ni ₆ Sn ₂	[1, 0, 0]	-0.17	mp-20112
Ni ₃ SnN	[2, 1, 1]	-0.40	mp-1017632
Ni ₄ W	[0, 0, 1]	-0.40	mp-30811
OsAl	[1, 0, 0]	-0.29	mp-875
Os ₄ Ga ₁₂	[0, 0, 1]	-0.29	mp-570844
OsV ₃	[1, 1, 1]	-0.41	mp-866121

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
PbMnRh2	[2, 1, 0]	-0.36	mp-4652
Pb4Pt2	[1, 0, 1]	-0.27	mp-21318
Pb10Rh8	[0, 1, 0]	-0.18	mp-569678
Pd	[1, 1, 0]	-0.41	mp-2
Pd	[1, 1, 1]	-0.45	mp-2
Pd4Cu16	[1, 0, 0]	-0.19	mp-30594
PdCu3	[1, 1, 1]	-0.22	mp-580357
PdCu3	[2, 2, 1]	-0.20	mp-580357
Pd2CuIn	[1, 0, 0]	-0.21	mp-867308
Pd2CuIn	[1, 1, 0]	-0.24	mp-867308
Pd2CuIn	[2, 1, 0]	-0.20	mp-867308
Pd3In	[1, 0, 2]	-0.36	mp-510436
Pd3In	[1, 1, 1]	-0.42	mp-510436
Pd3In	[2, 0, 1]	-0.39	mp-510436
Pd3In	[2, 0, 1]	-0.36	mp-510436
Pd3In	[2, 1, 0]	-0.39	mp-31337
Pd3In	[2, 1, 0]	-0.36	mp-31337
Pd3In	[2, 1, 2]	-0.46	mp-510436
Pd3In	[2, 1, 2]	-0.40	mp-510436
Pd8In4	[1, 0, 2]	-0.36	mp-22646
Pd8In4	[1, 1, 0]	-0.38	mp-22646
Pd8In4	[2, 1, 0]	-0.23	mp-22646
Pd2Ti	[1, 1, 0]	-0.42	mp-1018121
Pd2Ti	[1, 1, 2]	-0.37	mp-1018121
Pd2Ti	[2, 1, 2]	-0.47	mp-1018121
Pd2V	[1, 0, 1]	-0.26	mp-11549

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Pd2V	[1, 1, 0]	-0.44	mp-11549
Pd2V	[2, 1, 2]	-0.41	mp-11549
Pd3V	[1, 0, 0]	-0.40	mp-873
Pd3V	[1, 0, 1]	-0.34	mp-873
Pd3V	[1, 1, 0]	-0.32	mp-873
Pd3V	[2, 1, 1]	-0.40	mp-568711
Pd3V	[2, 1, 1]	-0.33	mp-568711
Pd3V	[2, 2, 1]	-0.39	mp-568711
Pd6V2	[1, 0, 0]	-0.46	mp-979980
Pd6V2	[1, 0, 2]	-0.24	mp-979980
Pd6V2	[1, 1, 1]	-0.37	mp-979980
Pt	[1, 1, 1]	-0.36	mp-126
Pt7Cu	[1, 0, 0]	-0.44	mp-12608
Pt7Cu	[1, 0, 0]	-0.36	mp-12608
PtCu3	[1, 1, 0]	-0.34	mp-12086
PtCu3	[1, 1, 0]	-0.28	mp-12086
PtCu3	[1, 1, 1]	-0.32	mp-12086
PtCu3	[2, 1, 1]	-0.27	mp-12086
PtCu3	[2, 2, 1]	-0.25	mp-12086
Pt2CuIn	[1, 0, 0]	-0.35	mp-639659
Pt2CuIn	[2, 0, 1]	-0.44	mp-639659
Pt2CuIn	[2, 1, 1]	-0.46	mp-639659
Pt2FeCu	[0, 0, 1]	-0.31	mp-3702
Pt2FeCu	[1, 0, 1]	-0.42	mp-3702
Pt13In9	[1, 0, 0]	-0.44	mp-571060
Pt13In9	[2, 0, -1]	-0.37	mp-571060

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Pt3In	[1, 0, 0]	-0.46	mp-20516
Pt3In	[1, 0, 0]	-0.25	mp-20516
Pt5In5	[0, 0, 1]	-0.36	mp-510438
Pt5In5	[1, 1, -1]	-0.30	mp-510438
Pt8Ti	[2, 1, 1]	-0.38	mp-30852
Pt2V	[0, 0, 1]	-0.35	mp-12108
Pt2V	[0, 1, 2]	-0.27	mp-12108
Pt2V	[1, 1, 0]	-0.45	mp-12108
Pt2V2	[1, 0, 0]	-0.45	mp-1017531
Pt3V	[0, 0, 1]	-0.44	mp-2705
Pt3V	[0, 0, 1]	-0.21	mp-2705
Pt3V	[1, 1, 2]	-0.26	mp-2705
Pt3V	[2, 1, 0]	-0.31	mp-372
Re2Al12	[0, 1, 0]	-0.17	mp-16528
Re2Al12	[1, 1, 0]	-0.46	mp-16528
ReAl12	[1, 0, 0]	-0.22	mp-1648
Re6Pt2	[0, 0, 1]	-0.46	mp-862589
Re4Si4	[1, 0, 0]	-0.44	mp-7948
Rh	[1, 0, 0]	-0.45	mp-74
Rh	[1, 1, 1]	-0.39	mp-74
Rh4Al18	[0, 0, 1]	-0.40	mp-1645
Rh4Al18	[0, 0, 1]	-0.20	mp-1645
Rh4Al18	[0, 1, 1]	-0.37	mp-1645
Rh4Al18	[1, 0, -1]	-0.19	mp-1645
Rh4Al18	[1, 0, 0]	-0.35	mp-1645
Rh4Al18	[1, 0, 0]	-0.30	mp-1645

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Rh4Al18	[1, 0, 1]	-0.36	mp-1645
Rh4Al18	[1, 1, 0]	-0.31	mp-1645
Rh4Al18	[1, 1, 0]	-0.30	mp-1645
Rh8Al20	[0, 0, 1]	-0.45	mp-1791
Rh8Al20	[1, 1, 0]	-0.42	mp-1791
RhAl	[1, 0, 0]	-0.37	mp-364
RhAl	[1, 1, 0]	-0.26	mp-364
RhAl	[2, 1, 0]	-0.18	mp-364
RhAl	[2, 1, 1]	-0.17	mp-364
RhAl	[2, 2, 1]	-0.34	mp-364
Rh2AlTi	[2, 1, 0]	-0.45	mp-866153
Rh2AlTi	[2, 1, 0]	-0.41	mp-866153
Rh2CoSn	[1, 0, 0]	-0.42	mp-1018085
Rh2CuGa	[1, 1, 1]	-0.47	mp-862485
Rh2CuV	[1, 0, 0]	-0.43	mp-979910
Rh2CuV	[1, 1, 1]	-0.37	mp-979910
Rh2CuV	[2, 1, 0]	-0.42	mp-979910
Rh2FeAl	[2, 1, 0]	-0.42	mp-861953
Rh6Ga10	[0, 0, 1]	-0.23	mp-30923
Rh6Ga10	[1, 0, -2]	-0.29	mp-30923
Rh6Ga10	[1, 2, -1]	-0.20	mp-30923
RhGa	[1, 0, 0]	-0.37	mp-2444
RhGa	[1, 0, 0]	-0.36	mp-2444
Rh4In12	[1, 0, 0]	-0.28	mp-18614
Rh4In12	[1, 0, 1]	-0.28	mp-18614
Rh4In12	[2, 1, 1]	-0.31	mp-18614

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Rh3Pt	[1, 0, 0]	-0.45	mp-974370
Rh3Pt	[1, 0, 0]	-0.39	mp-974370
Rh3Pt	[1, 0, 0]	-0.34	mp-974376
Rh3Pt	[1, 0, 1]	-0.42	mp-974376
Rh3Pt	[1, 0, 1]	-0.39	mp-974376
Rh3Pt	[1, 1, 1]	-0.43	mp-974370
Rh3Pt	[2, 1, 0]	-0.41	mp-974376
RhPt3	[0, 0, 1]	-0.41	mp-974616
RhPt3	[1, 1, 2]	-0.37	mp-974616
Rh3Sn	[1, 1, 2]	-0.42	mp-978974
Rh4Sn4	[1, 0, 0]	-0.23	mp-317
Rh4Sn4	[1, 1, 0]	-0.37	mp-317
Rh4Sn4	[1, 1, 1]	-0.33	mp-317
Rh4Sn4	[2, 1, 0]	-0.25	mp-317
Rh2SnTi	[1, 0, 0]	-0.34	mp-865707
Rh2SnTi	[2, 1, 0]	-0.45	mp-865707
Rh2V2	[0, 0, 1]	-0.34	mp-971751
Rh2V2	[1, 0, 0]	-0.32	mp-1251
Rh2V2	[2, 2, 1]	-0.46	mp-1251
Rh2V2	[2, 2, 1]	-0.41	mp-1251
Rh2V6	[1, 1, 0]	-0.44	mp-1578
Rh3V	[1, 0, 0]	-0.43	mp-1185
Ru2	[1, 0, 0]	-0.44	mp-33
Ru2As4	[1, 2, 0]	-0.19	mp-766
Ru2As4	[2, 1, 0]	-0.43	mp-766
RuCoAl2	[1, 1, 1]	-0.41	mp-862695

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Ru4In12	[1, 0, 0]	-0.31	mp-672326
Ru2SiGa	[2, 1, 0]	-0.35	mp-865615
Ru2Zn6	[1, 0, 1]	-0.46	mp-1380
Sb2Al2	[1, 1, 1]	-0.45	mp-1018100
Sb2Al2	[1, 1, 1]	-0.28	mp-1018100
Sb2Al2	[2, 1, 1]	-0.35	mp-1018100
Sb2Al2	[2, 1, 2]	-0.32	mp-1018100
SbAl	[1, 1, 1]	-0.38	mp-2624
SbAl	[2, 2, 1]	-0.34	mp-2624
Sb4Co2	[2, 1, 2]	-0.30	mp-9835
Sb8Co4	[2, 1, 0]	-0.20	mp-755
Sb8Co4	[2, 1, 1]	-0.22	mp-755
Sb2CuNi	[1, 0, 0]	-0.36	mp-11834
Sb2CuNi	[1, 0, 1]	-0.29	mp-11834
Sb2Ga2	[2, 1, 1]	-0.24	mp-1018059
SbGa	[2, 1, 0]	-0.31	mp-1156
Sb8Ir4	[1, 1, 0]	-0.37	mp-1247
SbMnRh	[2, 1, 0]	-0.37	mp-4846
SbMnRh2	[1, 1, 2]	-0.42	mp-571163
SbMnRu2	[2, 1, 0]	-0.41	mp-864957
Sb2Ni2	[1, 0, 0]	-0.43	mp-810
Sb2Ni2	[1, 1, 1]	-0.25	mp-810
Sb2Ni2	[1, 1, 1]	-0.21	mp-810
Sb2Ni2	[2, 1, 1]	-0.23	mp-810
Sb2Ni2	[2, 1, 1]	-0.21	mp-810
Sb4Os2	[0, 1, 2]	-0.43	mp-2695

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Sb4Os2	[0, 1, 2]	-0.40	mp-2695
Sb2Pt2	[1, 0, 0]	-0.33	mp-2845
SbPt7	[1, 0, 0]	-0.33	mp-1030
SbPt7	[1, 1, 1]	-0.29	mp-1030
Sb4Rh4	[1, 0, 0]	-0.20	mp-20619
Sb4Rh8	[1, 0, 0]	-0.35	mp-21359
Sb4Rh8	[2, 0, 1]	-0.42	mp-21359
Sb8Rh4	[1, 1, -2]	-0.41	mp-2682
Sb8Rh4	[1, 2, -1]	-0.20	mp-2682
Sb8Rh4	[2, 1, -1]	-0.21	mp-2682
SbRh2Cu	[2, 1, 0]	-0.44	mp-867753
Sb8Si8Pt8	[1, 0, 0]	-0.45	mp-11152
Sb8Si8Pt8	[1, 1, 0]	-0.43	mp-11152
SbSiPt5	[0, 0, 1]	-0.30	mp-1025366
SbSiPt5	[1, 0, 0]	-0.31	mp-1025366
SbSiPt5	[2, 1, 2]	-0.47	mp-1025366
Sb4V2	[2, 1, 1]	-0.41	mp-2851
SiAgPt5	[1, 0, 0]	-0.44	mp-1025220
SiAgPt5	[1, 0, 1]	-0.44	mp-1025220
SiAgPt5	[2, 1, 2]	-0.47	mp-1025220
Si2Co2Al	[1, 1, 0]	-0.41	mp-10010
Si2Co2Al	[1, 1, 1]	-0.42	mp-10010
Si2Co2Al	[2, 1, 0]	-0.42	mp-10010
Si2Co2Al	[2, 1, 0]	-0.42	mp-10010
Si2Fe4	[1, 1, 1]	-0.43	mp-22787
Si8Fe4Al12	[1, 1, 0]	-0.38	mp-505229

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Si8Fe6Al4	[0, 0, 1]	-0.40	mp-29111
SiFe2Al	[1, 1, 1]	-0.37	mp-867878
Si2Ir6	[0, 0, 1]	-0.30	mp-1841
Si2Ir6	[2, 1, 1]	-0.47	mp-1841
Si4Ir4	[0, 1, 1]	-0.44	mp-1128
Si2Mn6Al18	[1, 1, 0]	-0.37	mp-15819
Si16Os8	[0, 0, 1]	-0.47	mp-17123
SiOs2V	[2, 1, 0]	-0.46	mp-865506
SiOs2V	[2, 1, 0]	-0.45	mp-865506
Si3Pd6	[1, 0, 1]	-0.19	mp-697068
Si4Pd12	[0, 0, 1]	-0.34	mp-20622
Si4Pd12	[0, 1, 1]	-0.25	mp-20622
Si4Pd12	[1, 0, 0]	-0.31	mp-20622
Si4Pd12	[1, 0, 2]	-0.35	mp-20622
Si4Pd12	[1, 0, 2]	-0.32	mp-20622
Si4Pd12	[1, 2, 0]	-0.33	mp-20622
Si4Pd12	[2, 1, 0]	-0.31	mp-20622
Si2Pt6	[0, 1, 0]	-0.42	mp-13363
Si2Pt6	[2, 0, -1]	-0.33	mp-13363
Si4Pt12	[0, 0, 1]	-0.44	mp-21163
Si4Pt12	[1, 2, 0]	-0.39	mp-21163
Si4Pt12	[1, 2, 1]	-0.40	mp-21163
Si4Pt4	[2, 0, 1]	-0.46	mp-696
SiPt2	[1, 1, 2]	-0.44	mp-1299
SiPt5In	[1, 0, 0]	-0.41	mp-1025370
SiPt5In	[2, 0, 1]	-0.34	mp-1025370

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
SiPt5In	[2, 1, 0]	-0.40	mp-1025370
Si6Rh10	[1, 0, 0]	-0.38	mp-21012
SiRh2Cu	[1, 1, 0]	-0.47	mp-978532
SiRh2Zn	[1, 0, 0]	-0.46	mp-977384
SiRh2Zn	[2, 1, 0]	-0.42	mp-977384
SiSnPt5	[1, 0, 0]	-0.42	mp-1025324
SiSnPt5	[1, 1, 1]	-0.45	mp-1025324
Si6V3	[1, 0, 0]	-0.34	mp-10711
Si6V3	[1, 1, 0]	-0.21	mp-11190
Si6V3	[2, 1, 1]	-0.36	mp-11190
Si2W	[1, 1, 0]	-0.38	mp-1620
Si6W3	[1, 1, 0]	-0.34	mp-8939
Sn2Cu2	[1, 0, 0]	-0.39	mp-10598
Sn4Pd4	[0, 1, 0]	-0.26	mp-2369
Sn4Pd4	[1, 0, 1]	-0.17	mp-2369
Sn4Pd4	[2, 1, 2]	-0.26	mp-2369
Sn4Pd8	[1, 0, 1]	-0.28	mp-1851
Sn4Pd8	[1, 1, 0]	-0.23	mp-1851
Sn4Pd8	[1, 1, 2]	-0.24	mp-1851
Sn4Pd8	[1, 2, 2]	-0.29	mp-1851
SnPd3	[1, 0, 0]	-0.39	mp-718
SnPdTi	[2, 1, 0]	-0.46	mp-961682
Sn2Pt2	[1, 1, 1]	-0.36	mp-19856
Sn2Pt2	[2, 1, 0]	-0.17	mp-19856
SnPt3	[1, 0, 0]	-0.46	mp-20971
Sn2Ti4	[1, 0, 0]	-0.20	mp-30875

Continued on next page...

Table A.2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
TiPd2In	[1, 0, 0]	-0.32	mp-866168
TiRh2In	[1, 0, 0]	-0.39	mp-866170
TiRh2In	[2, 1, 0]	-0.47	mp-866170
Zn2As4Sn2	[1, 0, 0]	-0.38	mp-5190
Zn3Co	[1, 1, 2]	-0.20	mp-971948
Zn13Fe	[2, 0, 1]	-0.41	mp-1722
ZnPt3	[1, 0, 0]	-0.45	mp-30856
ZnPt3	[2, 1, 0]	-0.39	mp-30856
ZnPt3	[2, 1, 0]	-0.38	mp-30856
Zn2RhPd	[2, 1, 0]	-0.43	mp-864839
ZnRh2V	[1, 0, 0]	-0.35	mp-865487
ZnRh2V	[2, 1, 0]	-0.45	mp-865487
ZnRh2V	[2, 1, 0]	-0.43	mp-865487
Zn3Ti	[2, 2, 1]	-0.34	mp-21289

Appendix B

Supplementary Information for Chapter 4

This work originally appeared as the Supplementary Information for: Miao Zhong, Kevin Tran,* Yimeng Min,* Chuanhao Wang,* Ziyun Wang, Cao-Thang Dinh, Phil De Luna, Zongqian Yu, Armin Sedighian Rasouli, Peter Brodersen, Song Sun, Oleksandr Voznyy, Chih-Shan Tan, Mikhail Askerka, Fanglin Che, Min Liu, Ali Seifitokaldani, Yuanjie Pang, Shen-Chuan Lo, Alexander Ip, Zachary Ulissi, and Edward H. Sargent. Accelerated discovery of CO₂ electrocatalysts using active machine learning. Nature, 581(7807):178–183, 2020.*

B.1 Computational Methods

B.1.1 DFT/ML screening methods

The automated DFT framework was constructed using various Python and shell software packages. The Materials Project[72] was used to establish bulk structures; the Atomic Simulation Environment[64] was used to manage the structures; pymatgen[136] was used to enumerate all facets with Miller indices between -2 and 2 and all symmetrically distinct facet terminations; pymatgen was also used to perform Delaunay triangulation on the surfaces to enumerate adsorption sites; and VASP[91, 92, 89, 90]

was used to perform the DFT calculations. Each surface/site enumeration, DFT calculation, and various administrative computational tasks were encoded as interdependent tasks so that they could be managed automatically by dependency management software, Luigi[15]. FireWorks[71] was also used to manage calculations across several computing clusters.

DFT results were then used by a machine learning workflow to predict CO adsorption energies for each adsorption site enumerated. To accomplish this, a method was developed to encode each adsorption site into a numerical array (Figure B-1). Each element present in the bulk structure was tabulated. Each element was described with a vector four numbers: the atomic number (Z), the Pauling electronegativity (χ), the number of atoms of that element coordinated with the CO molecule (CN) as determined by a cut-off radius of 5 Å and a Voronoi polyhedral angle cutoff tolerance of 0.8,[136] and the median monometallic adsorption energy of CO on that element ($\Delta\tilde{E}$) as calculated from our database of results. The vector creation process was then repeated on the “second shell” of atoms. Given that our database included alloys with no more than three components, the final number of features was $4 \times 2 \times 3$. These features were fed to an automated machine learning tool, TPOT[135], which automatically chose and tuned an appropriate regression method (typically a combination of random forest and boosted tree regressors). The MAE of the models we created with TPOT were ca. 0.18 eV using 90/10 train/test splits and ca. 0.29 eV using time-series splitting. You can refer to the Tran & Ulissi paper[171] for more details.

The resulting ML model and DFT framework were then coupled to create an active machine learning workflow. The ML model was used to predict all the adsorption energies of every adsorption site enumerated by the DFT framework. The sites whose ML-predicted adsorption energies were closest to the optimal value of -0.6 eV were then automatically simulated by the DFT framework to yield a DFT-predicted adsorption energy. The additional DFT data was then used to retrain a new ML model, which created new predictions and prioritizations. Thus, the ML model used the DFT framework to query its own training data continuously, resulting in a database

of DFT data that grew automatically and systematically. In total, the active learning workflow performed over 300 ML regressions to guide DFT calculations of CO binding energies on *ca.* 4,000 different adsorption sites on Cu-containing surfaces, and *ca.* 1,000 of those were performed on CuAl surfaces.

B.1.2 DFT settings for screening

We performed all DFT screening calculations using the Vienna Ab initio Simulation Package (VASP)[91, 92, 89, 90] implemented in ASE.[64] The RPBE functionals[57] were used along with k-point grids of $4 \times 4 \times 1$ and an energy cutoff 350 eV. The default pseudopotentials supplied by VASP version 5.4 were also used. Bulk relaxations were performed with a $10 \times 10 \times 10$ k-point grid and a 500 eV cutoff, and only isotropic relaxation was allowed. All surfaces were replicated in the X/Y directions so that each cell vector was at least 4.5 Å. Corrections for spin magnetism or dispersion were not included, so we excluded from this study all materials that contained Mn, Fe, Ni, Co, or O. All slabs were replicated in the Z direction to a minimum height of 7 Å with at least 20 Å of vacuum between slabs. Generally, the bottom layers were fixed and defined as those atoms more than 3 Å from the top of the surface.

B.1.3 Calculating the optimal ΔE_{CO} value

The literature-sourced scaling relationships[105] calculate an optimal adsorption free energy of CO (ΔG_{CO}) of -0.17 eV. This value was based on microkinetic modeling of the full reaction pathway from CO₂ to methane on single metal terrace (111) and step (211) surfaces and kinetic/thermodynamic scaling relationships. For Cu, hydrogenation of CO was predicted to be the rate-limiting step. If we assume that CO is still involved in the rate-limiting step of other hydrocarbon formation reactions, such as CO-CO or HCO-HCO coupling, then similar ΔG_{CO} targets should apply. If we do not make this assumption, then the ΔG_{CO} target becomes a necessary-but-not-sufficient criterion of activity, because we still know that CO will be an intermediate in the CO₂ reduction pathway and will therefore need to have a quasi-stable ΔG_{CO}

of ca. -0.17 eV. Thus, we need a way to calculate ΔG_{CO} .

Equation B.1 can be used to calculate the change in chemical potential (ΔG_{CO})[105]. $\Delta\mu_{CO*}$ is the chemical potential of CO bound on a surface, $\Delta\mu_*$ is the chemical potential of the surface, and $\Delta\mu_{CO}$ is the chemical potential of CO in the gas phase.

$$\Delta G_{CO} = \mu_{CO*} - \mu_* - \mu_{CO} \quad (B.1)$$

Chemical potentials can be calculated using Equation B.2[134, 105], where E is the electronic energy as calculated by DFT, ZPE is the zero-point energy, C_p is the heat capacity, T is temperature, S is entropy, $\Delta\mu_{solv}$ is the change in chemical potential from solvent de-stabilization, and $\Delta\mu_{corr}$ is any experimental offset required to account for differences between experimental chemical potentials and DFT-based chemical potentials.

$$\mu = E + ZPE + \int C_p dT - TS + \Delta\mu_{solv} + \Delta\mu_{corr} \quad (B.2)$$

Literature[165] reports that ZPE_{CO} is 0.130 eV, $C_{p,CO}dT$ is 0.091 eV at 298 K, S_{CO} is 0.002092 eV K⁻¹, and $\Delta\mu_{corr,CO}$ is 0.02 eV for gas phase CO[165]. Thus, the chemical potential of CO in the gas phase is:

$$\begin{aligned} \mu_{CO,g} &= E_{CO,g} + ZPE_{CO,g} + \int C_{p,CO} dT - TS_{CO,g} + \Delta\mu_{solv,g} + \Delta\mu_{corr,g} \\ \mu_{CO,g} &= E_{CO,g} + 0.130 \text{ eV} + 0.091 \text{ eV} - (298 \cdot 0.002092) \text{ eV} + 0 \text{ eV} + 0.02 \text{ eV} \\ \mu_{CO,g} &= E_{CO,g} - 0.362 \text{ eV} \end{aligned} \quad (B.3)$$

The same method is used to calculate the adsorbed-state chemical potential of CO (μ_{CO*}), given literature values[165] for ZPE , C_p , and S as well as a solvation correction term[105] (μ_{solv}) and an empirical correction term[7] (μ_{corr}) as shown in Equation B.4. Note that we assumed a CO stretch frequency of 2,000 cm⁻¹, and deviations from this frequency may cause errors.

$$\begin{aligned}
\mu_{CO*} &= E_{CO*} + ZPE_{CO*} + \int C_p dT - TS_{CO*} + \Delta\mu_{solv} + \Delta\mu_{corr} \\
\mu_{CO*} &= E_{CO*} + 0.192 \text{ eV} + 0.085 \text{ eV} - (298 \cdot 0.000452) \text{ eV} - 0.2 \text{ eV} + 0.2 \text{ eV} \quad (\text{B.4}) \\
\mu_{CO*} &= E_{CO*} + 0.142 \text{ eV}
\end{aligned}$$

Assuming that μ_* is equal to E_* . By combining Equations A.1, B.3, and B.4, we can calculate the change in chemical potential of adsorbing CO:

$$\begin{aligned}
\Delta G_{CO} &= \mu_{CO*} - \mu_* - \mu_{CO} \\
\Delta G_{CO} &= [E_{CO*} + 0.142 \text{ eV}] - E_* - [E_{CO,g} - 0.362 \text{ eV}] \\
\Delta G_{CO} &= E_{CO*} - E_* - E_{CO,g} + 0.50 \text{ eV} \\
\Delta G_{CO} &= \Delta E_{CO} + 0.50 \text{ eV}
\end{aligned} \quad (\text{B.5})$$

where the change in electronic energy, ΔE_{CO} , is defined using Equation B.6 and is calculated with DFT[134].

$$\Delta E_X = E_{X*} - E_* - E_{X,g} \quad (\text{B.6})$$

We combined Equation B.5 with the scaling relationship's optimal ΔG_{CO} value of -0.17 eV to calculate an optimal ΔE_{CO} value of -0.67 eV. Note that we used RPBE[57] while the source of the scaling relationships used the BEEF-vdW[179] functionals though, so any differences that may be caused by using different functionals may impart errors into our estimates of activity.

B.1.4 Method for creating 2-dimensional activity & selectivity volcanoes

We thank Xinyan Liu for providing us the code that she used to create the 2-dimensional activity map in her manuscript[105] to generate Figure 4-1a and Figure B-64. to visualize our DFT/ML predicted results. The code combines adsorption

energy scaling relationships with microkinetics of elementary reactions for both CO₂ reduction and H₂ evolution to predict adsorbate coverages, reaction rates, and thus activity and selectivity for CO₂ reduction vs. H₂ evolution. Additional details can be found in the Supplementary Information of Liu et al.[105]. Note that we set the scaling between H and CH_x such that they would be linearly independent; we did this to address convergence issues with the microkinetic models. Refer to Figure B-64 for additional results from the microkinetic modeling.

B.1.5 Method for creating t-SNE diagrams

To represent each adsorption site numerically, we developed two vectors: a coordination vector and a neighbours' coordination vector. Figure B-2 illustrates a simplified example of these vectors. The coordination vector contains one item for each of the 31 elements considered in the screening[171], and each item in this vector is the sum of the number of atoms of that element that are coordinated with CO. The neighbours' coordination vector is a flattened array contains 31×31 items for each elemental pairing, and each item in this vector is the sum of the number of atoms of one element that are coordinated with the all adsorbate neighbours that belong to another element. Algorithm 2 outlines explicitly how we calculated the neighbours' coordination number (ncn).

Algorithm 2 Calculating array of neighbors' coordination numbers for an adsorption site

```

1:  $ncn := \text{zeros}(n, n)$             $\triangleright$  where  $n=31$ , the total number of elements we are
   investigating
2: for all neighbors do
3:    $i := \text{index of the element of neighbor}$ 
4:   for all neighbor's neighbors do
5:      $j := \text{index of the element of neighbor's neighbor}$ 
6:      $ncn_{i,j} := ncn_{i,j} + 1$ 

```

The coordination vector and the neighbors' coordination vector were then concate-

nated into one vector, scaled such that each item in the new vector had a mean of zero and a variance of one, and then sent through SKLearn’s[138] principle component analyzer where the dimensionality was reduced until only 85% of the variance was accounted for, yielding a vector of 113 items instead of 31×32 items.

This reduced vector was processed by the t-SNE algorithm[113] with a perplexity of 120 and a learning rate of (ϵ) of 200 and then stopped after 2,000 iterations, yielding a 2-dimensional reduction of the adsorption sites that our workflow performed DFT calculations for.

B.1.6 DFT calculations

The DFT calculations were performed using a VASP package in the Generalized Gradient Framework formulated by Perdew et al[139], including the RPBE flavour of DFT and the projector augmented wave (PAW) method to account for core-valence interactions[18, 88, 139, 90]. The cutoff of kinetic energy for plane wave expansions was set to 400 eV and the reciprocal space was sampled by the Γ -centered Monkhorst-Pack scheme[127] with a grid of $4 \times 4 \times 1$ and an energy cutoff 350 eV. The Cu (111) and (100) surface slabs were constructed with three Cu layers using ASE (Atomic Simulation Environment)[64] using the RPBE-optimized lattice parameter with a vacuum layer of at least 15 Å.

Zero-point energies (ZPE), enthalpy and entropy contributions to free energies at room temperature (298.15 K) were calculated from vibrational modes of surface species using numerical six-point derivatives in VASP and ASE. Note that very low-frequency modes were obtained in some cases because the explicit water molecules are not properly constrained by the hydrogen bonding network presented in water bulk. Such low-frequency modes can cause unphysically large entropy contributions, so they were reset to a threshold value of 60 cm⁻¹ and excluded frequencies smaller than the threshold, corresponding to the acoustic translational mode of the six-member rings in water bulk. For the surface reaction (i.e., $A^* \rightarrow B^*$), the change in Gibbs free energy at temperature T and 1 atmospheric pressure is given by:

$$\Delta G_{A^* \rightarrow B^*}(T, P^0) = \Delta E_{rxn} + \Delta H^0(T) - T\Delta S^0(T, P^0) \quad (\text{B.7})$$

where ΔE_{rxn} is the calculated reaction energy of $A^* \rightarrow B^*$. $\Delta H^0(T)$ and $\Delta S^0(T, P^0)$ are the enthalpy and entropy differences between the initial and final states. In addition, $\Delta H^0(T)$ gives the zero-point energy and the temperature dependence of enthalpy change at a standard pressure of 1 atm for adsorption of molecule A (here, it is room temperature of 298.15 K), which is given by:

$$\begin{aligned} \Delta H^0(T) &= H_{trans}^A + H_{rot}^A + H_{vib}^A + H_{vib}^{A^*} \\ \Delta S^0(T) &= S_{trans}^A + S_{rot}^A + S_{vib}^A + S_{vib}^{A^*} \end{aligned} \quad (\text{B.8})$$

where H_{trans}^A , H_{rot}^A , H_{vib}^A , $H_{vib}^{A^*}$ is the enthalpy at the temperature T and a standard pressure of 1 atm for the translational, rotational, vibrational modes of the A. S_{trans}^A , S_{rot}^A , S_{vib}^A , $S_{vib}^{A^*}$ are entropy contributions from the 3-D translational, 2-D rotational, vibrational modes.

B.2 Experimental Methods

B.2.1 Preparation of evaporated Cu on gas diffusion layers (GDLs)

Cu was evaporated on the GDL (Fuel Cell Store, Sigracet 39 BC) by a thermal evaporation process. GDL is made of an air-brushed polytetrafluoroethylene (PTFE) on carbon nanoparticles layer as a conductive and hydrophobic layer atop a carbon fibre layer as a conductive and supporting layer. 0.5 g Cu foils were placed in a crucible inside the evaporation chamber (Edwards AUTO 360 Thermal Evaporator). A thin Cu layer (*sim*500 nm) was deposited at an evaporation rate of approximately 1-2 nm s⁻¹ under a base pressure of 10⁻⁶ Torr. GDLs were kept rotating at a slow speed of 50 rpm during evaporation.

B.2.2 Preparation of ion-implanted Al-on-Cu on GDLs

Evaporated Cu on GDL samples were used as substrates and placed in ion implantation chamber (50 keV Aluminum Implantation at Western University). A 99.999% purity aluminium (Al) rod is installed in a copper target in the caesium (Cs) sputter source. A 100 keV Al^{2+} molecules are produced in the sputter source injector, selected by the injector magnet and injected into the Tandetron accelerator. Al^{2+} is preferred because it has six times the flux of Al^+ . For ion implants below 100 keV, the Tandetron terminal pump and stripper gas are turned off. The Tandetron is used as a large lens to focus a negative beam to target in the implant chamber. The high energy magnet steers the Al^{2+} down the beamline with the implant chamber. As the beam travels down the beamline it passes through NEC Electrostatic Raster/Scanner which sweeps the beam in the X direction at 517Hz and 64Hz in the Y direction over an aperture which defines the implant area on the implant stage. The Al^{2+} molecule breaks into Al each having an energy of 50 keV after the collisions with the sample. The implant stage has 4 sides. Each side can be rotated to face the beam for implantation. The implant stage in the Implant Chamber is suspended in a Faraday cage to suppress the secondary electrons. The ion charge is collected from the implant stage and fed to an Ortec 439 current integrator. The charge is converted into pulses which are counted by a computer to determine the dose. To avoid substantial ion-beam induced damage and over-heating to the samples, we ion implanted 2 w% Al into Cu which corresponded to 6×10^{15} ions cm^2 . It took 112 minutes to finish this experiment. Auger spectroscopic analysis determined that the $\text{Al}/(\text{Al}+\text{Cu})$ molar concentration on the surface is 4.5%. We implanted 5 wt% and 10 wt% Al into Cu and it took 180 and 360 minutes, respectively. Surface Al concentrations are determined to be 5% and 5.5 %, indicating making more Al on Cu surface is difficult by the ion implantation method. This may be due to the ion-beam-induced damage and heating effect. Al may gradually migrate into the bulk at a large implantation dose.

B.2.3 Preparation of evaporated-etched Al-on-Cu on GDLs

Evaporated-etched Al-on-Cu was synthesized via a two-step process of evaporation and etching. First, 0.3 g Al foils were placed in a crucible inside the evaporation chamber. A thin Al layer (~ 100 nm) was deposited on evaporated Cu-on-GDL samples at an evaporation rate of approximately $1\text{-}2\text{ nm s}^{-1}$ under a base pressure of 10^{-6} Torr. The evaporated Al-on-Cu sample was then immersed in a 5 wt% hydrochloric acid solution to remove the excessive Al. The solution was kept stirring at a low speed of 250 rpm during the 5 minutes etching. Deionized water was used to wash off remained hydrochloric acid and other residual ions from the sample surface. An airbrush was used to dry the samples.

B.2.4 Preparation of de-alloyed Cu-Al and nanoporous Cu on GDLs

De-alloyed Cu-Al was synthesized via an evaporation and etching process. First, 0.5 g Cu foils and 0.25 g Al foils were placed in a crucible inside the deposition chamber. They were melted under a base pressure of 10^{-6} Torr for 2 minutes to form greyish Cu-Al alloys. Then, a thin layer of Cu-Al alloy (~ 500 nm) was deposited on GDLs at an evaporation rate of approximately $1\text{-}2\text{ nm s}^{-1}$ under a base pressure of 10^{-6} Torr. After cooling down to room temperature naturally, the evaporated Cu-Al alloy on GDL samples were transferred to a 5 wt% hydrochloric acid solution to fabricate de-alloyed Cu-Al catalyst at a mild stirring speed of 250 rpm. Deionized water was used to wash off remained hydrochloric acid and other residual ions from the sample surface. An airbrush was used to carefully dry the samples.

Nanoporous Cu samples were prepared by immersing de-alloyed Cu-Al catalysts into 10 mM CuCl_2 solution for 10 minutes with a mild stirring speed of 50 rpm. Then, the samples were washed with deionized water and dried by an airbrush carefully.

B.2.5 Preparation of de-alloyed Cu-Al on polytetrafluoroethylene (PTFE) substrates

De-alloyed Cu-Al was fabricated on the PTFE (pore size: $0.45\ \mu\text{m}$) substrates via a co-sputtering and etching process. Co-sputtering is used to avoid the PTFE melting at high temperatures during evaporation. First, a 200 nm thick Cu-Al layer was co-sputtered on the PTFE using a Cu target (99.99%) and an Al target (99.99%) at a sputtering rate of $3.1\ \text{\AA}/\text{s}$ and $1.1\ \text{\AA}/\text{s}$ respectively under a base pressure of 10^{-6} Torr. Following the same chemical etching process in a 5 wt% hydrochloric acid solution at a stirring speed of 250 rpm, the co-sputtered Cu-Al were de-alloyed. The remained hydrochloric acid and residual ions on the surface was washed off by deionized water. The samples were then carefully dried with an air gun. For the stability test, carbon black NPs (Sigma-Aldrich, $< 100\ \text{nm}$) and graphite particles were dispersed in a mixture of isopropanol, water and Nafion solution (50 microliter of Nafion solution in 0.5 mL of isopropanol:water mixture). The mixture was sonicated for 1 hour and then spray coated on the Cu-Al/PTFE electrodes.

B.2.6 Characterizations

The morphologies of the prepared samples were investigated using scanning electron microscope (SEM) on a Hitachi SU 5000 VPSEM, transmission electron microscope (TEM) on a Hitachi HF-3300 instrument with an acceleration voltage of 200 kV and high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) in a Cs-corrected STEM (JEOL, JEM-ARM200F) at an accelerating voltage of 200 kV. Compositions of the prepared samples were studied using Energy-dispersive X-ray spectroscopy (EDX) on a Bruker Quantax EDX in SEM and TEM, respectively, and electron energy loss spectroscopy (EELS) elemental analysis by GIF Quantum 965. Al concentrations on surfaces of the prepared samples were measured using Auger electron spectroscopy (AES) on a 710 Scanning Auger Nanoprobe instrument (Ulvac-PHI, Chigasaki, Japan). The beam settings used for the acquisition are 10 keV and 10 nA. We performed AES analyses at different locations on the surface

for each sample to trace the Al concentrations and homogeneity. We found that ion implantation gave rather uniform distributions of Al on the surfaces. In contrast, Al concentrations on those chemical etched samples may vary $\pm 5\text{-}10\%$ especially for those samples with high Al concentrations, indicating that the chemical etching condition using HCl solutions might be further optimized. The Cu and Al ion concentrations in the testing solutions were measured using an Inductively Coupled Plasma Atomic Emission Spectrometer (ICP-AES Agilent Dual-View 720 with CCD for full wavelength coverage between 167 to 785 nm).

Ex situ and in situ X-ray absorption spectra at Cu K-edge on catalysts and standard references were collected at the beamline of 1W1B from Beijing synchrotron radiation facility, China. The electron storage ring was operated at 2.5 GeV with an average current of 200 mA. A Si (111) double crystal was used as a monochromator and the data of absorption were collected in fluorescence mode. The energy of the absorption spectra was calibrated by measuring the X-ray absorption near edge spectroscopy (XANES) of a Cu metal foil. The obtained data were processed by established methods with the ATHENA software package. The normalized extended X-ray absorption fine-structure spectroscopy (EXAFS) was converted from energy to k-space and weighted by k^3 . These data were then Fourier transformed to R-space.

For in situ measurements, a self-built flow cell consisting of two chambers for CO₂ diffusion (chamber I) and holding the reaction liquid (chamber II) was used. The sample is sandwiched between the two chambers with pressed the Teflon spacers (sample size: $1.2 \times 1.2 \text{ cm}^2$; thickness: 500 nm). CO₂ gas is introduced into the chamber I with the gas controller and diffuses and passes through the samples and finally reaches the liquid to participate in the reaction. The thickness of the liquid can be controlled from 1000 to 200 μm . After the liquid cell filled with 1.0 M aqueous KOH, 30 mL KOH solution with the same concentration is flowed into the liquid cell at 0.5 mL/min by using an automatic pump syringe. The measurement of the Cu K-edge is started simultaneously with the stop of the flowing.

B.2.7 Electrochemical reduction of CO₂

All CO₂ electrolysis experiments were performed using a three-electrode set-up in a flow-cell configuration connected to an electrochemical workstation (Autolab PG-STAT302N). An Ag/AgCl (in saturated KCl solution) and a platinum wire were used as reference and counter electrodes, respectively. 1 M KOH was used as the electrolyte. To quantify FEs for each product, we performed CO₂ electrolysis in a chronopotentiometry mode. For Tafel analysis, we performed CO₂ electrolysis in a Chronoamperometry mode. For linear sweep voltammogram (LSV) analyses, we cycled the LSV until it was stable.

The cathodic chamber was separated from the gas chamber by a 1.5 cm × 1.5 cm gas diffusion electrode (GDE). GDEs were made of catalysts on GDLs as discussed in detail above. A plastic plate with a 1 cm × 1 cm window is placed between the GDE and cathode chamber. The cathodic chamber was separated from the anodic chamber by an anion exchange membrane (Fumasep FAA-3-PK-130). Rubber spacers were placed in between different chambers to avoid mixture of electrolytes or gaseous products.

CO₂ gas was delivered into the gas chamber at a rate of 56 standard cubic centimetres per minute (s.c.c.m.) and was routed into a gas chromatograph (Perkin Elmer Clarus 680) to quantify the gaseous products. The liquid products were quantified by NMR (600 MHz Agilent DD2 spectrometer), in which electrolyte was mixed with D₂O (deuterated water) and dimethyl sulfoxide (DMSO, Sigma, 99.99%) was used as an internal standard.

Faradaic efficiencies (FEs) for different products can be calculated as follows: $FE = F \times m \times n / Q = F \times m \times n / (I \times t)$, where F is the Faraday constant, m is the electron numbers needed for one CO₂ molecule reduction to the desired product and n is the amount of the desired product (in moles). Half-cell electricity-to-ethylene power conversion efficiency (Half-cell C₂H₄ PCE) can be estimated as follows: Half-cell C₂H₄ PCE = $FE \times V_{theoretic} / V_{real}$, where $V_{theoretic} = 1.17$ (VRHE), $V_{real} = (1.23 - V_{applied})$ (VRHE).

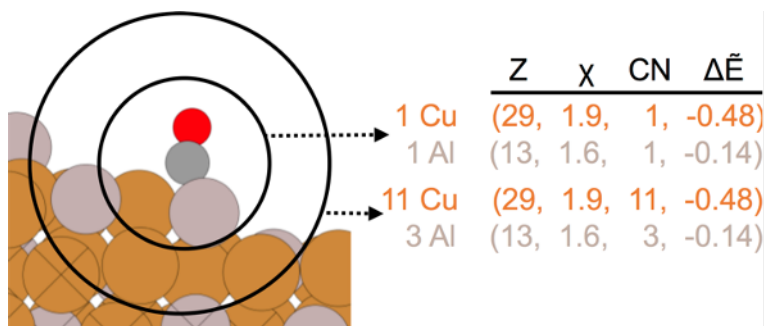


Figure B-1: An example of the numerical encoding of an adsorption site. Vectors are created for each element present within the first and second neighbouring shells of CO. Each vector contains the atomic number of the element (Z), the Pauling electronegativity of the element (χ) the number of atoms of that element within each respective shell (CN), and the median monometallic adsorption energy of CO on that element ($\Delta\tilde{E}$). Color codes for elements: orange is Cu, light gray is Al, dark gray is C, and red is O.

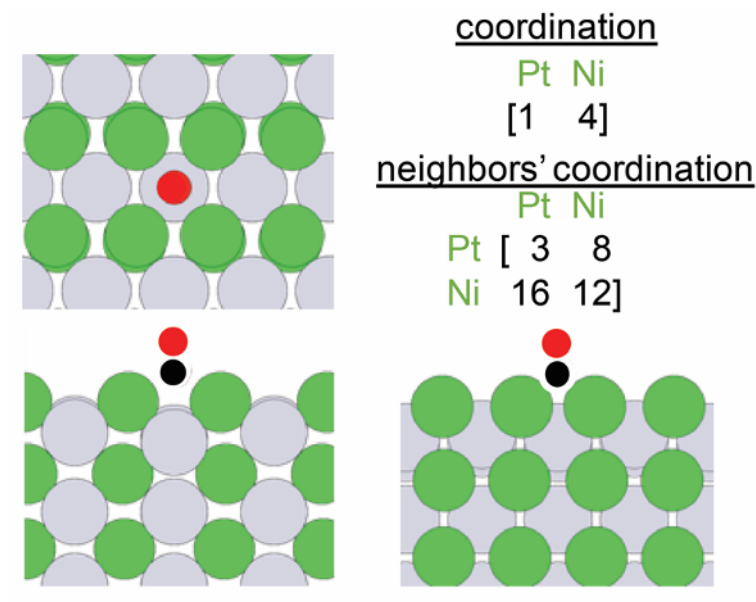


Figure B-2: A simplified example of a numerical representation of a coordination site which was used for t-SNE analysis. Each item in the “coordination” vector represents the coordination number for a particular element, e.g., Pt or Ni. Each item in the “neighbours” coordination” array represents this same coordination vector, but for each of the adsorbate’s neighbours. Note that this example is for illustrative purposes only. The vectors and arrays actually used contained enough items to represent 31 different elements, totalling in 31×32 features. Color codes for elements: green is Ni, light gray is Pt, black is C, and red is O.

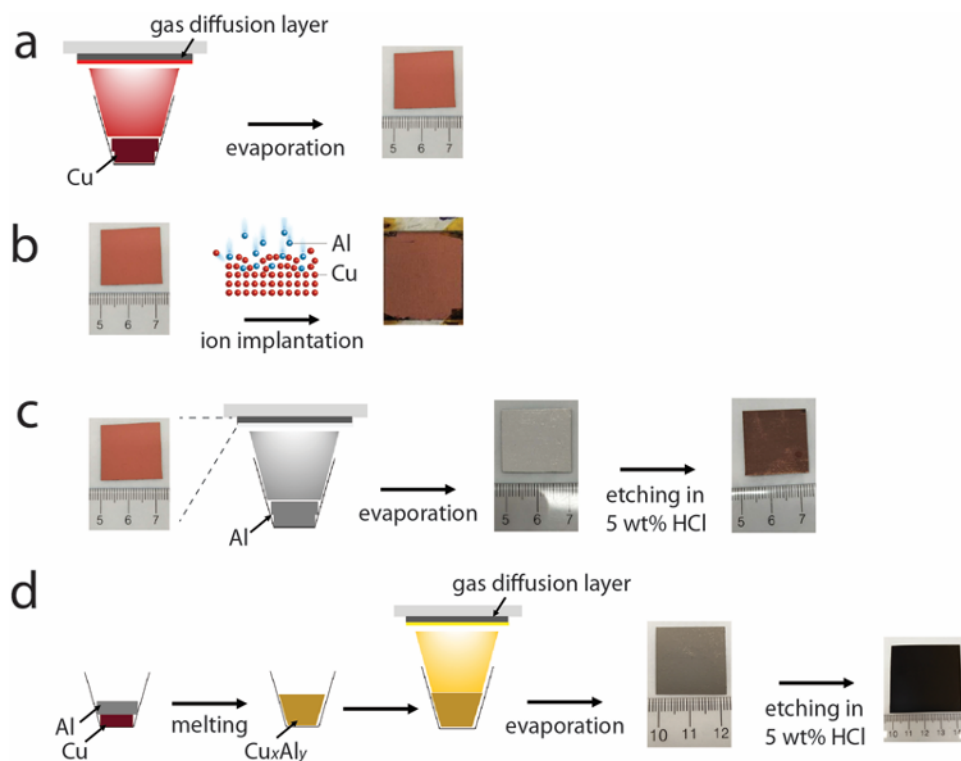


Figure B-3: Schematic for the synthesis of different catalysts on gas diffusion layers. a, evaporated Cu. b, ion-implanted Al-on-Cu. c, evaporated-etched Al-on-Cu. d, de-alloyed Cu-Al catalysts.

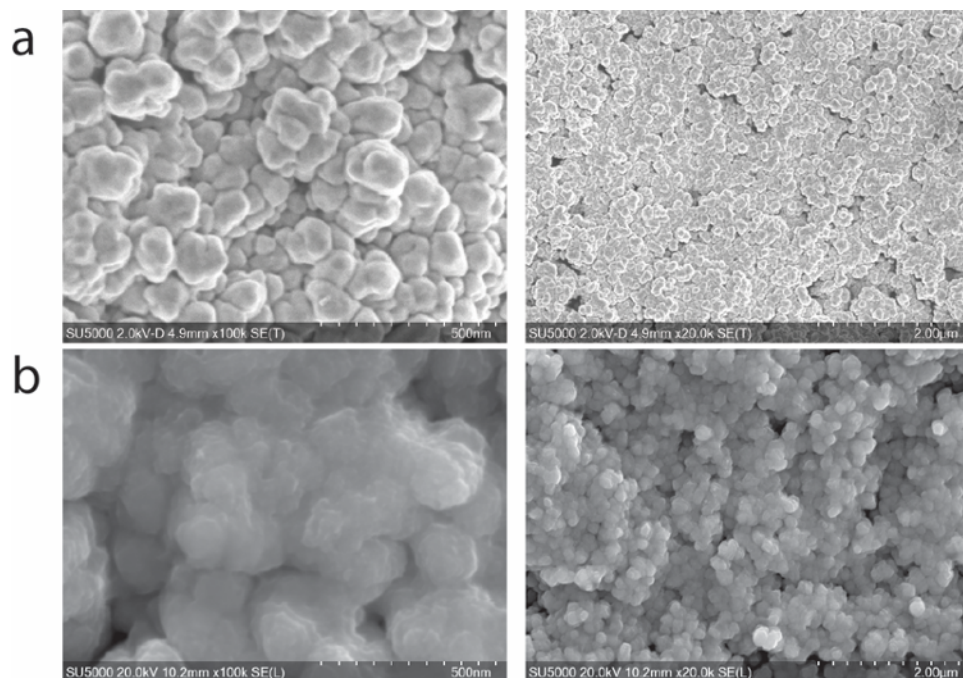


Figure B-4: Morphologies of evaporated Cu catalysts on gas diffusion layers. a, Top-view SEM images before CO₂ electroreduction. b, Top-view SEM images after 5 hours CO₂ electroreduction in 1 M KOH at an applied current density of 600 mA cm⁻² in a flow cell.

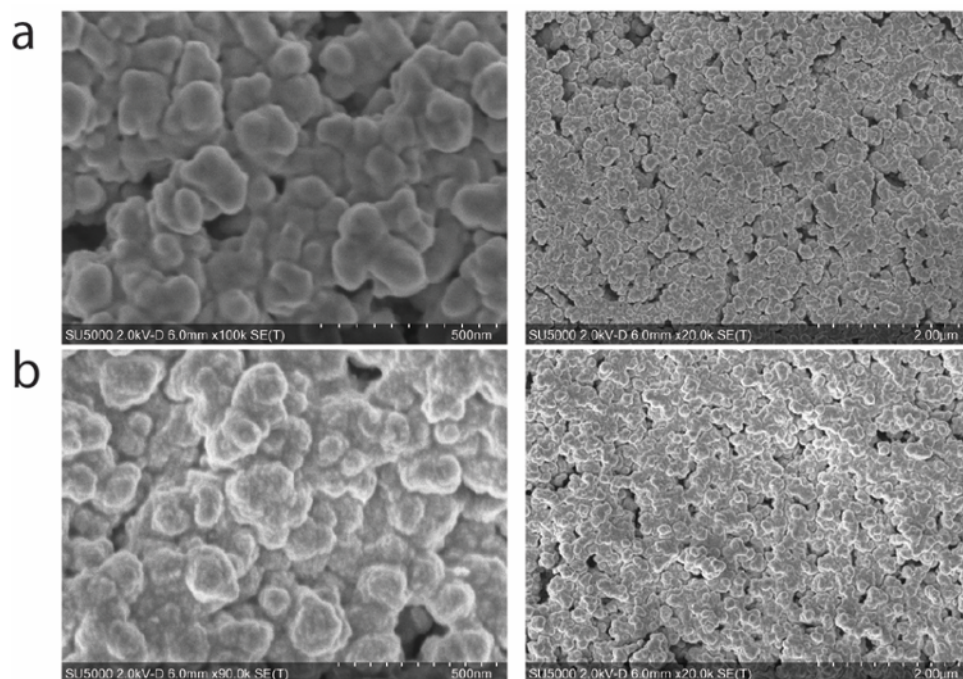


Figure B-5: Morphologies of ion-implanted Al-on-Cu catalysts on gas diffusion layers. a, Top-view SEM images before CO₂ electroreduction. b, Top-view SEM images after 5 h CO₂ electroreduction in 1 M KOH at an applied current density of 600 mA cm⁻² in a flow cell.

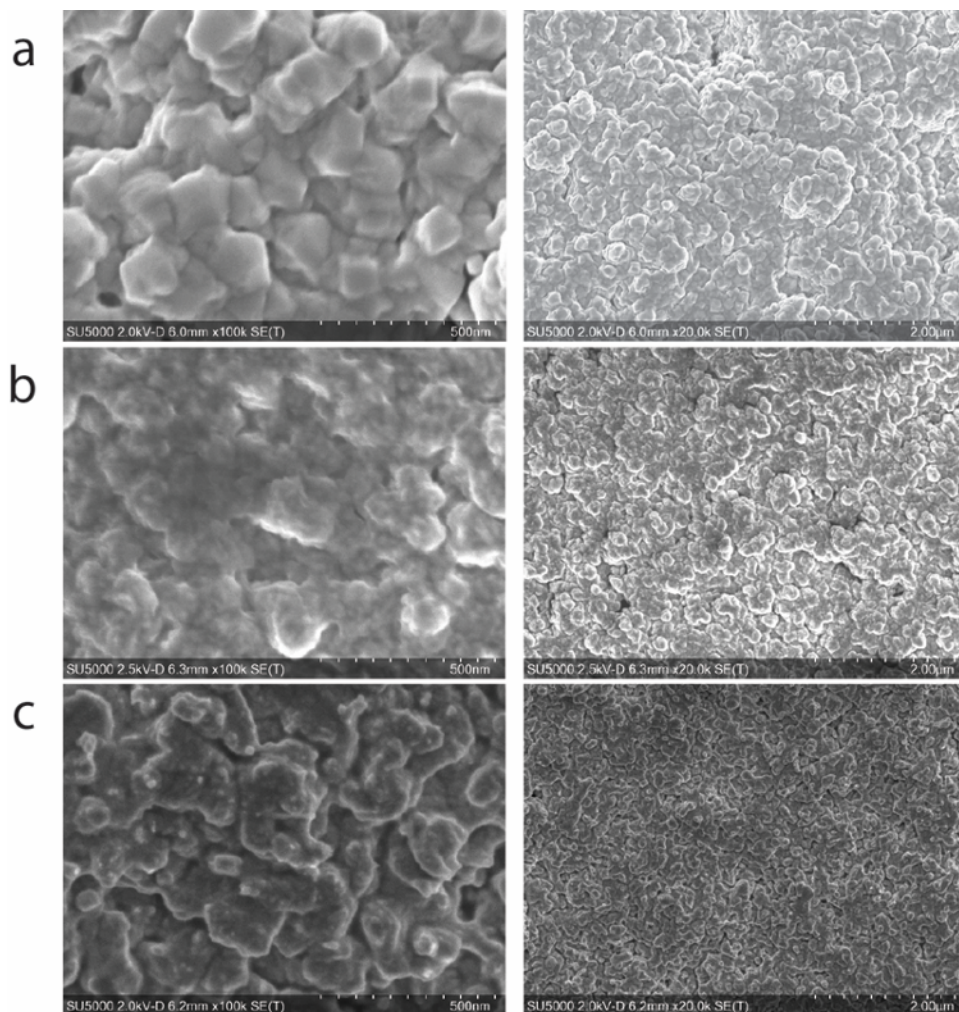


Figure B-6: Morphologies of evaporated and evaporated-etched Al-on-Cu samples on gas diffusion layers. a, Top-view SEM images of the as-evaporated Al-on-Cu samples. b, Top-view SEM images of the evaporated-etched Al-on-Cu catalyst before CO_2 electroreduction. c, Top-view SEM images of the evaporated-etched Al-on-Cu catalyst after 5 hours CO_2 electroreduction in 1 M KOH at an applied current density of 600 mA cm^{-2} in a flow cell.

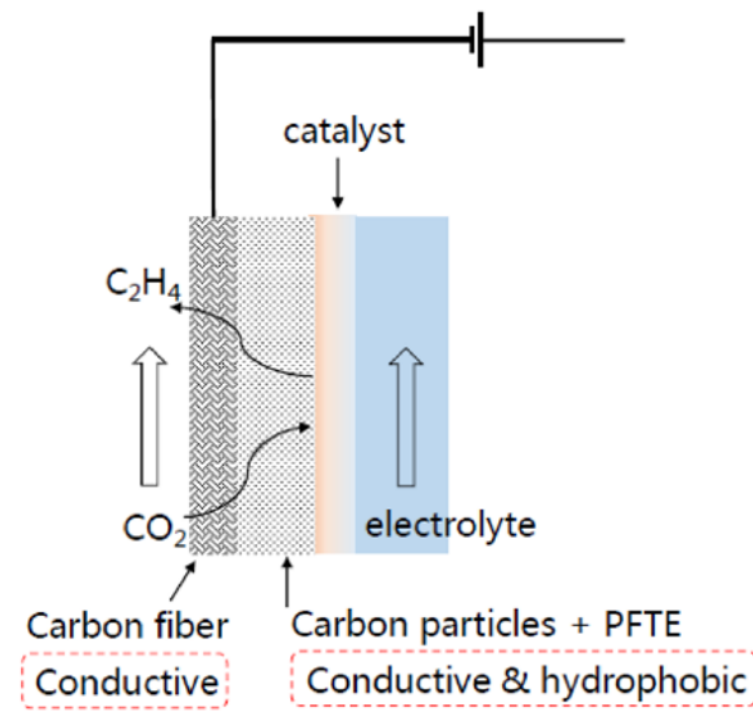


Figure B-7: Schematic of the cathode side in a flow-cell configuration.

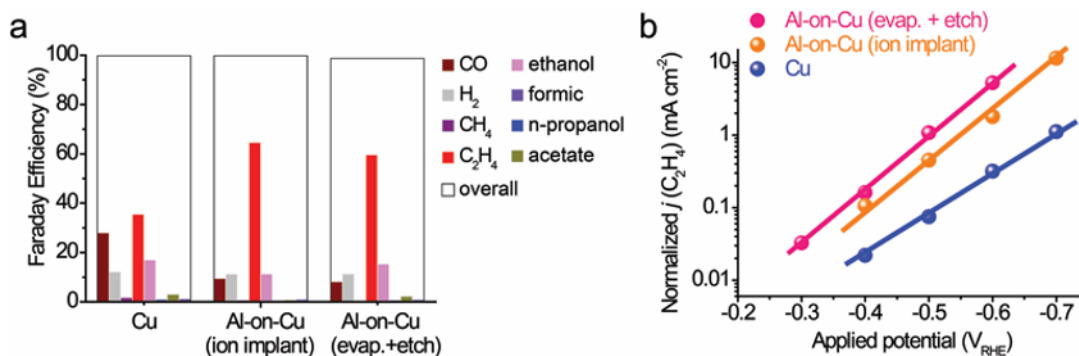


Figure B-8: CO₂ electroreduction performances on pure Cu, ion-implanted Al-on-Cu, and evaporated-etched Al-on-Cu catalysts. a, Faradaic efficiencies of gaseous products on pure Cu, ion-implanted, and evaporated-etched Al-on-Cu obtained from chronopotentiometry tests at an applied current density of 600 mA cm⁻². b, C₂H₄ production partial current density versus potential on pure Cu, ion-implanted, and evaporated-etched Al-on-Cu.

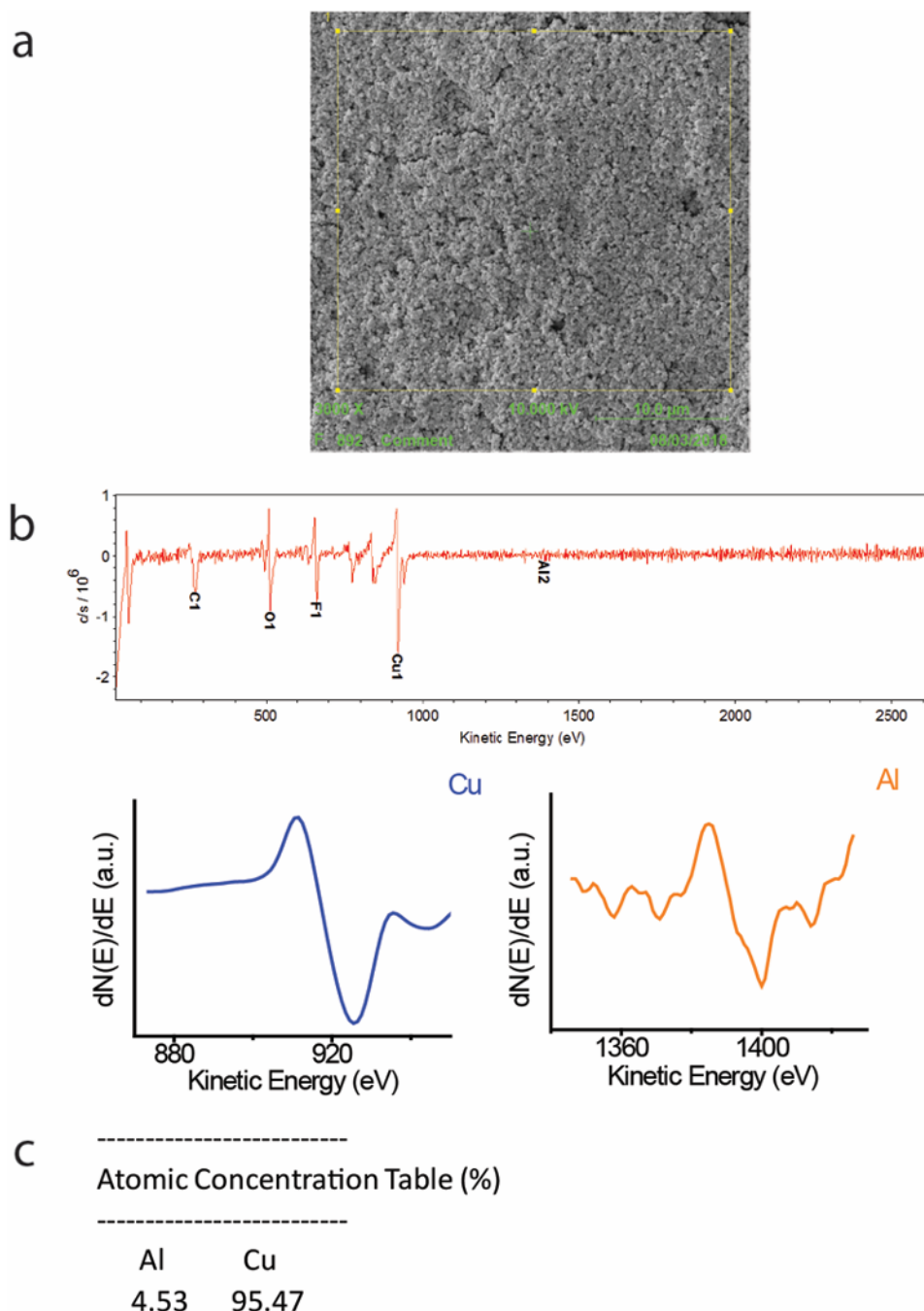


Figure B-9: Auger electron spectroscopic analysis of an ion-implanted Al-on-Cu catalyst. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.

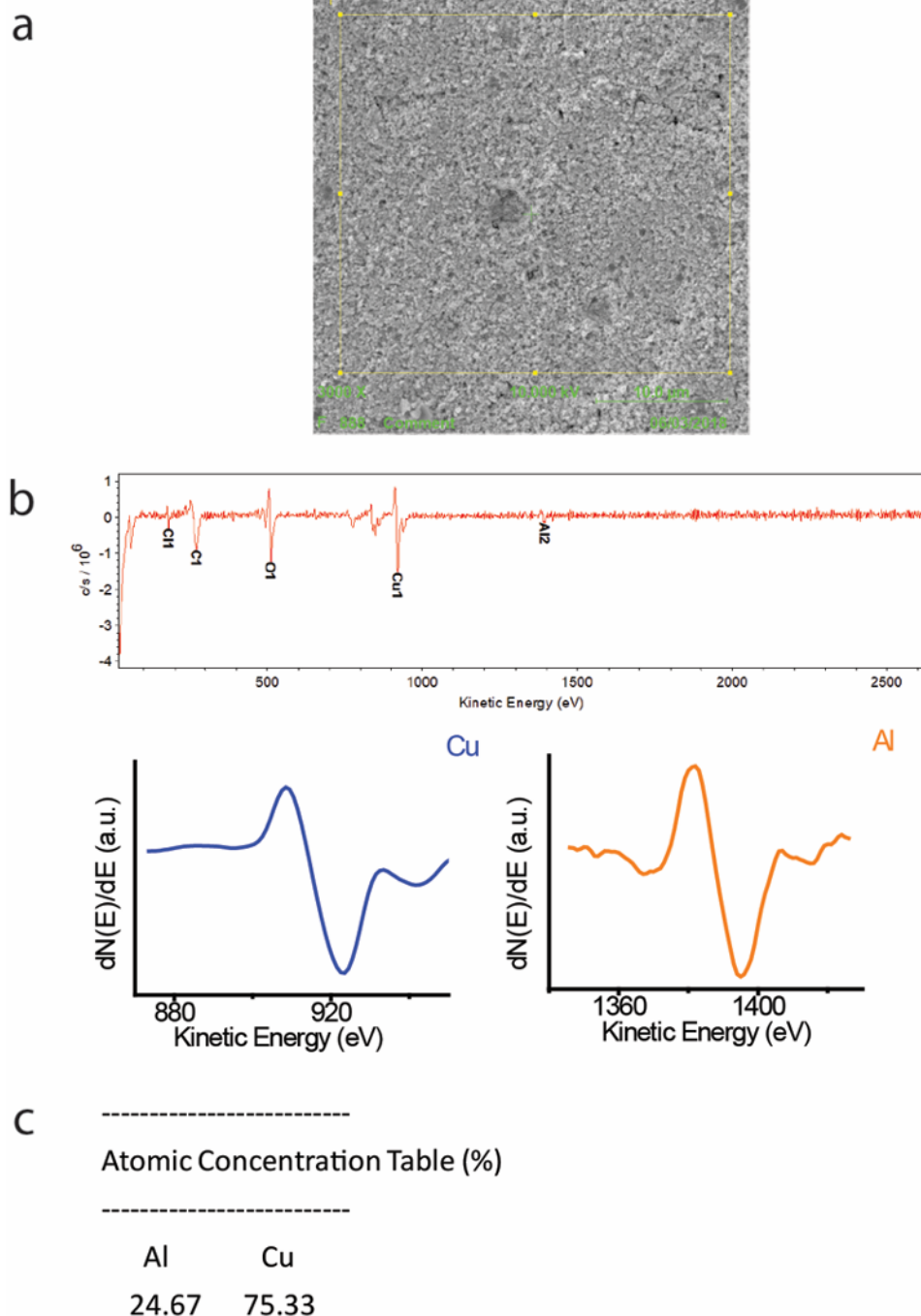


Figure B-10: Auger electron spectroscopic analysis of an evaporated-etched Al-on-Cu catalyst. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.

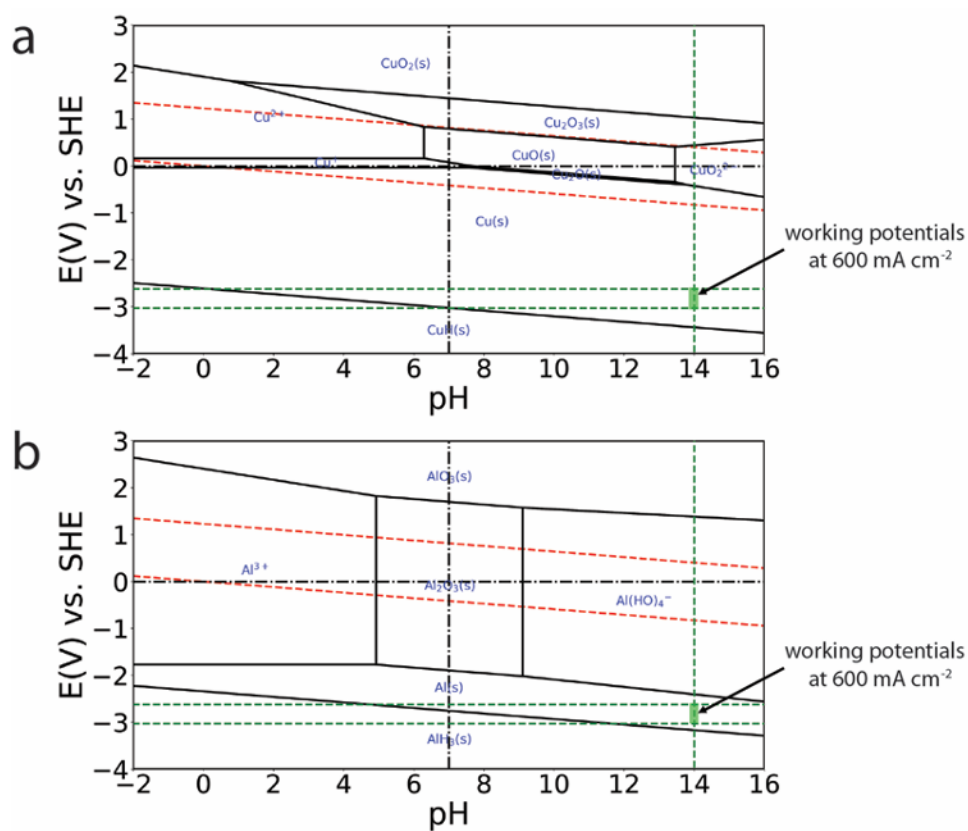


Figure B-11: Pourbaix diagrams. Pourbaix diagrams of a, Cu and b, Al at ionic concentrations of $1 \mu\text{M}$. The potentials versus the standard hydrogen electrode (V vs. SHE) can be converted to the reversible hydrogen electrode scale (V vs. RHE) according to the Nernst equation, $V \text{ vs. RHE} = V \text{ vs. SHE} + 0.059 \times \text{pH}$. According to the Pourbaix diagrams, Cu metal is cathodically protected at potentials more negative than ca. -0.4 VSHE which corresponds to ca. 0.43 VRHE at pH 14 (1 M KOH) and ca. 0.48 VRHE at pH 15 (10 M KOH). Al metal is cathodically protected at potentials more negative than -2.3 VSHE , which corresponds to -1.47 VRHE at pH 14 and -1.41 VRHE at pH 15.

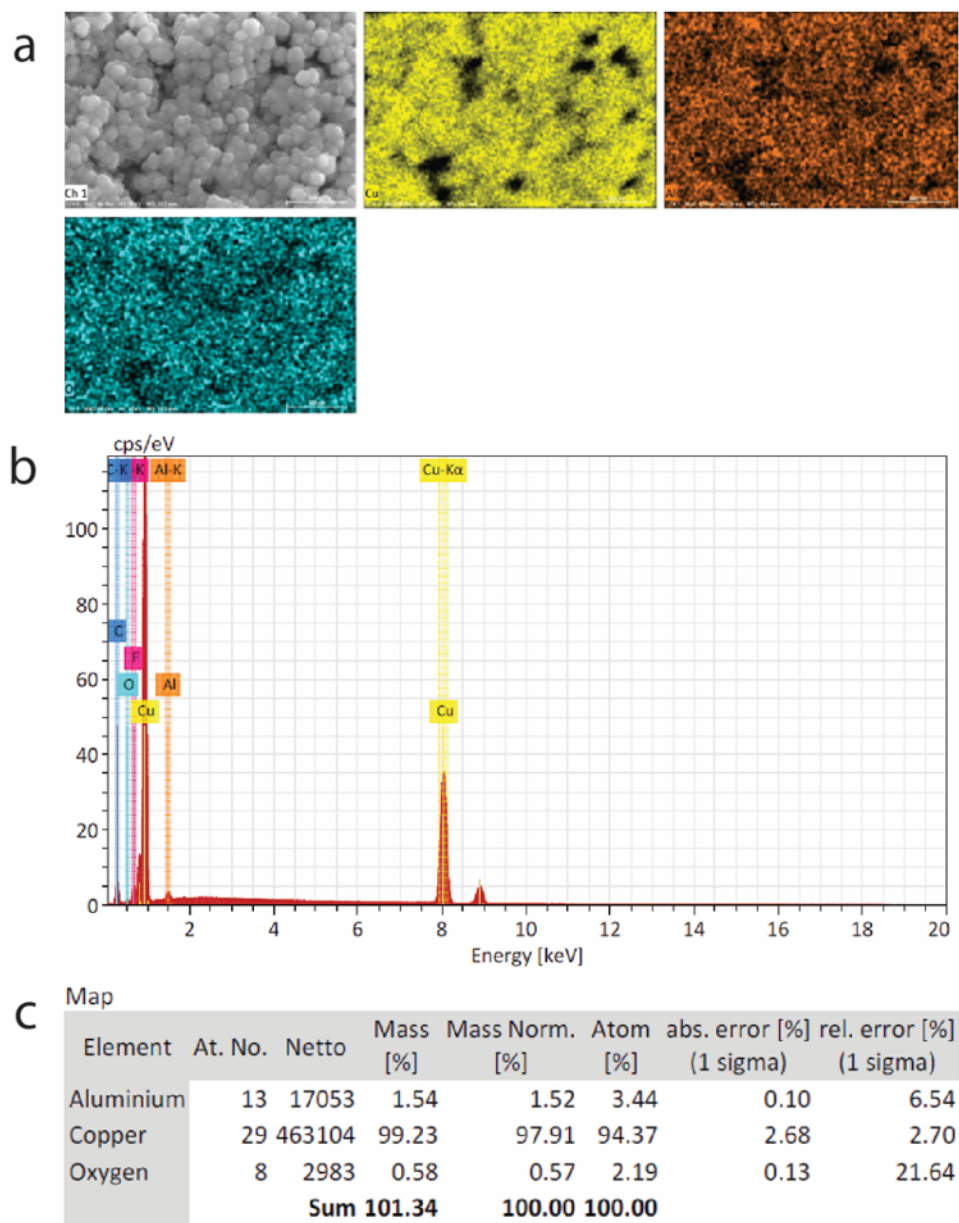


Figure B-12: EDX analysis of an as-prepared ion-implanted Al-on-Cu sample before CO_2 electroreduction. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations.

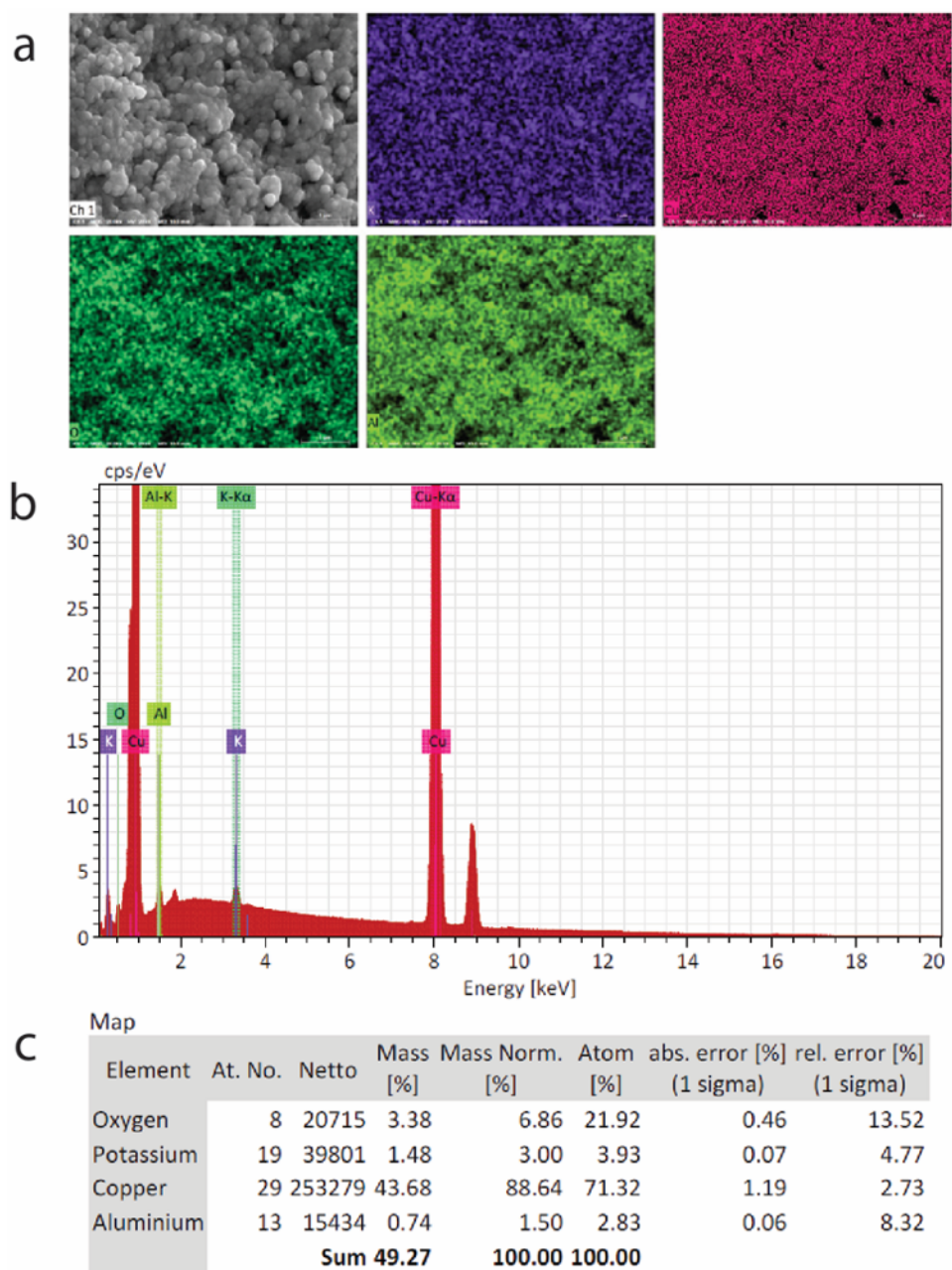


Figure B-13: EDX analysis of an ion-implanted Al-on-Cu sample after 5 hours CO_2 electroreduction at 600 mA cm^{-2} (-1.8 to -2.0 VRHE) in the 1 M KOH electrolyte. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Potassium is observed on the surface after the reaction.

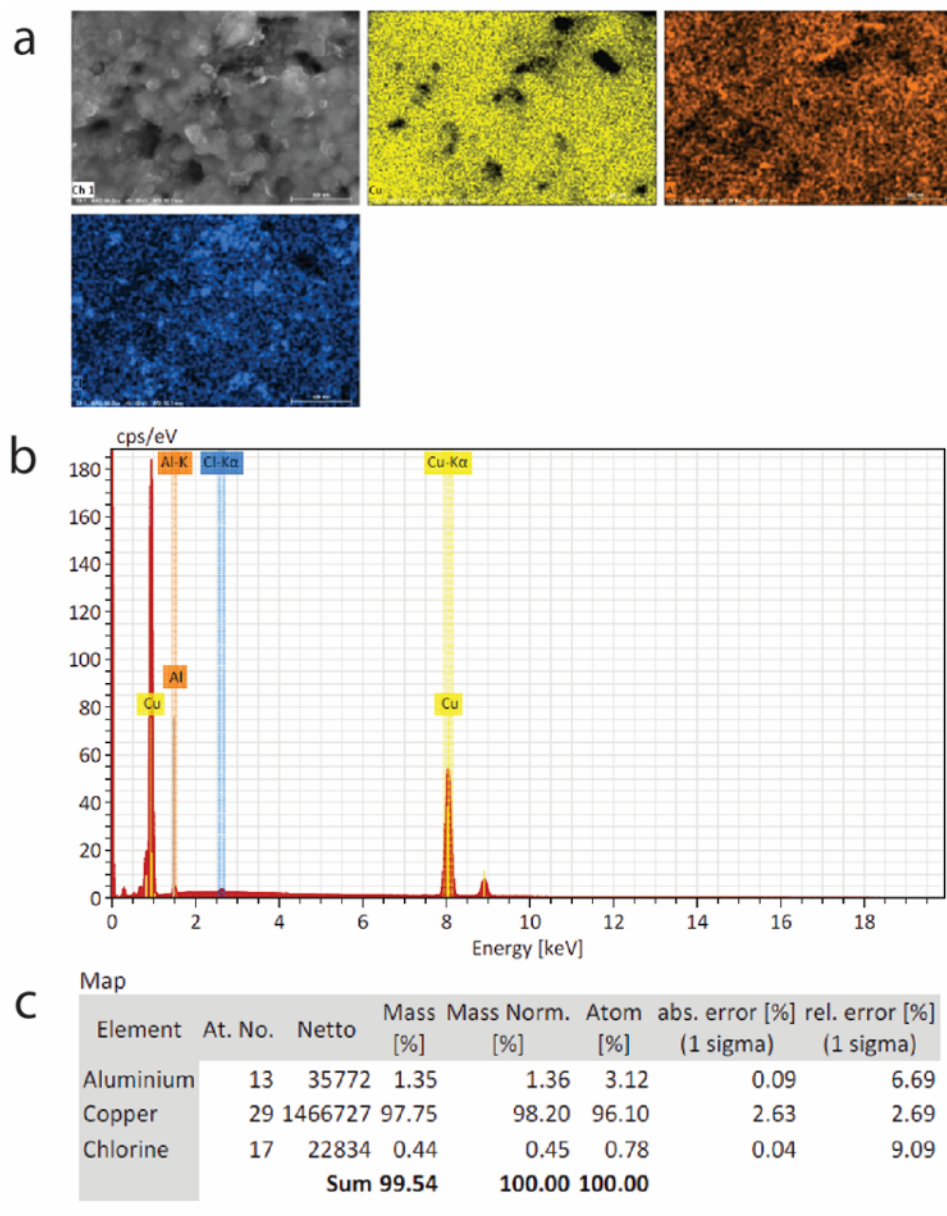


Figure B-14: EDX analysis of an evaporated-etched Al-on-Cu sample before CO₂ electroreduction. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Chlorine is observed on the surface due to the use of HCl solution in the etching process.

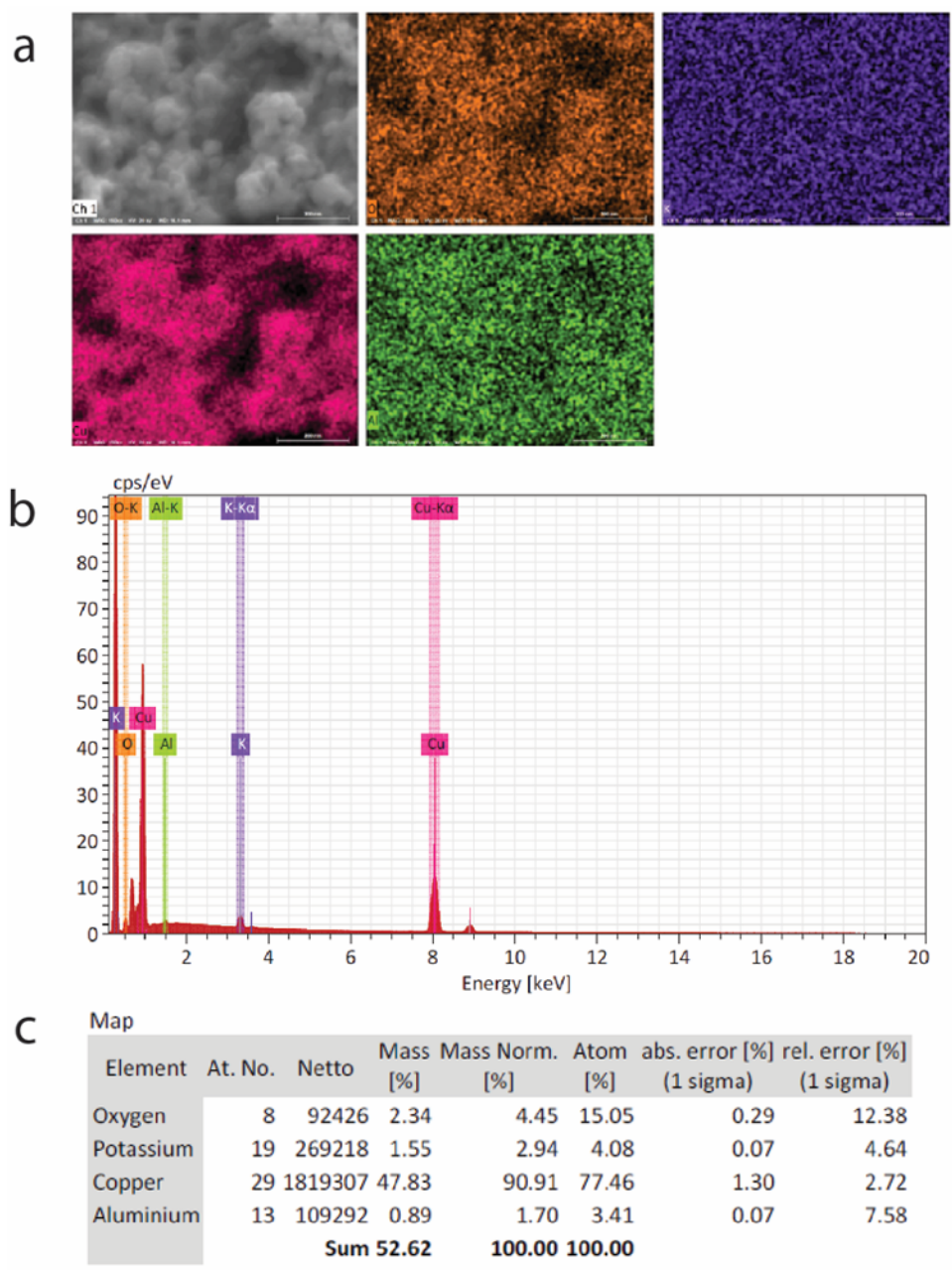


Figure B-15: EDX analysis of an evaporated-etched Al-on-Cu sample after 5 hours CO_2 electroreduction at 600 mA cm^{-2} (-1.8 to -2.0 VRHE) in the 1 M KOH electrolyte. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Potassium was observed on the surface after the reaction.

B.2.8 The stability of Al for CO₂ electroreduction under our testing condition

To further test the electrochemical stability of Al, we performed a control experiment of electroreducing CO₂ using a pure Cu catalyst in 1 M KOH solution in the presence of 1 mM Al(OH)⁴⁻ anions. Energy-dispersive X-ray spectroscopy (EDX) analysis showed ~4% Al after 30 minutes of CO₂ electroreduction at 600 mA cm⁻² (Figure B-16). AES analysis confirmed over 80% Al on the surface, indicating electrodeposition of Al (Figure B-17). Thus, the reverse reaction of dissolving Al into the solution to form Al(OH)⁴⁻ anions was avoided. However, C₂H₄ production activity suffered at the expense of dramatically increased H₂ production (Figure B-18), highlighting the importance of realizing a homogeneous distribution of Al on Cu surface. As another control, we immersed a pure Cu electrode in the same solution of 1 M KOH and 1 mM Al(OH)⁴⁻ for half hour without applying electrical bias. Al concentration was below EDX detection limit (Figure B-19).

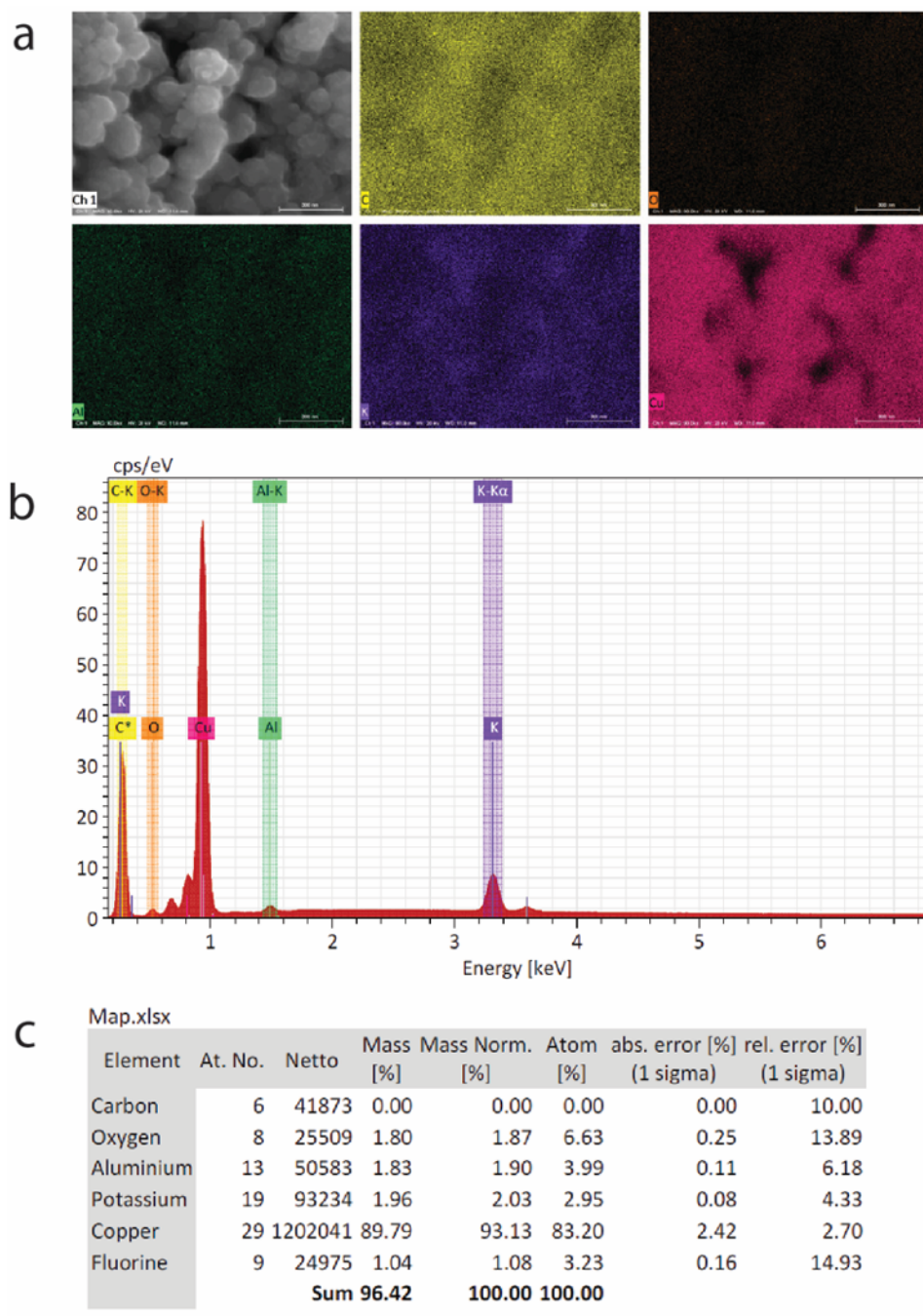


Figure B-16: EDX analysis of a pure Cu catalyst after half-hour CO_2 electroreduction in 1 M KOH with 1 mM $\text{Al}(\text{OH})_4^-$. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Fluorine was from the PTFE/carbon layer in the gas diffusion electrode. Al was electrodeposited on Cu at an applied current density of 600 mA cm^2 . Therefore, dissolving Al into solution was avoided.

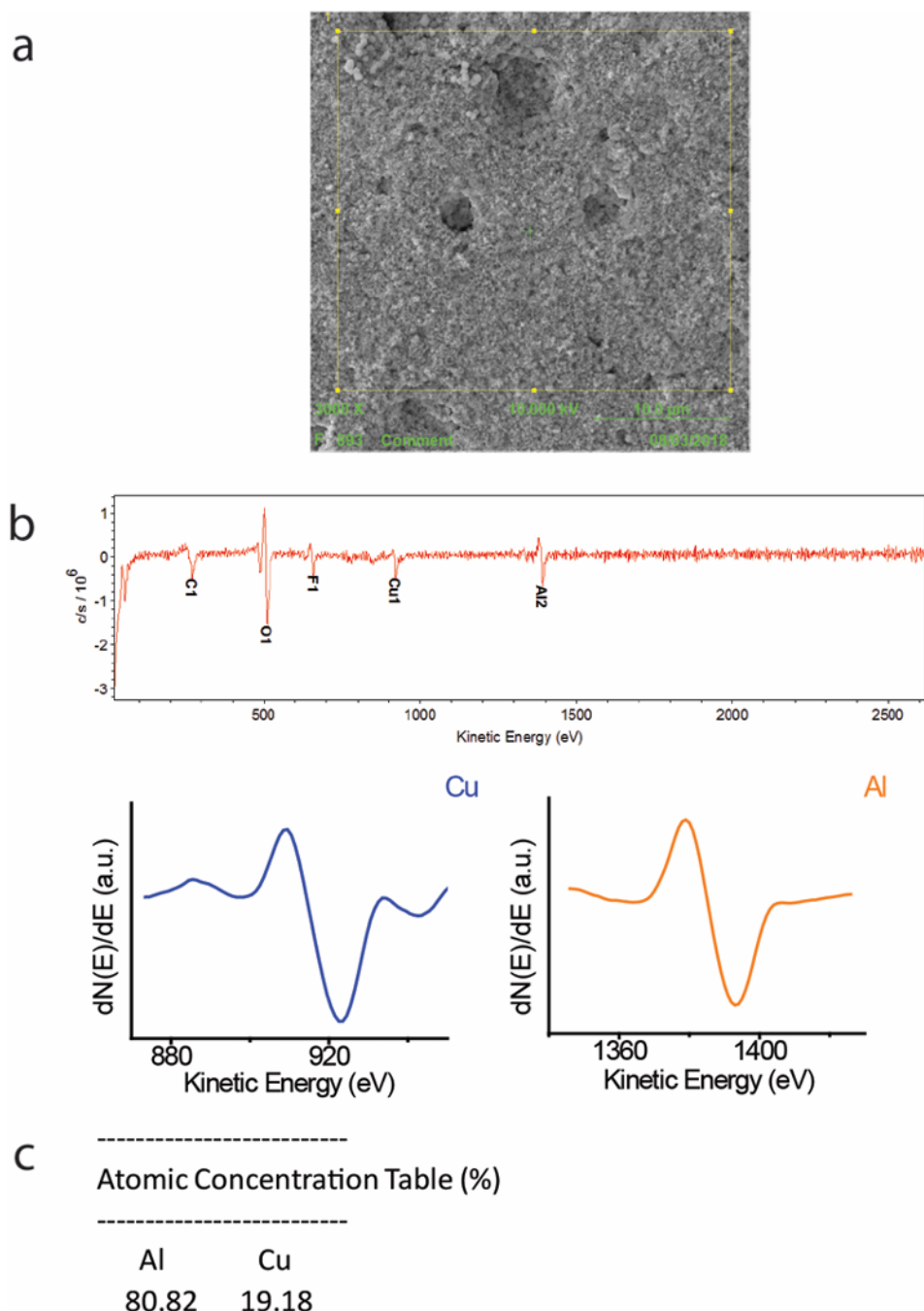


Figure B-17: Auger electron spectroscopic analysis of a pure Cu catalyst after 0.5-hour CO_2 electroreduction at 600 mA cm^{-2} in 1 M KOH with 1 mM $\text{Al}(\text{OH})^-$. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.

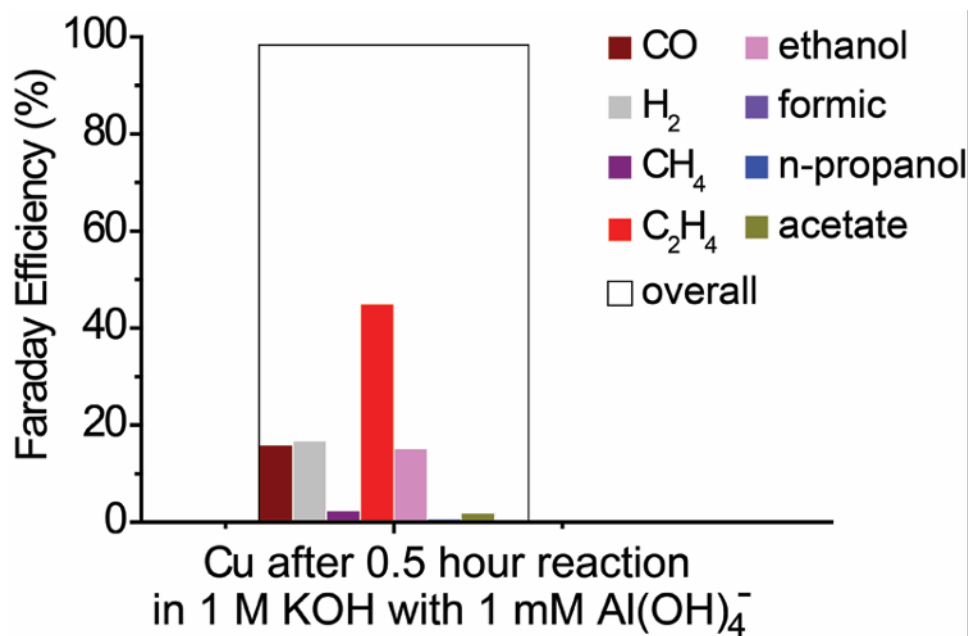


Figure B-18: CO₂ electroreduction performance. Faradaic efficiencies for gaseous products with a pure Cu catalyst at a constant current density of 600 mA cm² in 1 M KOH with the presence of 1 mM Al(OH)₄⁻ obtained from Chronopotentiometry tests.

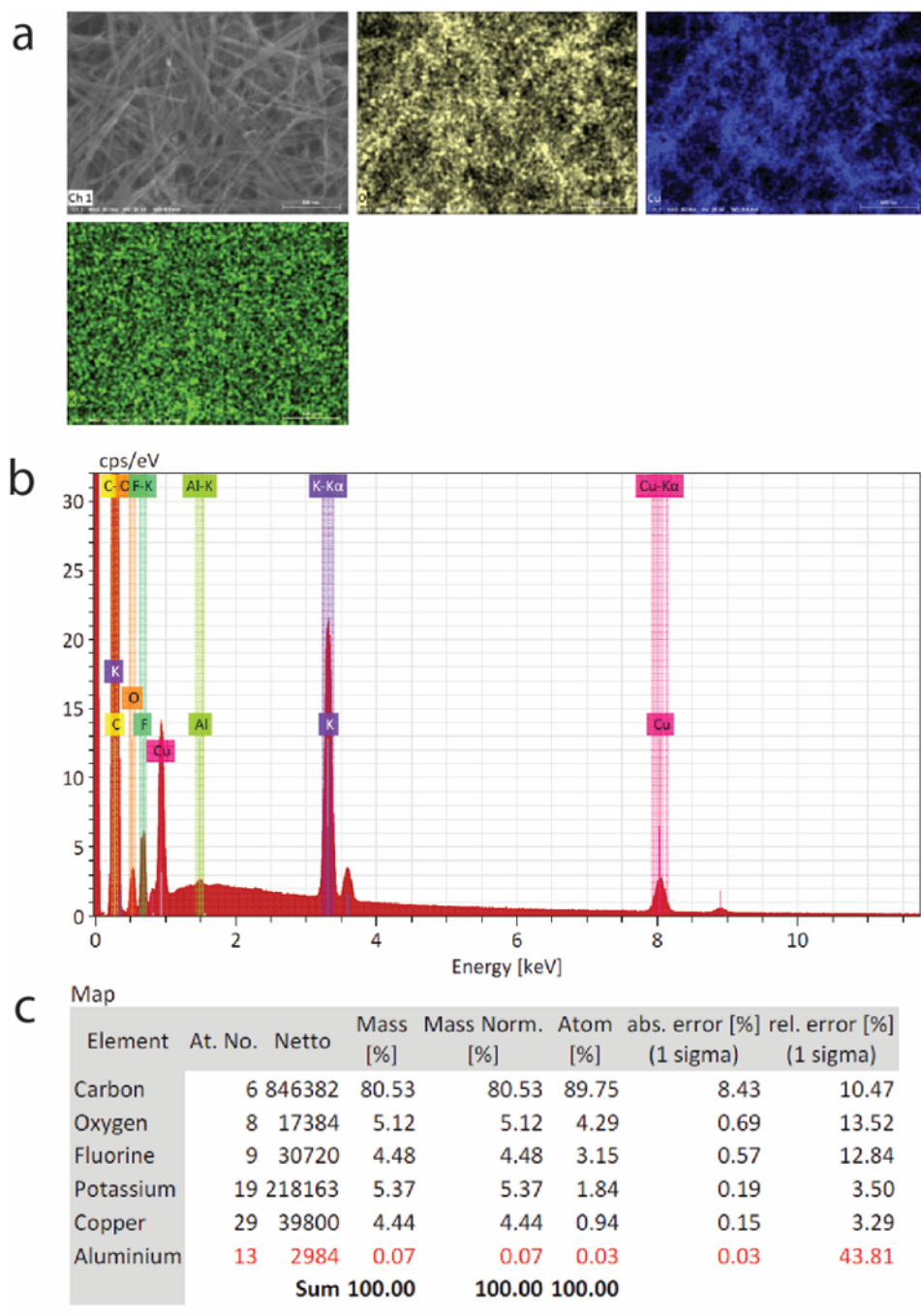


Figure B-19: EDX analysis of a pure Cu after immersing in 1 M KOH with 1 mM $\text{Al}(\text{OH})_4^-$ for 0.5 hour. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. We manually selected Al which was automatically marked in red by EDX software (Esprit 2.1) with a large error of 43.81%, indicating Al was actually out of the EDX detection limit. In SEM image in Figure B-14a, the morphology was also changed. Most of Cu dissolved into 1 M KOH leaving thin $\text{Cu}(\text{OH})_2$ nanowires to be dissolved.

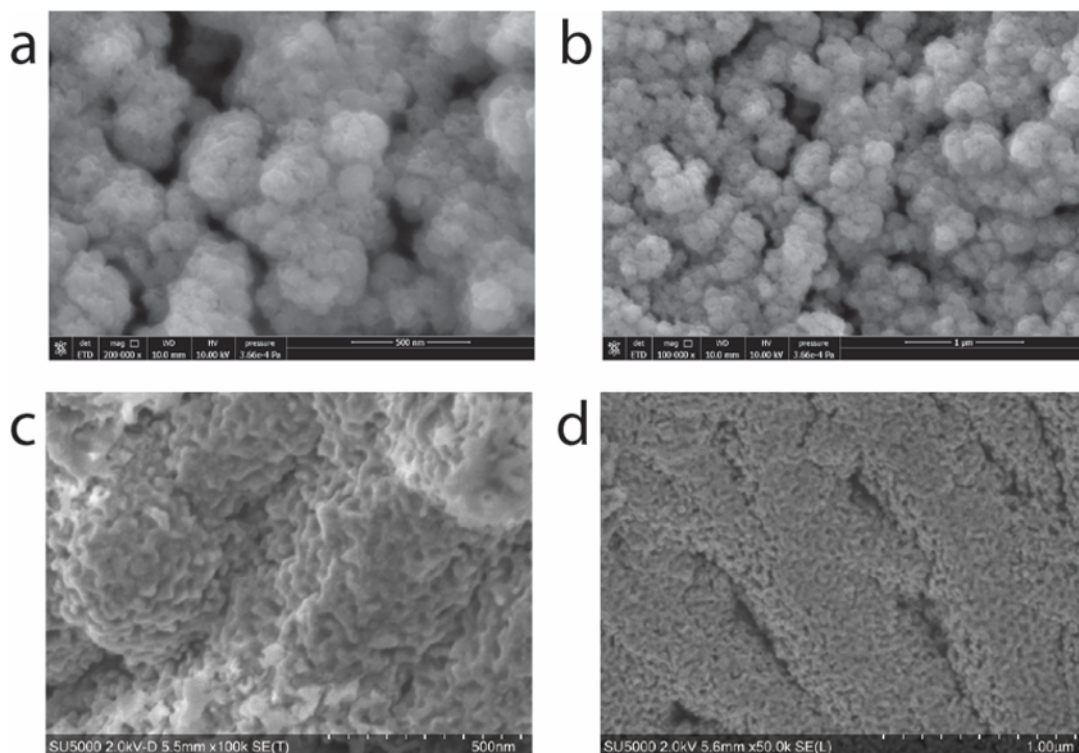


Figure B-20: Top-view SEM images of the de-alloyed Cu-Al. a-b, De-alloyed Cu-Al catalyst prepared on a C-GDL substrate by physical evaporation and chemical etching. c-d, De-alloyed Cu-Al catalyst prepared on a PTFE substrate by co-sputtering and chemical etching.

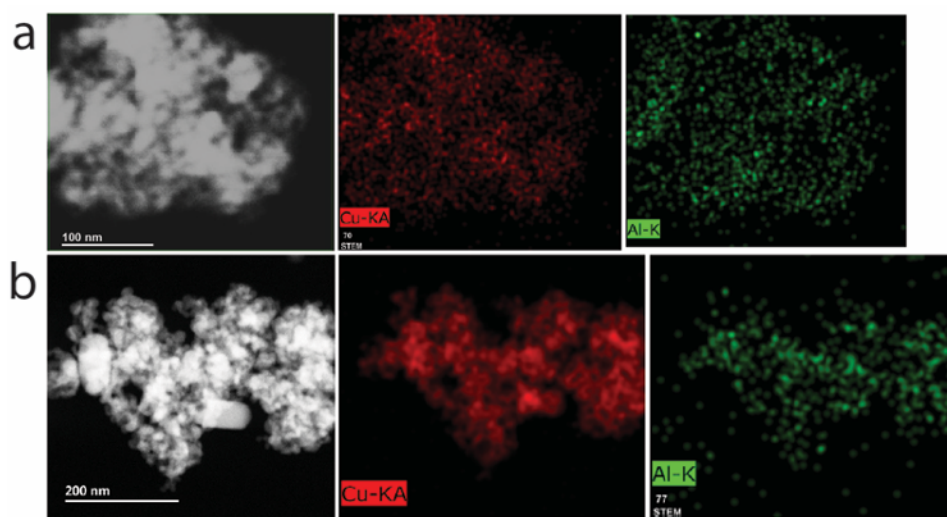


Figure B-21: EDX analyses in TEM of de-alloyed Cu-Al catalysts. a, as-prepared de-alloyed Cu-Al catalyst. b, de-alloyed Cu-Al catalyst after 5 hours CO_2 electroreduction in 1 M KOH at 600 mA cm^2 in a flow-cell configuration.

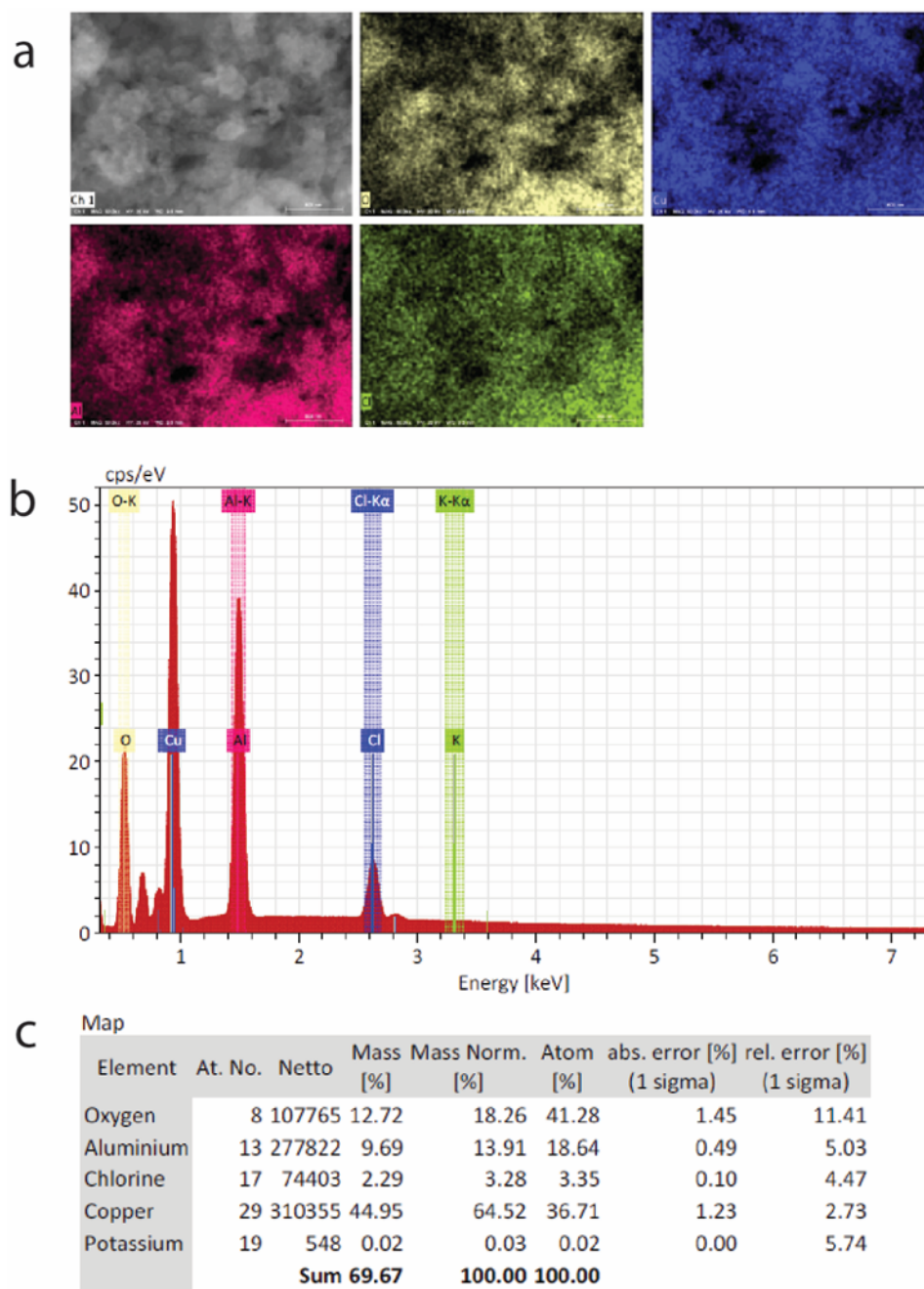


Figure B-22: EDX analysis in SEM of the as-prepared de-alloyed Cu-Al catalyst before CO₂ electroreduction. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Chlorine was detected because of the use of a 5 wt% HCl solution in the etching process to prepare a de-alloyed Cu-Al catalyst.

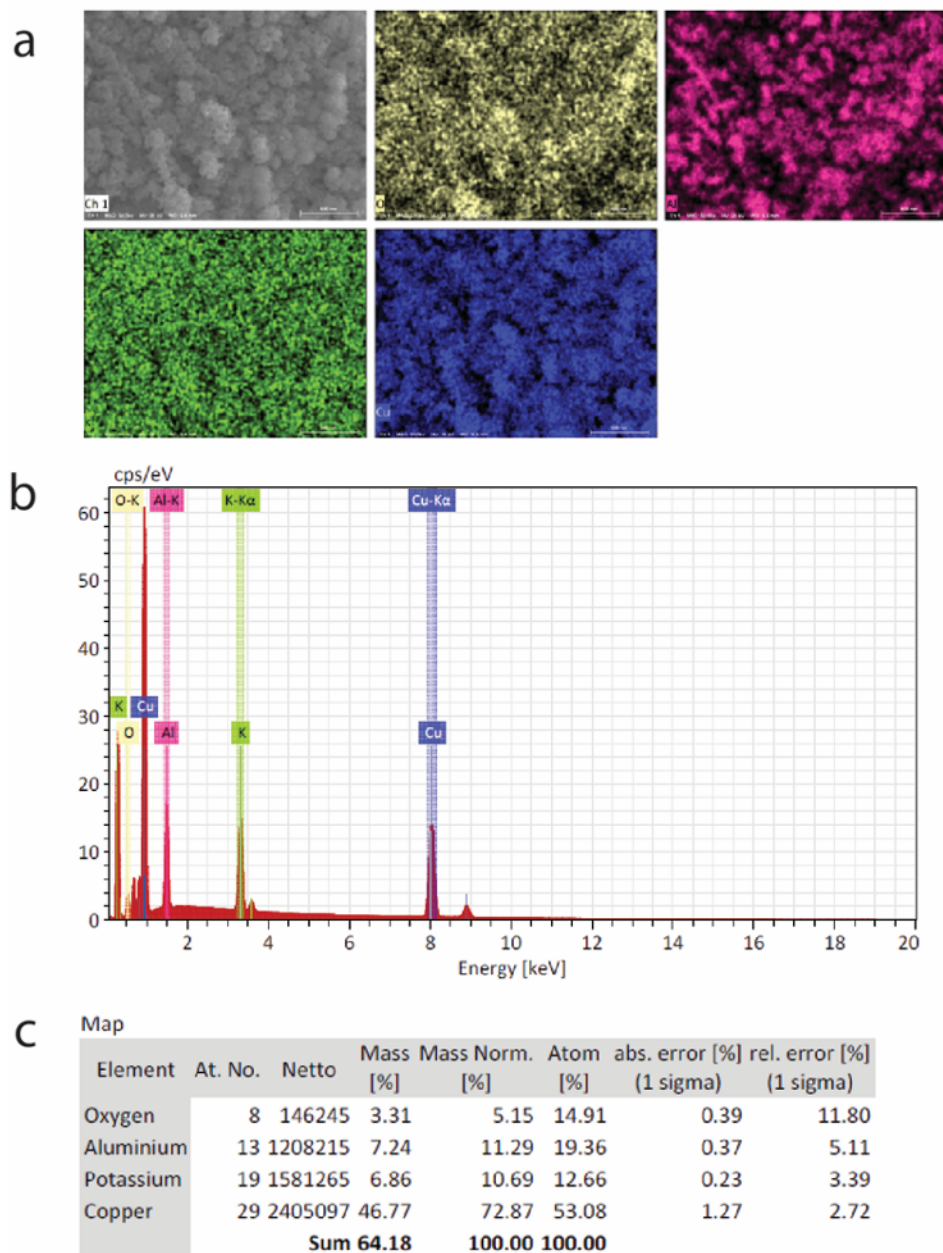


Figure B-23: EDX analysis in SEM of a de-alloyed Cu-Al catalyst after 5 hours CO_2 electroreduction at 600 mA cm^2 (-1.8 to -2.0 VRHE) in 1 M KOH. a, EDX mapping. b, EDX spectrum. c, Elemental concentrations. Potassium was detected because of the use of KOH electrolyte in CO_2 electroreduction.

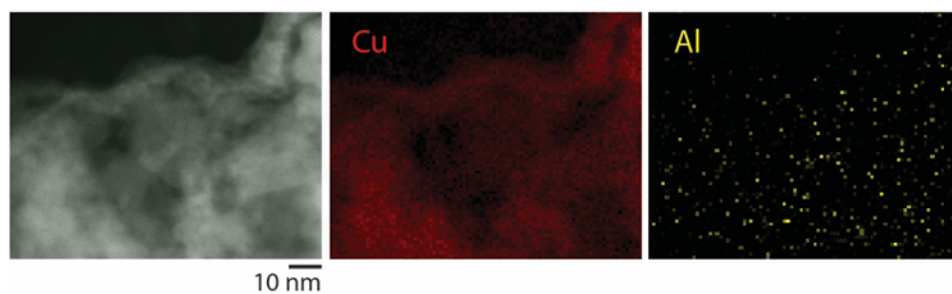


Figure B-24: HAADF-STEM with EELS mapping images of Cu and Al of a de-alloyed Cu-Al catalyst after 5 hours CO_2 electroreduction. A nanoporous structure was observed with the de-alloyed Cu-Al after the CO_2 electroreduction.

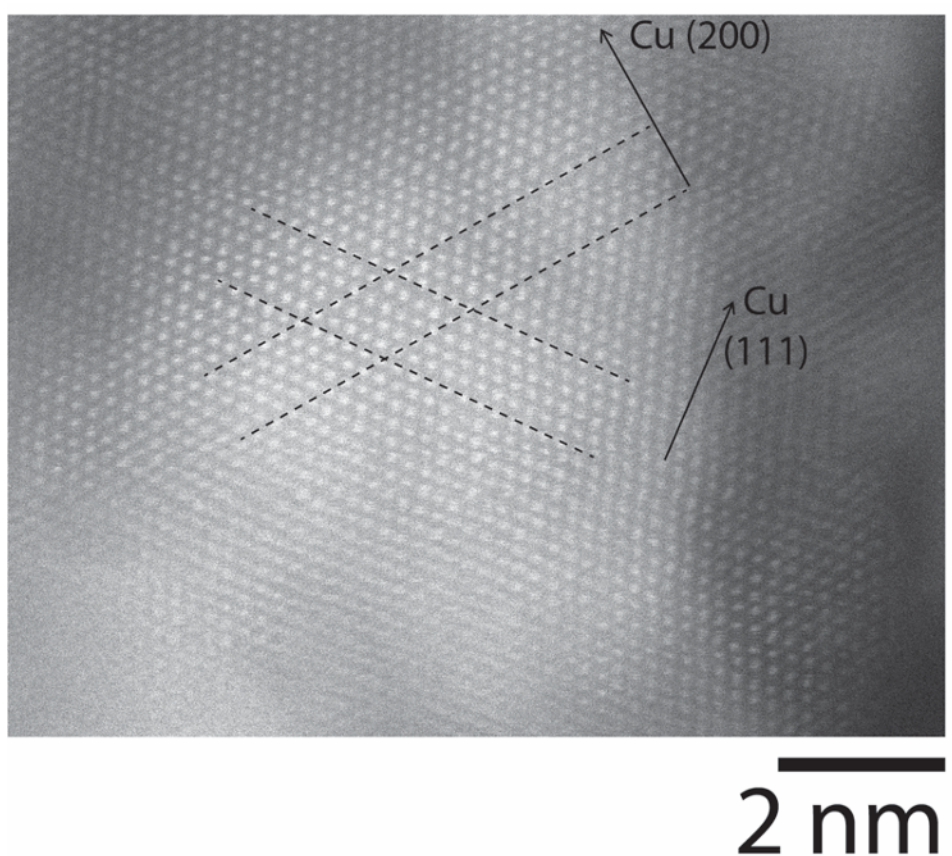


Figure B-25: STEM-ADF image of the de-alloyed Cu-Al catalyst. The observation was performed in the same area where we performed EELS analysis (Figure 4-2c)

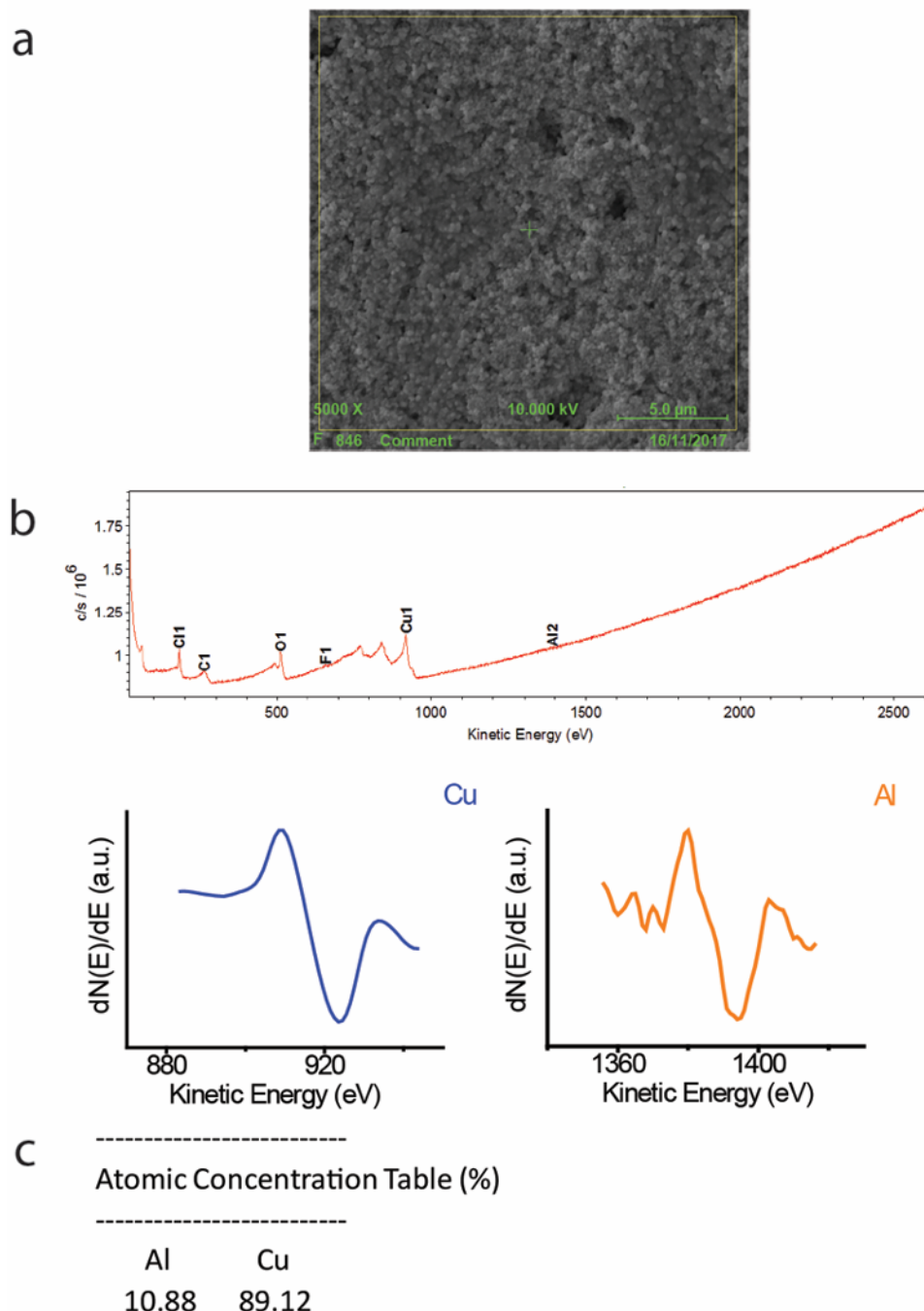


Figure B-26: Auger electron spectroscopic analysis of an as-prepared de-alloyed Cu-Al catalyst. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.

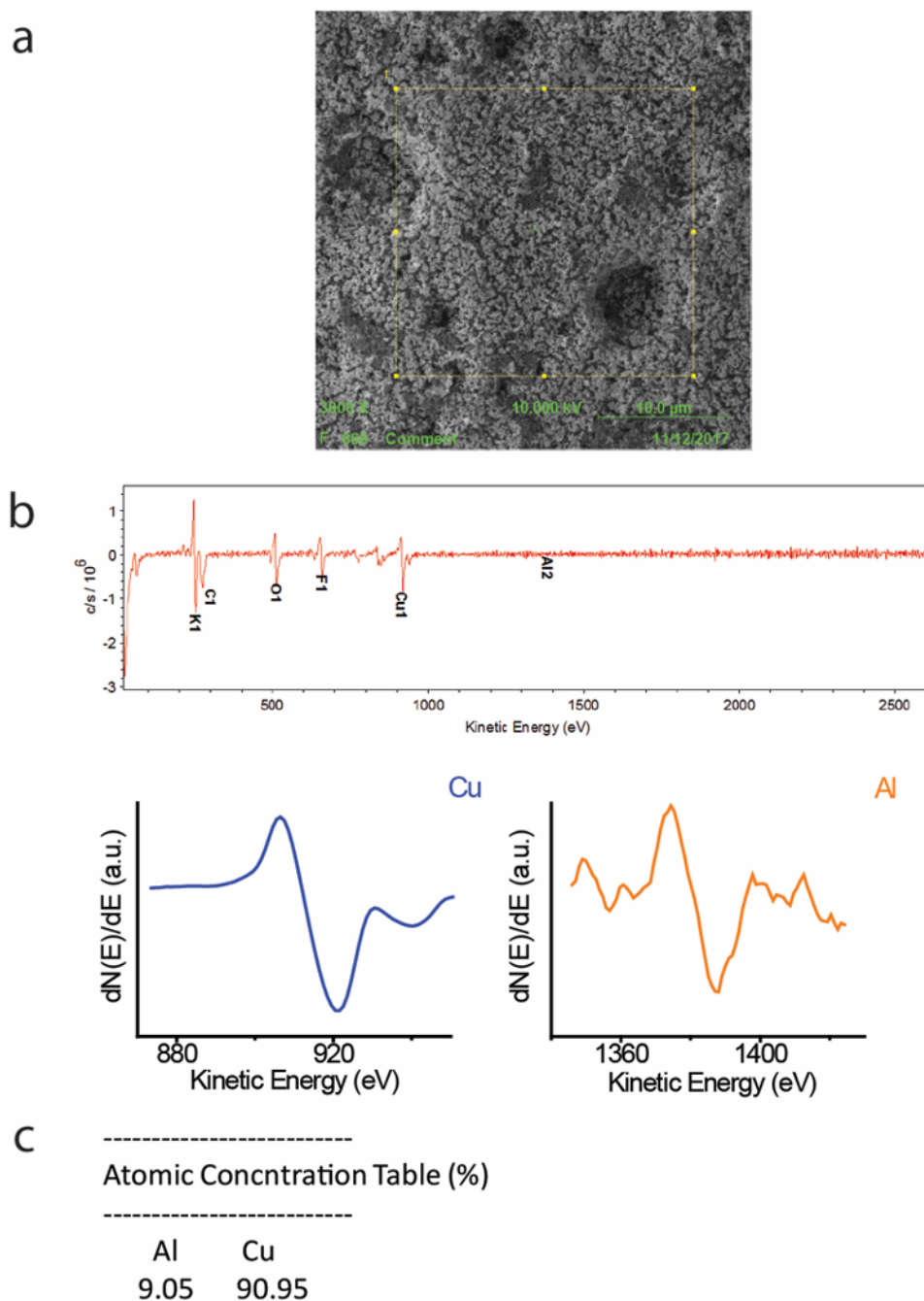


Figure B-27: Auger electron spectroscopic analysis of a de-alloyed Cu-Al catalyst after 5 hours CO_2 electroreduction at 600 mA cm^{-2} in 1 M KOH. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al.

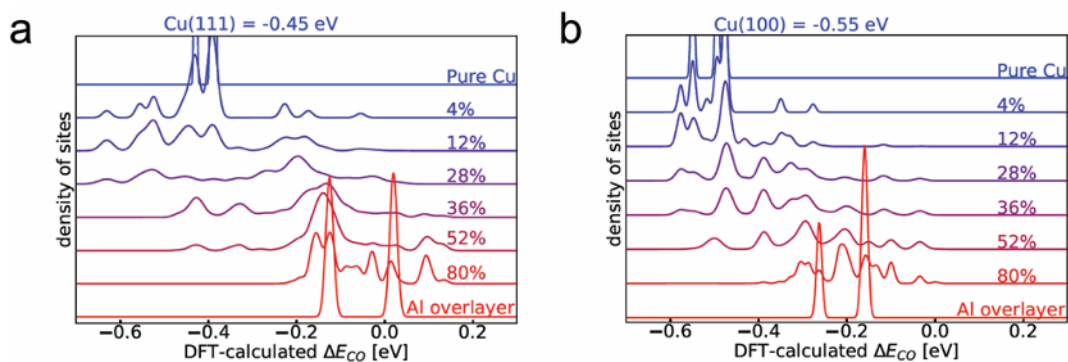


Figure B-28: ML and DFT analyses of the (111) and (100) facets across varying ratios of Al concentrations. a, b, The distribution of ΔE_{CO} values for the adsorption sites on Cu (111) and (100) surfaces with different amounts of Al replacement on the top layer of atoms.

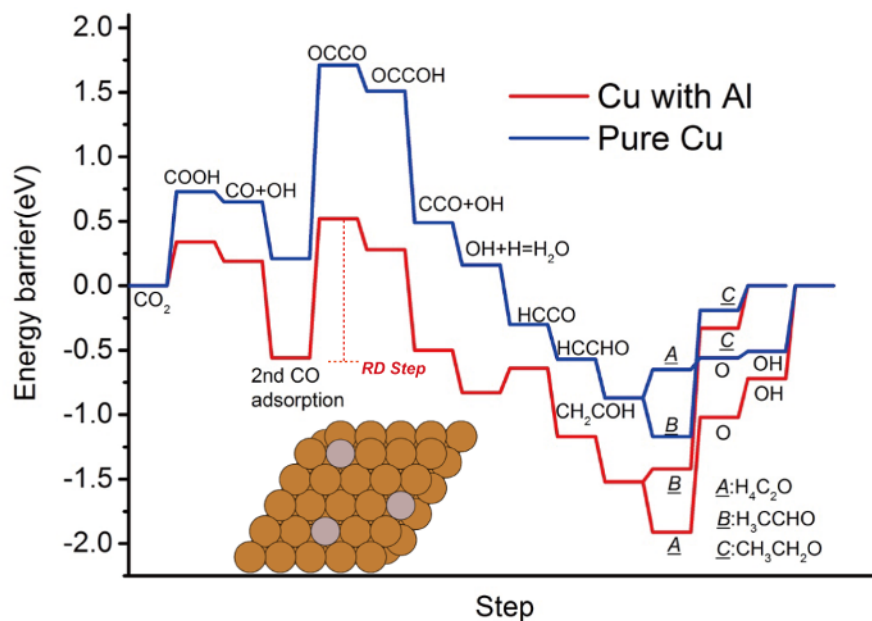


Figure B-29: Reaction Gibbs free energy diagram. Reaction Gibbs free energy diagram from adsorbed CO_2 to OC_2H_4 , an intermediate to C_2H_4 , and to CHOCH_3 , an intermediate to $\text{C}_2\text{H}_5\text{OH}$, on the pure Cu (111) surface (blue lines) and ML predicted 12% Al incorporated Cu (111) surface (red lines). The ML predicted structure of 12% Al incorporated Cu (111) is inserted in the lower left.

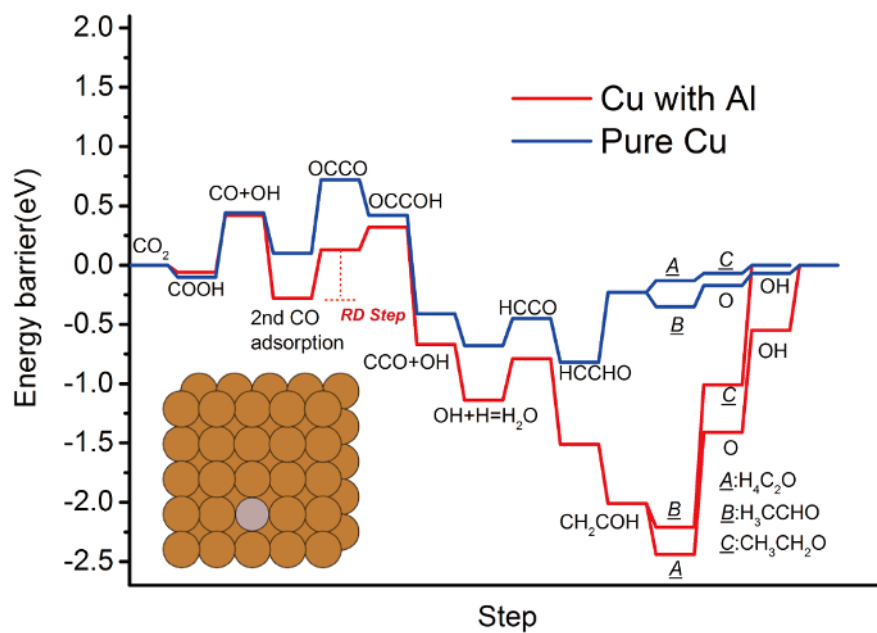


Figure B-30: Reaction Gibbs free energy diagram. Reaction Gibbs free energy diagram from adsorbed CO₂ to OC₂H₄, an intermediate to C₂H₄, and to CHOCH₃, an intermediate to C₂H₅OH, on the pure Cu (100) surface (blue lines) and ML predicted 4% Al incorporated Cu (100) surface (red lines). The ML predicted structure of 4% Al incorporated Cu (111) is inserted in the lower left.

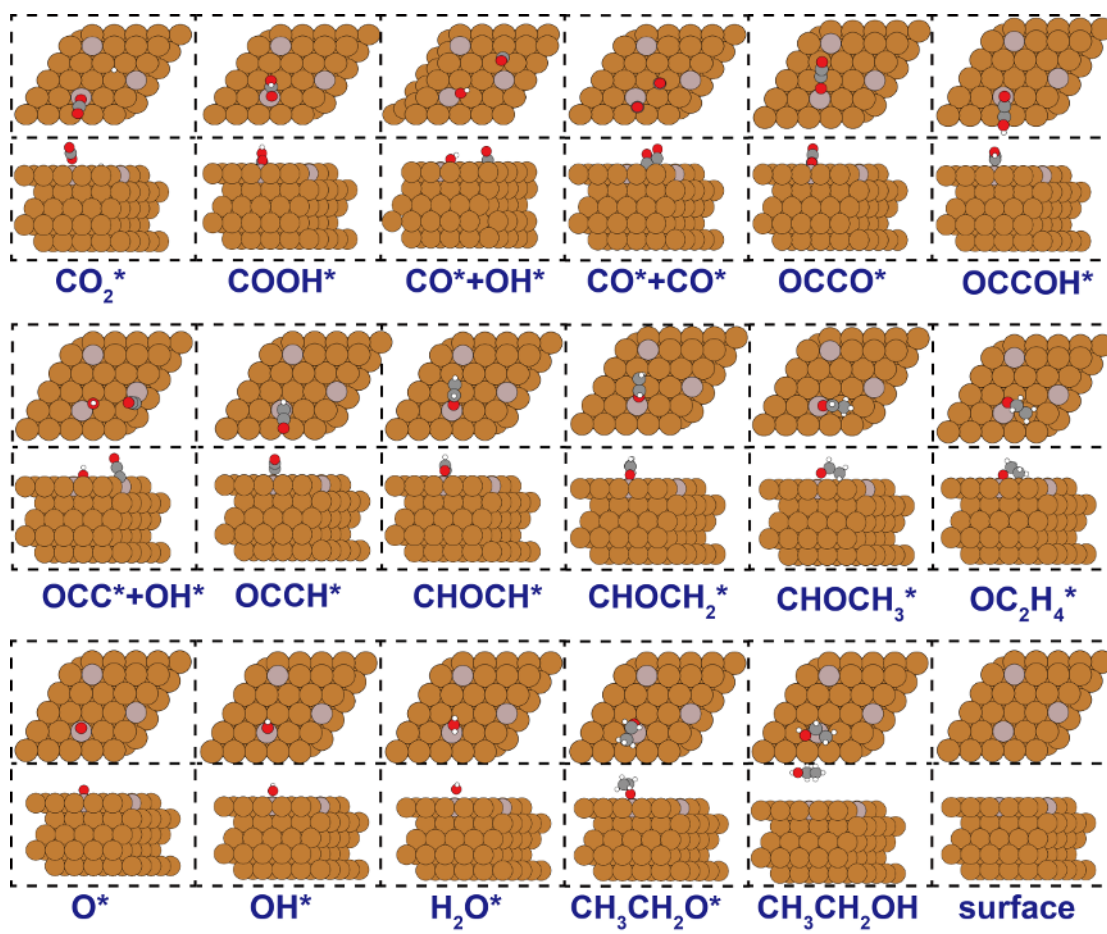


Figure B-31: A possible pathway with intermediates in the CO_2 electroreduction. Optimized geometries of CO_2 reduction intermediates from adsorbed CO_2^* to OC_2H_4 and CHOCH_3 over the 12% Al incorporated Cu (111) s

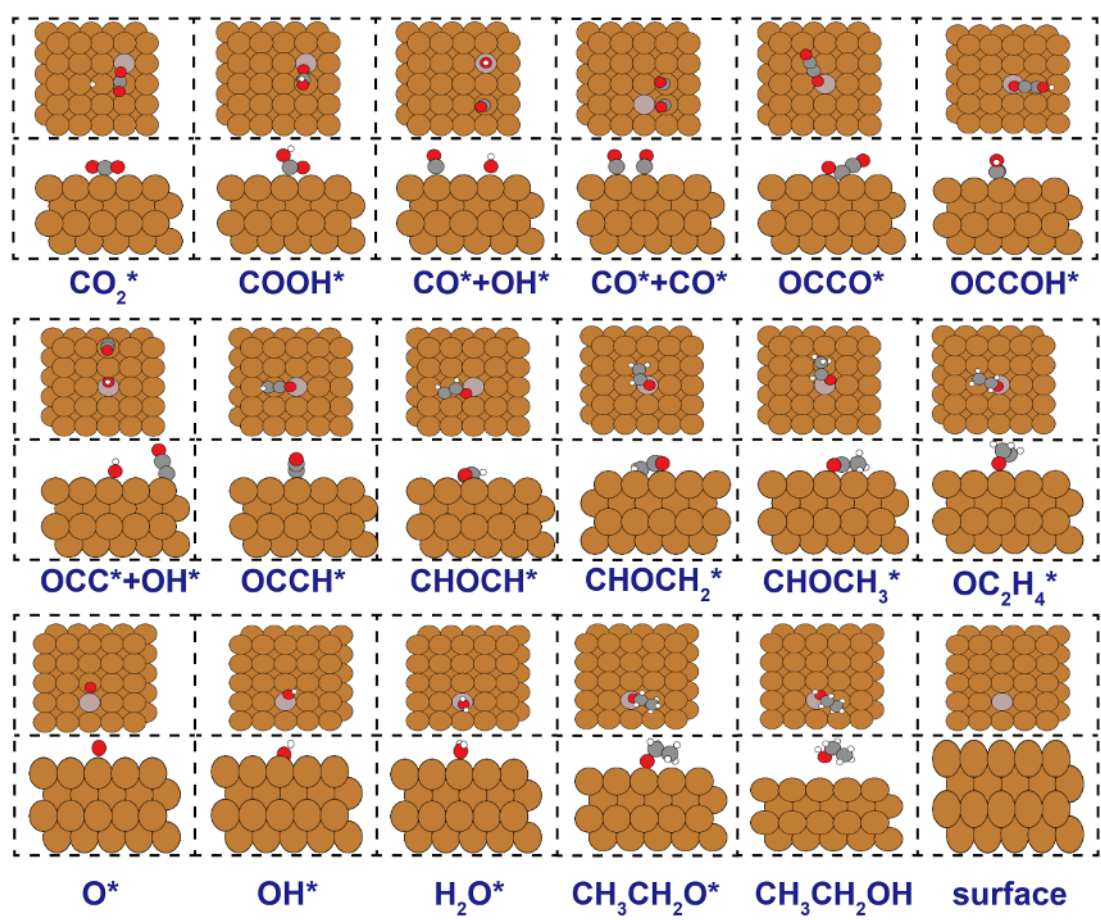


Figure B-32: A possible pathway with intermediates in the CO_2 electroreduction. Optimized geometries of CO_2 reduction intermediates from adsorbed CO_2^* to OC_2H_4 and CHOCH_3 over the 4% Al incorporated Cu (100) s

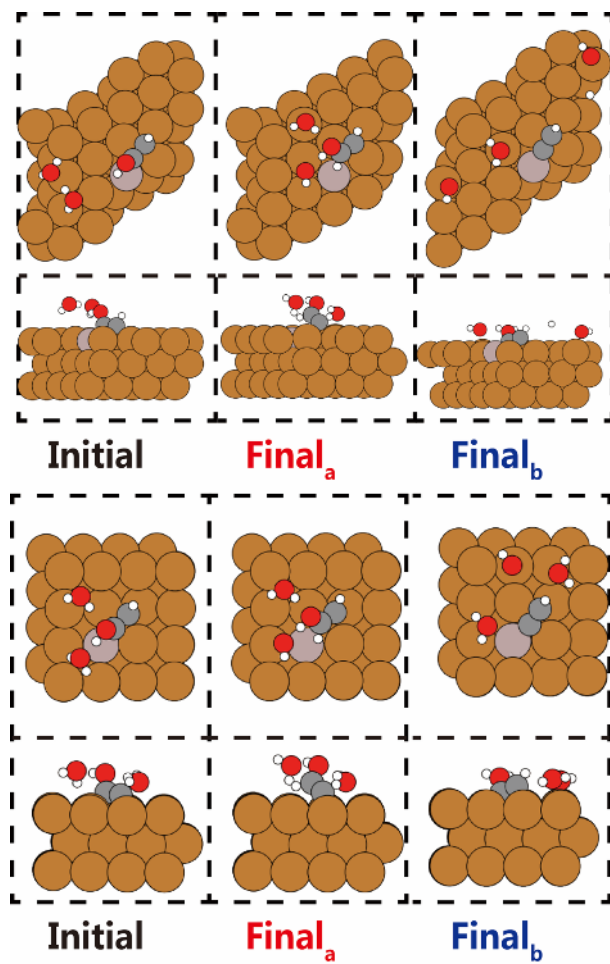


Figure B-33: Initial and final intermediates of forming ethanol (red) and C_2H_4 (blue) on the 12% Al incorporated Cu (111) and 4% Al incorporated (100) surfaces. A possible pathway with intermediates in the CO_2 electroreduction. According to the previous work reported by Goddard et al.[182], surface water may play an important role to determine the key intermediates that branch toward ethanol and C_2H_4 production especially at high pH conditions. We use their method to calculate the Gibbs free energy barriers of the key intermediates on the 12% incorporated Cu (111) and 4% Al incorporated Cu (100) surfaces. The energy barriers for forming ethanol are 0.343 eV and 0.223 eV larger than that for forming C_2H_4 on the Al incorporated Cu (111) and Cu (100) surfaces, respectively.

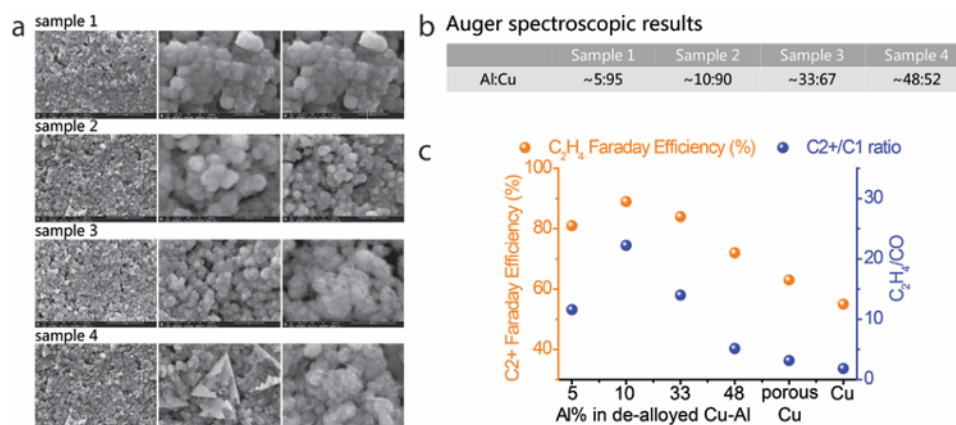


Figure B-34: De-alloyed Cu-Al with different surface Al concentrations. a, SEM images of de-alloyed Cu-Al with different etching time. b, Al concentrations on surfaces determined by surface-sensitive Auger spectroscopic analysis. Surface Al concentrations could be roughly varied from 4–28% by controlling the etching time. c, Faradaic efficiencies of C_2+ and C_2+/C_1 ratio with de-alloyed catalysts at different surface Al concentrations.

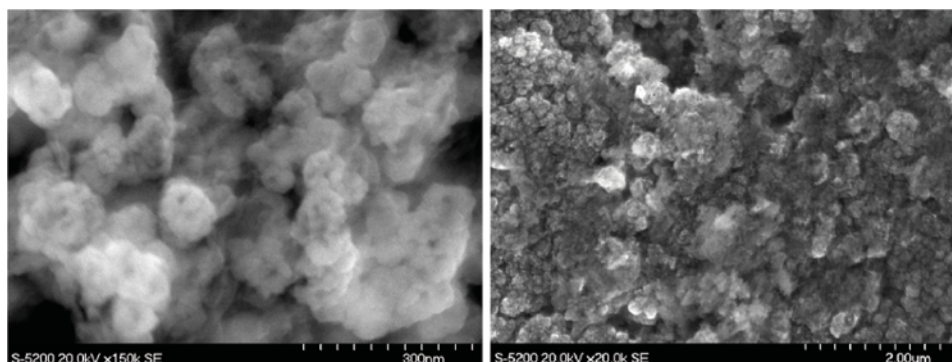


Figure B-35: SEM images of a de-alloyed Cu-Al catalyst soaked in 10 mM $CuCl_2$ for 10 minutes. 10 minutes is an optimal time that we used to roughly maintain nanoporosity while replacing surface Cu with Al by the displacement reaction of $2 Al + 3 CuCl_2 \rightarrow 2 AlCl_3 + 3 Cu$. We also found that nanopores were gradually blocked by the precipitated Cu with a prolonged $CuCl_2$ treatment.

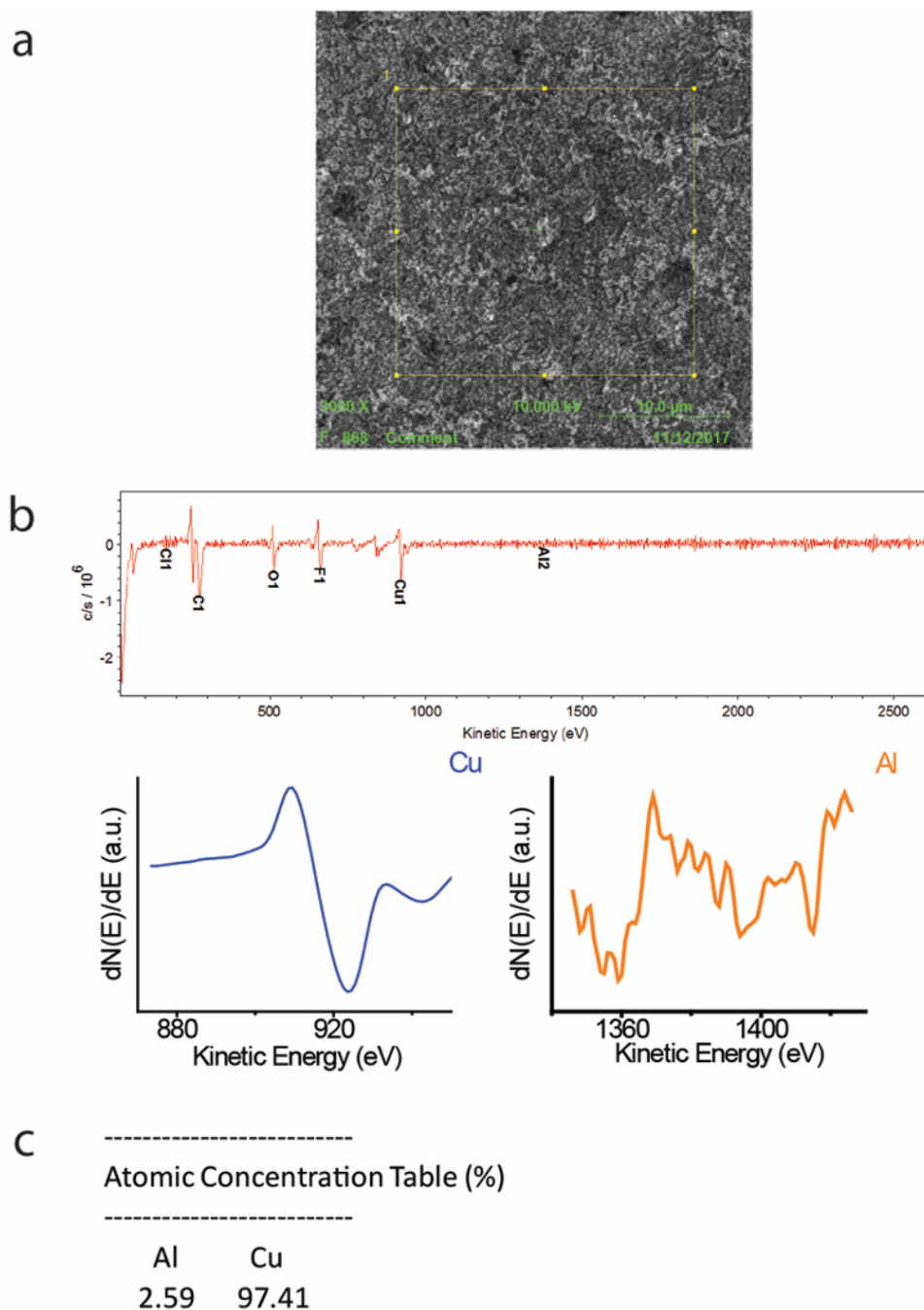


Figure B-36: Auger electron spectroscopic analysis of a de-alloyed Cu-Al catalyst soaked in 11 mM CuCl_2 solution for 0.5 hour. a, Auger secondary electron microscopic image. b, Auger spectroscopic survey and narrow-scan spectra. c, Concentrations of Cu and Al. The Al concentration calculated from Al narrow-scan spectrum is already close to the AES detecting limit.

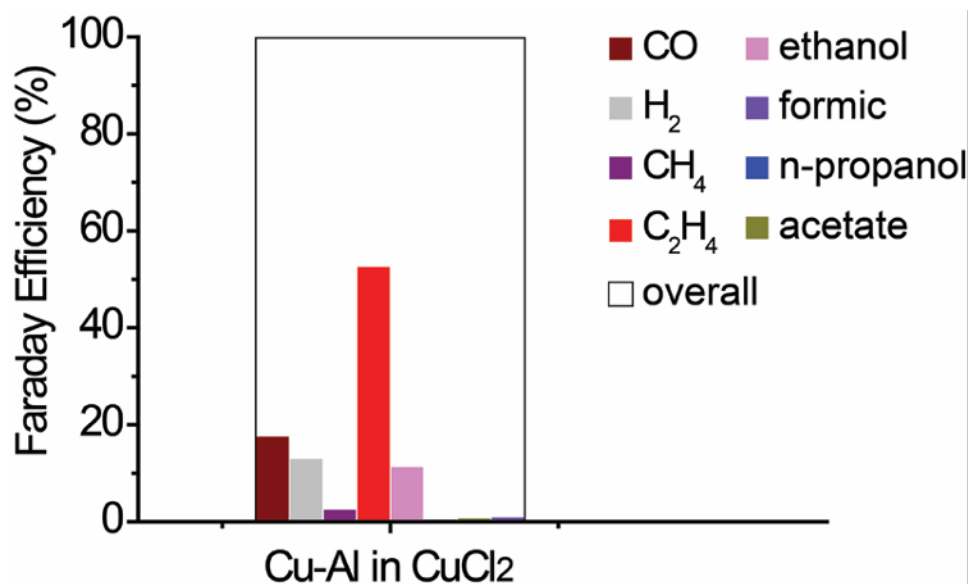


Figure B-37: CO₂ electroreduction performance. Faradaic efficiencies for all products at an applied current density of 600 mA cm² obtained with of a de-alloyed Cu-Al catalyst after soaking in 10 mM CuCl₂ solution for 10 minutes.

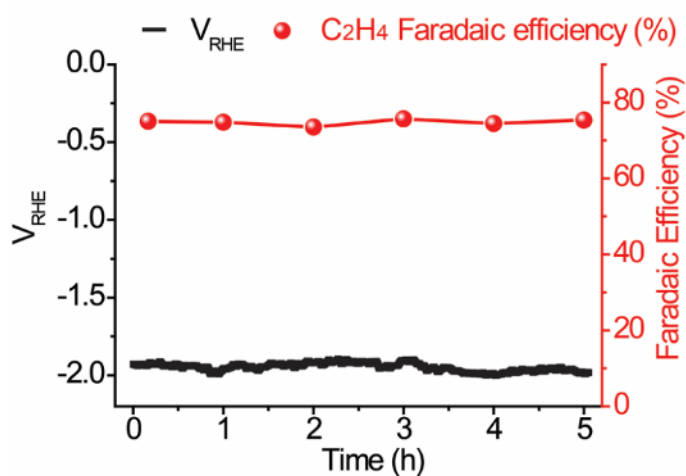


Figure B-38: CO₂ electroreduction stability of the de-alloyed Cu-Al/C-GDL electrode. The CO₂ electroreduction activity of a de-alloyed Cu-Al/C-GDL electrode at an applied current density of 600 mA cm². Left axis: potential (V vs. RHE) vs. time (s), right axis: C₂H₄ Faradaic efficiency (%) vs. time (s)

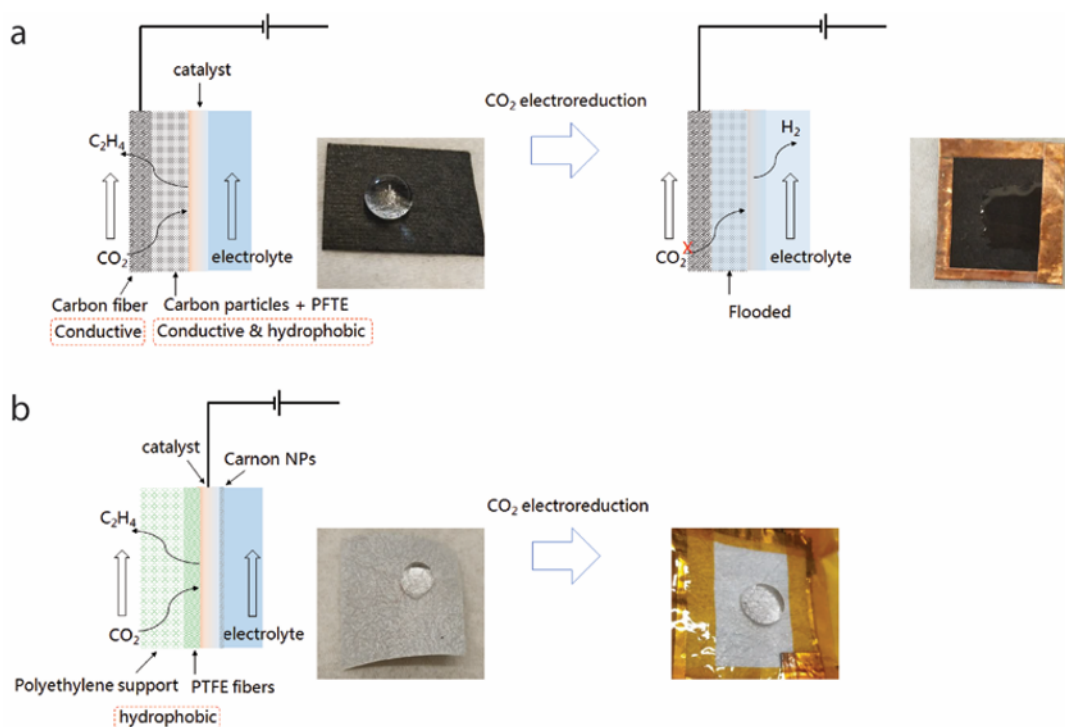


Figure B-39: Configuration of de-alloyed Cu-Al catalysts on (a) C-GDL and (b) PTFE substrates for CO₂ electrolysis in a flow cell system. Optical images were taken on the backside of the samples to show the hydrophobicity of a, Cu-Al/C-GDL before and after 5 hours stability test in 1 M KOH, and b, Cu-Al/PTFE before and after 50 hours stability test in 3 M KOH with 3M KI) solutions.

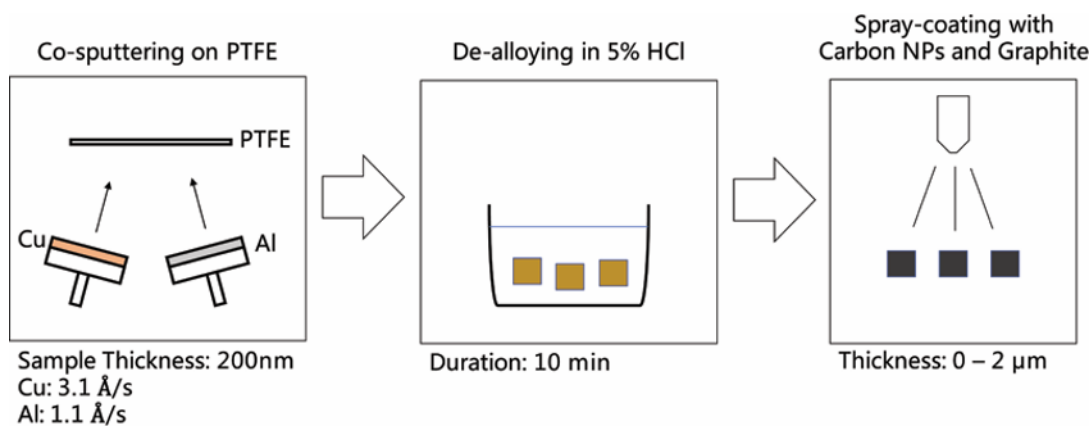


Figure B-40: Co-sputtering, chemical etching and spray coating to make graphit/carbon NPs/de-alloyed Cu-Al /PTFE substrates.

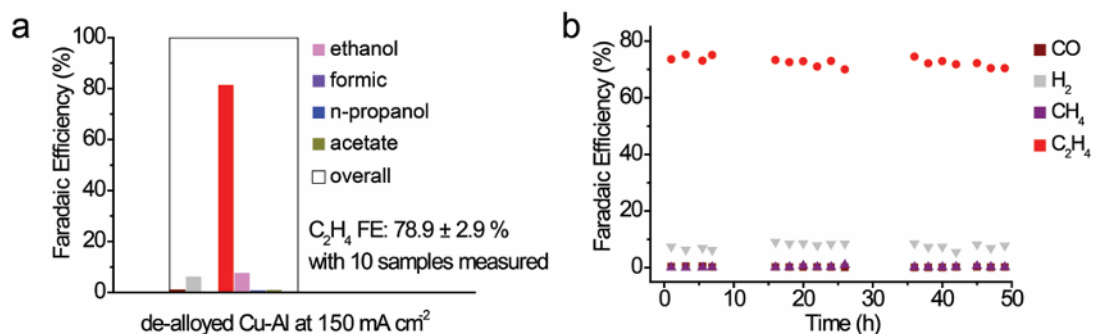


Figure B-41: CO₂ electroreduction performance with graphite/carbon NPs/de-alloyed Cu-Al/PTFE samples in a 1 M KOH electrolyte. a, Faradaic efficiencies for all products at an applied current density of 400 mA cm² with 10 distinct samples measured. b, Faradaic efficiency of gas products vs. time during the stability test.

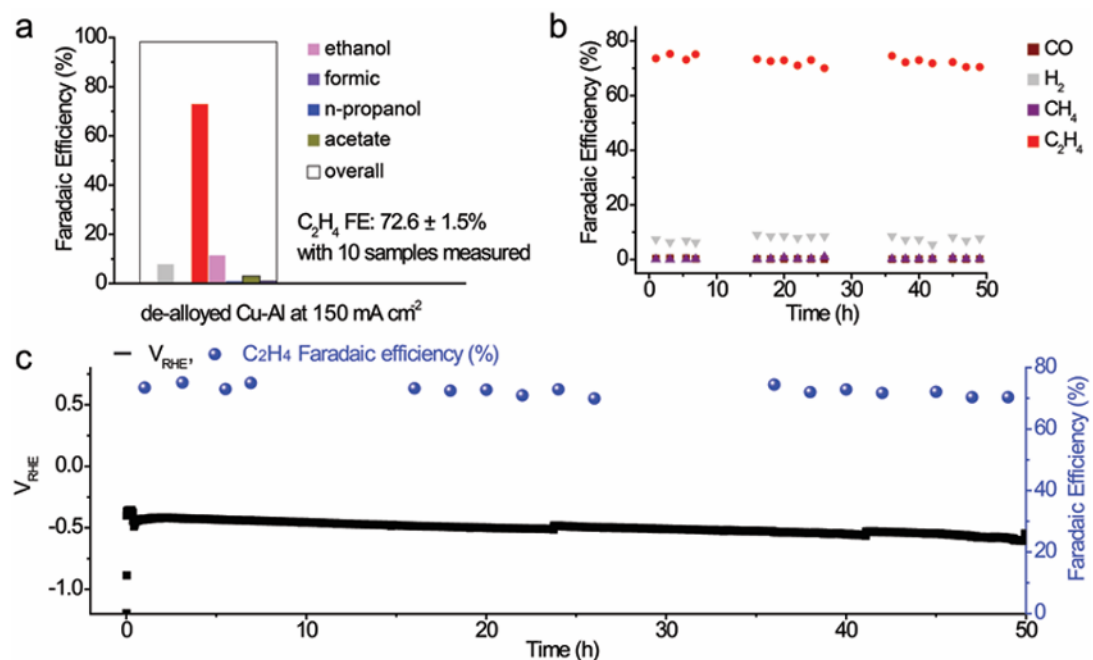


Figure B-42: CO₂ electroreduction performance with graphite/carbon NPs/de-alloyed Cu-Al/PTFE samples in a 3 M KOH and 3 M KI electrolyte. a, Faradaic efficiencies for all products at an applied current density of 150 mA cm² with 10 distinct samples measured. b, Faradaic efficiency of gas products vs. time during the stability test. c, The CO₂ electroreduction stability at an applied current density of 150 mA cm². Left axis: potential (V vs. RHE) vs. time (s), right axis: C₂H₄ Faradaic efficiency (%) vs. time (s)

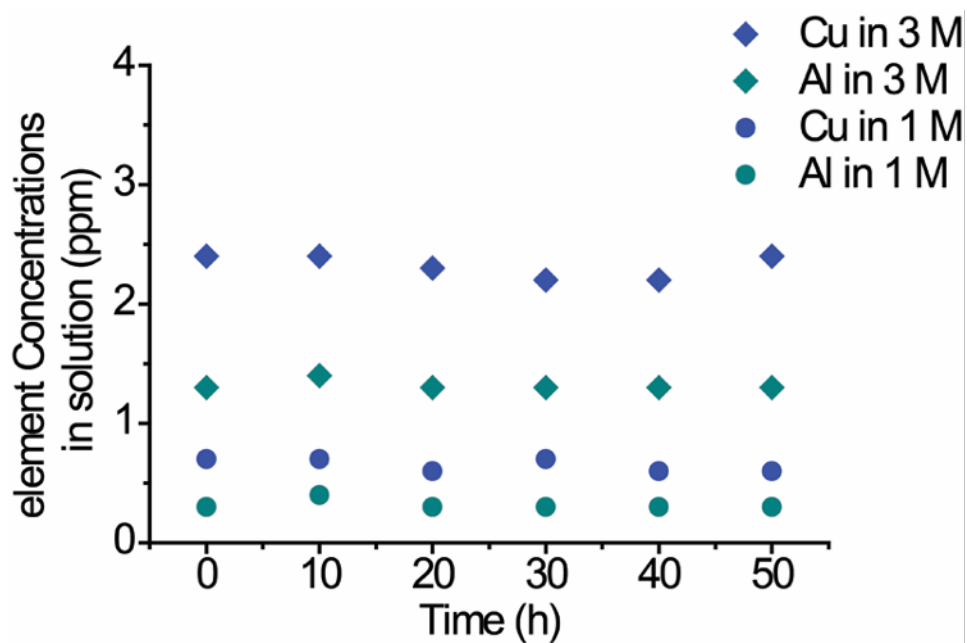


Figure B-43: ICP-AES results of Cu and Al ions in the testing solution with different the testing time. The de-alloyed Cu-Al catalysts on PTFE was measured at 400 mA cm^2 in 1 M KOH solution and at 150 mA cm^2 in 3 M KOH solution for 50 hours.

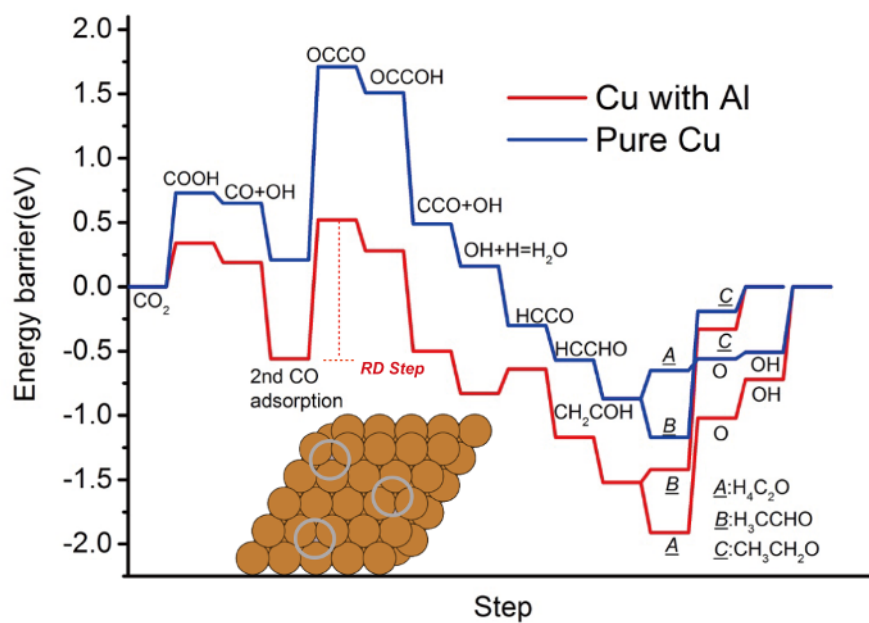


Figure B-44: Reaction Gibbs free energy diagram. Reaction Gibbs free energy diagram from adsorbed CO_2 to C_2H_4 and $\text{C}_2\text{H}_5\text{OH}$ on Cu (111) facet with (red lines) and without (blue lines) Al incorporation in the subsurface (the grey circle shows the Al atoms' positions). The model of Al incorporated Cu is described in lower left. The energy barrier of the rate determine step (OC-CO dimerization) decreases after Al incorporation.

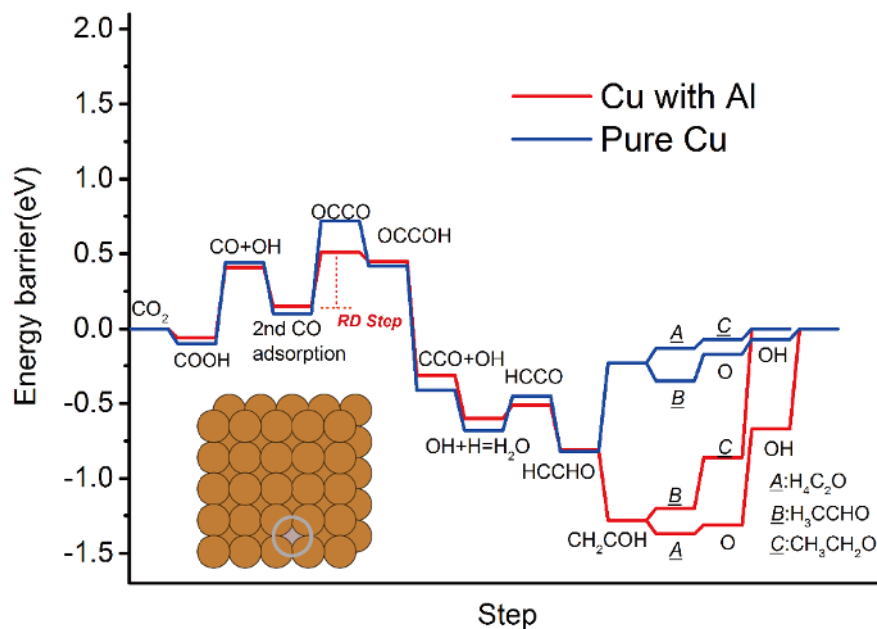


Figure B-45: Reaction Gibbs free energy diagram. Reaction Gibbs free energy diagram from adsorbed CO_2 to C_2H_4 and $\text{C}_2\text{H}_5\text{OH}$ on Cu (100) facet with (red lines) and without (blue lines) Al incorporation in the subsurface. The model of Al incorporated Cu is described in lower left. The energy barrier of the rate determine step (OC-CO dimerization) decreases after Al incorporation.

B.2.9 In situ and ex situ X-ray absorption studies

A large amount of Cu-O bonding was visible continuously in the ex situ and in situ XANES analyses with the de-alloyed Cu-Al catalyst before, during and after the reaction (Figure B-47). In situ Fourier transform extended X-ray absorption fine structure (FT EXAFS) spectra of the de-alloyed Cu-Al catalysts, together with ex-situ reference spectra of Cu foil, Cu_2O and CuO, gave precise coordination information. We observed a prominent peak at $2.2 \text{ \AA}$ from Cu-Cu contribution, which agrees with the prominent peak of Cu-Cu in Cu foils, and a weak peak at $1.5 \text{ \AA}$, which agrees with the prominent peak of Cu-O in Cu_2O and CuO. Moreover, the Cu-O peak intensity remained almost the same during in situ synchrotron measurement. It is smaller than that obtained with the same sample without CO_2 electroreduction (Figure B-

47), indicating Cu-O in de-alloyed Cu-Al catalysts was partially reduced to Cu. A stable balance of Cu-Cu and Cu-O in the de-alloyed catalyst was established within a short time right after starting CO₂ electroreduction. As a control, we measured in situ XANES with a pure Cu on GDL under the same condition. No obvious oxidation state of Cu was observed (Figure B-47).

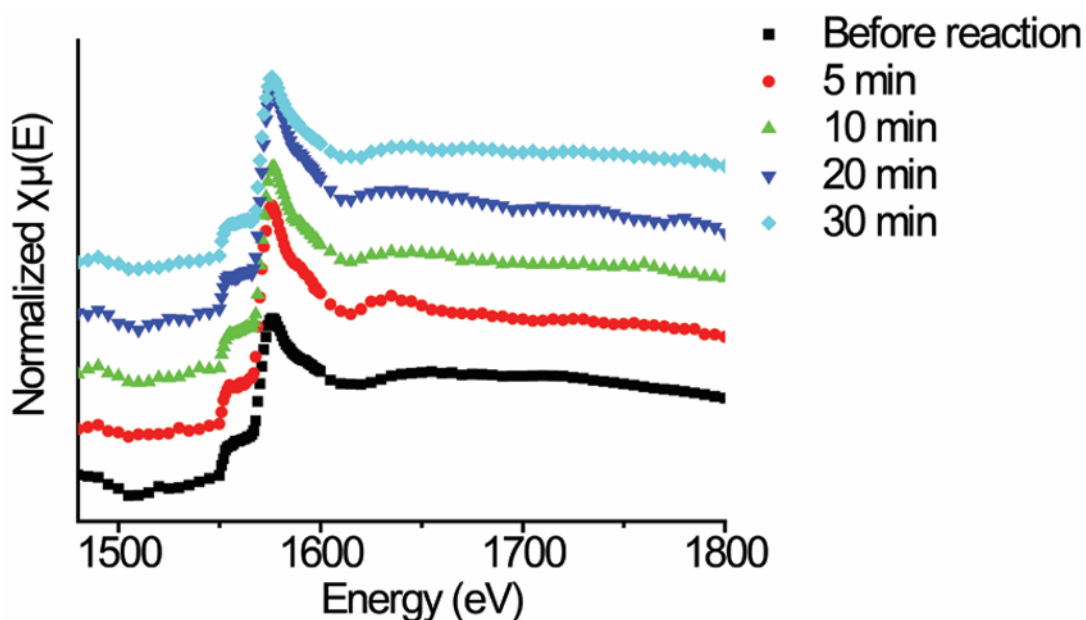


Figure B-46: Synchrotron analyses of the de-alloyed Cu-Al catalyst before and after 5, 10, 20 and 30 min. of the CO₂ electroreduction reactions.

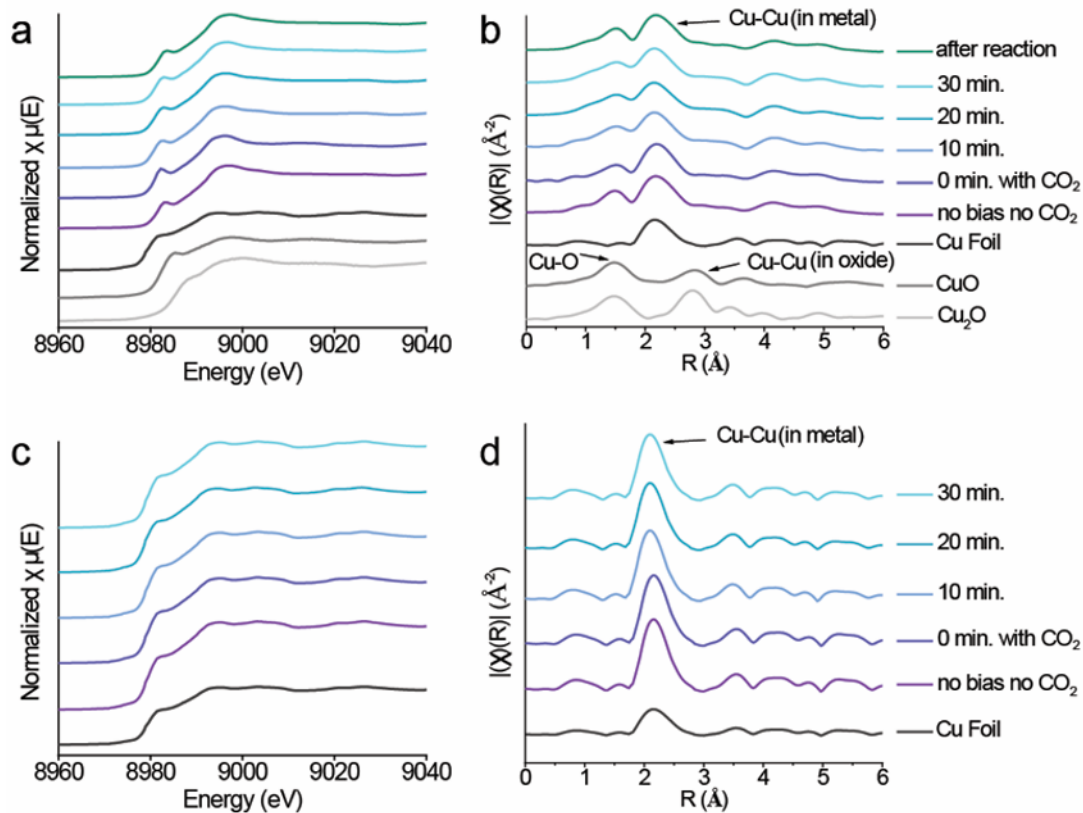


Figure B-47: Ex situ and in situ synchrotron analyses of the high-performing de-alloyed Cu-Al catalyst and pure Cu catalyst. a-b, In situ X-ray absorption near-edge structure (XANES) spectra of de-alloyed Cu-Al catalyst before, during and after CO_2 electroreduction tests and ex situ XANES spectra of Cu foil, Cu_2O , CuO. c-d, In situ X-ray absorption near-edge structure (XANES) spectra of pure Cu catalyst before and during CO_2 electroreduction tests and ex situ XANES spectra of Cu foil.

Table B.1: The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-48.

No.	1	2	3	4	5	Pure Cu
RD step barrier (eV)	0.51	0.52	0.51	0.56	0.53	0.62

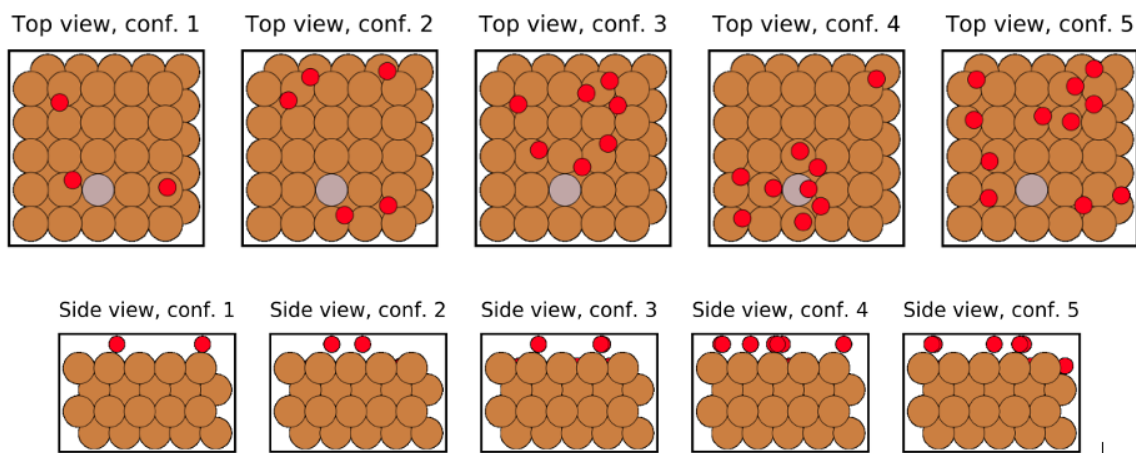


Figure B-48: Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO_2 reduction. 4% Al incorporated Cu (100) surface with different amounts of Cu-O bonds as shown in configurations 1–5.

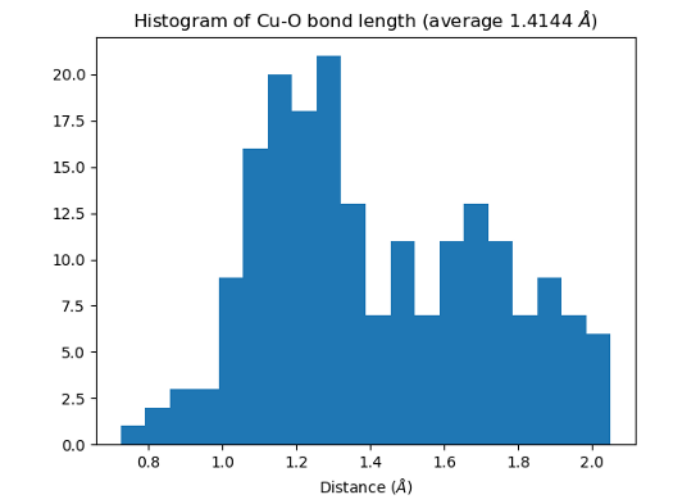


Figure B-49: The distribution of Cu-O bond length in the Figure B-48.

Table B.2: The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-50.

No.	1	2	3	4	5	Pure Cu
RD step barrier (eV)	0.45	0.47	0.45	0.48	0.49	0.62

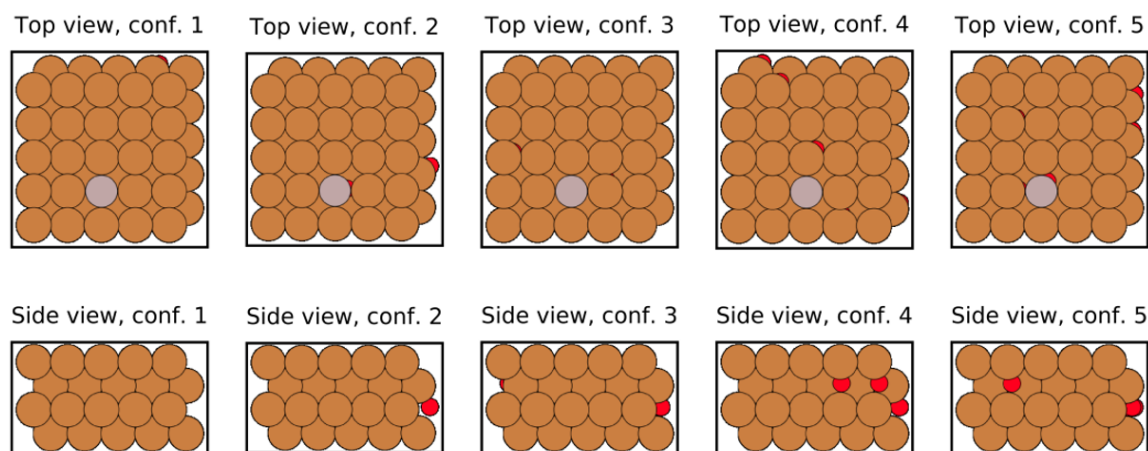


Figure B-50: Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO_2 reduction. 4% Al incorporated Cu (100) surface with different amounts of Cu-O bonds as shown in configurations 1–5.

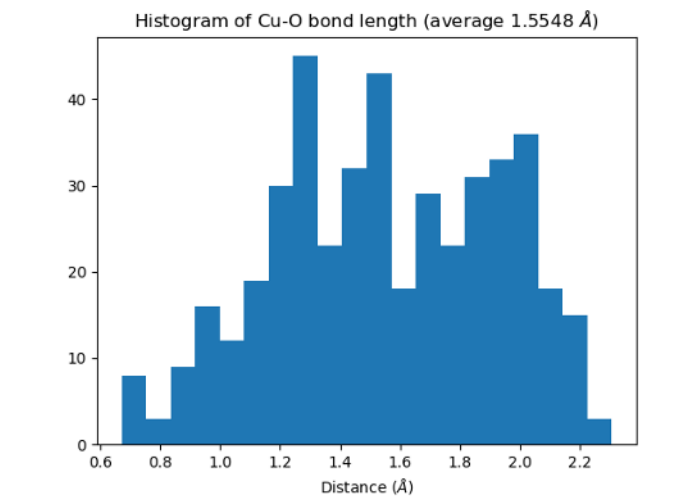


Figure B-51: The distribution of Cu-O bond length in the Figure B-50.

Table B.3: The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-52.

No.	1	2	3	4	5	Pure Cu
RD step barrier (eV)	0.53	0.57	0.59	0.51	0.60	0.62

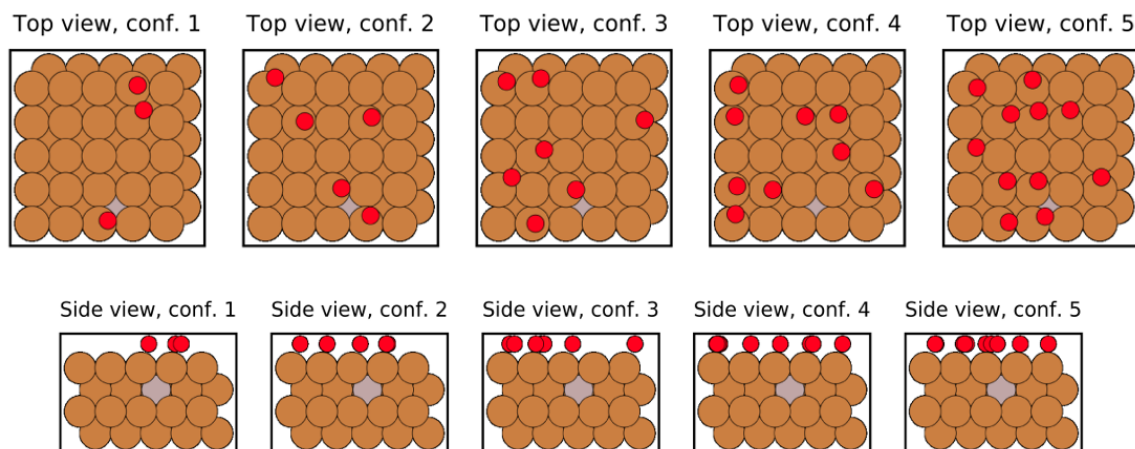


Figure B-52: Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO_2 reduction. 4% Al incorporated Cu (100) surface with different amounts of Cu-O bonds as shown in configurations 1–5.

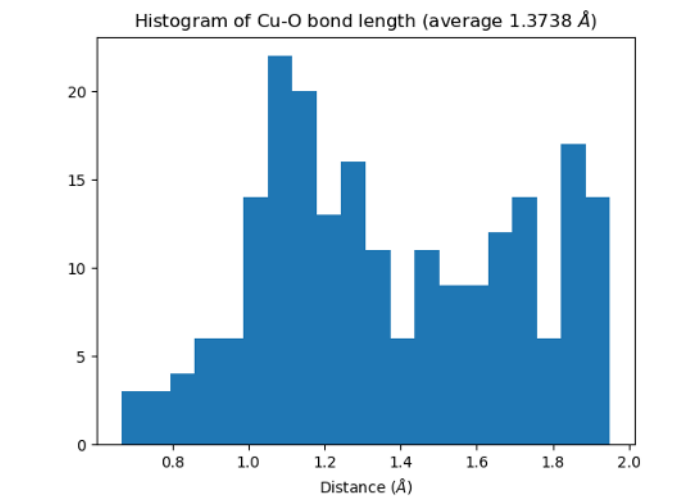


Figure B-53: The distribution of Cu-O bond length in the Figure B-52.

Table B.4: The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-54.

No.	1	2	3	4	5	Pure Cu
RD step barrier (eV)	0.46	0.42	0.56	0.52	0.51	0.62

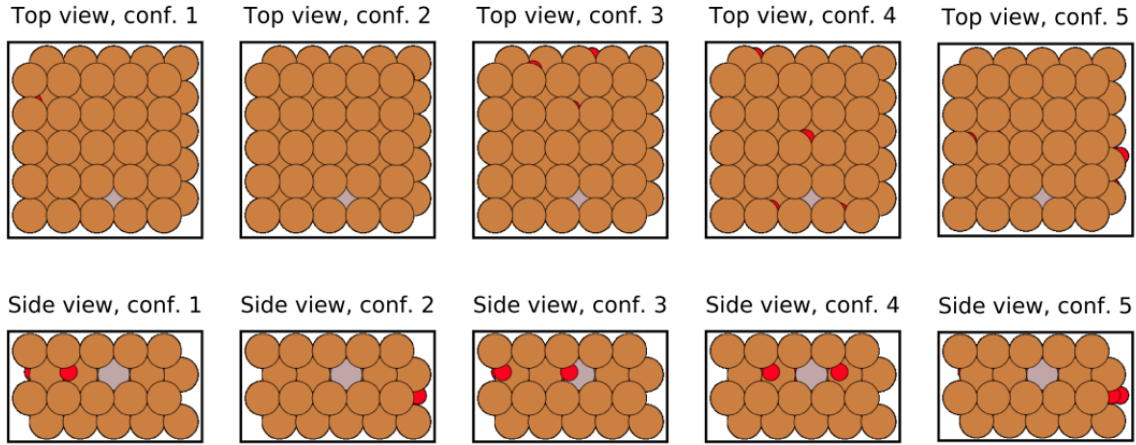


Figure B-54: Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO_2 reduction. 4% Al incorporated Cu (100) surface with different amounts of Cu-O bonds as shown in configurations 1–5.

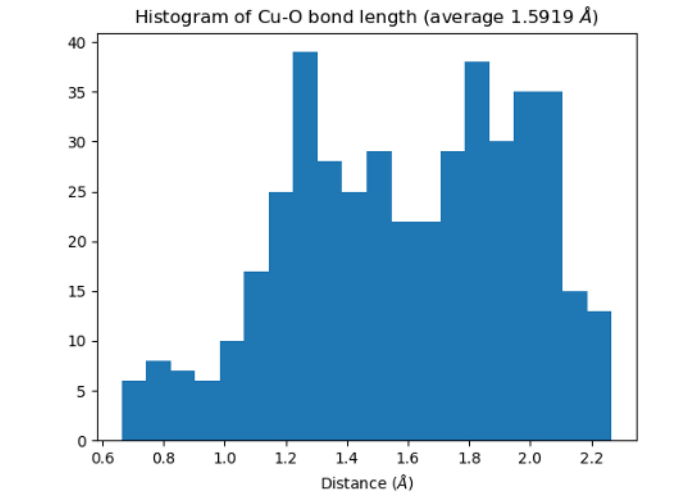


Figure B-55: The distribution of Cu-O bond length in the Figure B-54.

Table B.5: The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-56.

No.	1	2	3	4	5	Pure Cu
RD step barrier (eV)	1.103	1.02	0.98	0.95	1.07	1.47

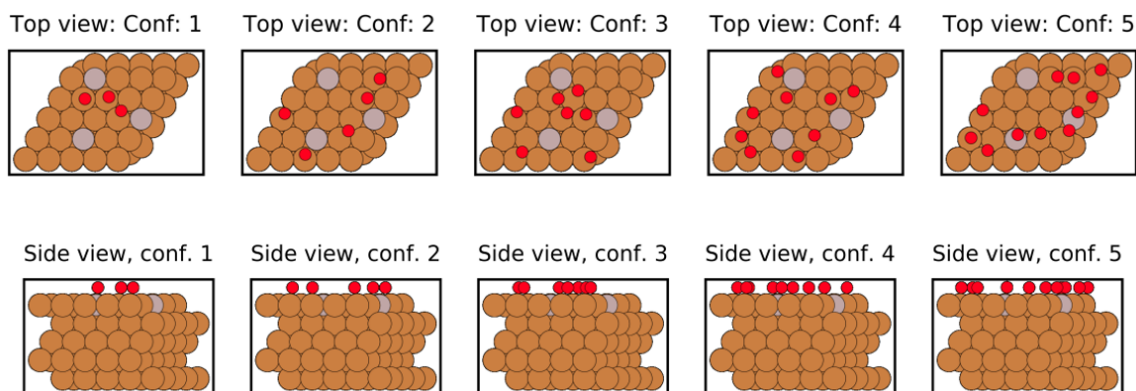


Figure B-56: Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO₂ reduction. 12% Al incorporated Cu (111) surface with different amounts of Cu-O bonds as shown in configurations 1–5.

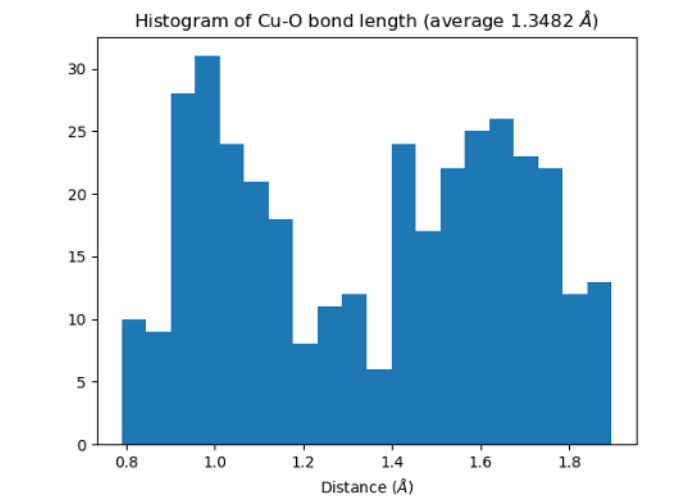


Figure B-57: The distribution of Cu-O bond length in the Figure B-56.

Table B.6: The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-58.

No.	1	2	3	4	5	Pure Cu
RD step barrier (eV)	0.98	0.99	1.17	1.21	1.03	1.47

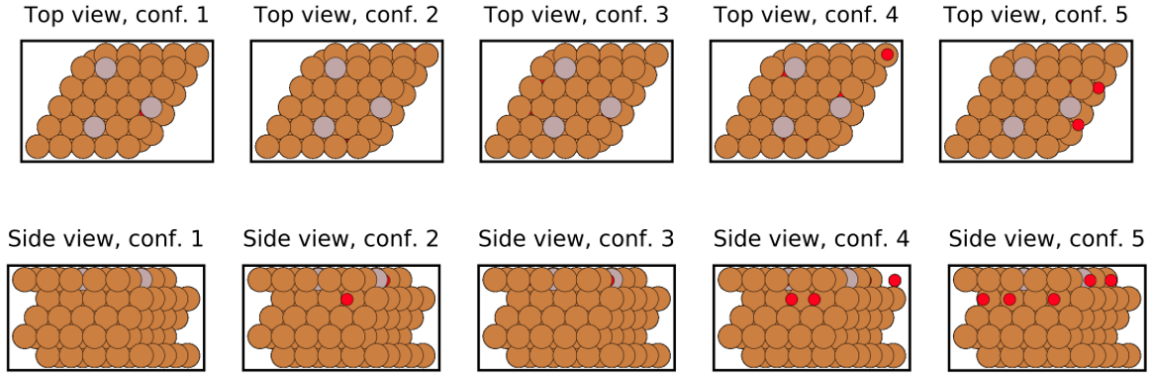


Figure B-58: Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO_2 reduction. 12% Al incorporated Cu (111) surface with different amounts of Cu-O bonds as shown in configurations 1–5.

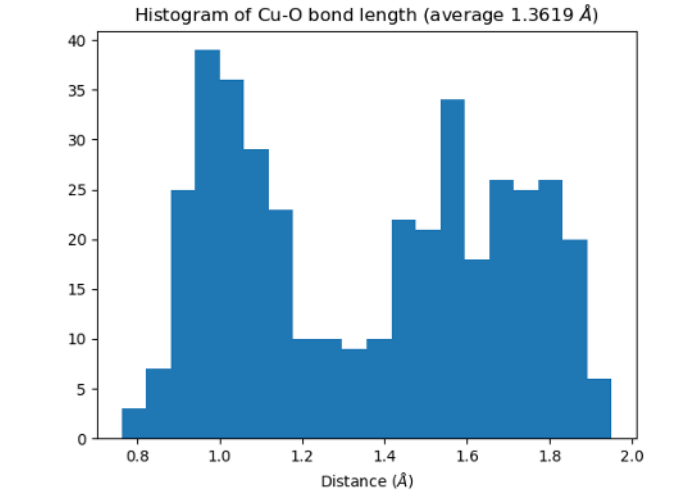


Figure B-59: The distribution of Cu-O bond length in the Figure B-58.

Table B.7: The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-60.

No.	1	2	3	4	5	Pure Cu
RD step barrier (eV)	0.92	1.03	1.03	1.10	1.13	1.47

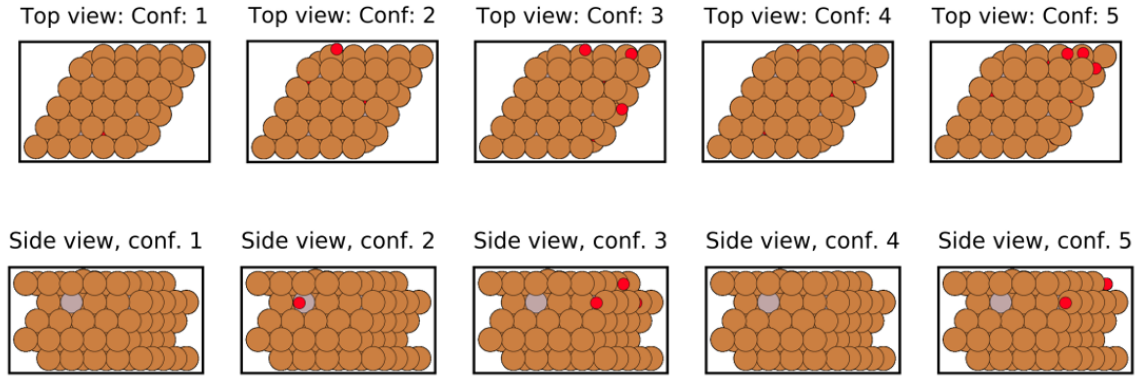


Figure B-60: Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO₂ reduction. 12% Al incorporated Cu (111) surface with different amounts of Cu-O bonds as shown in configurations 1–5.

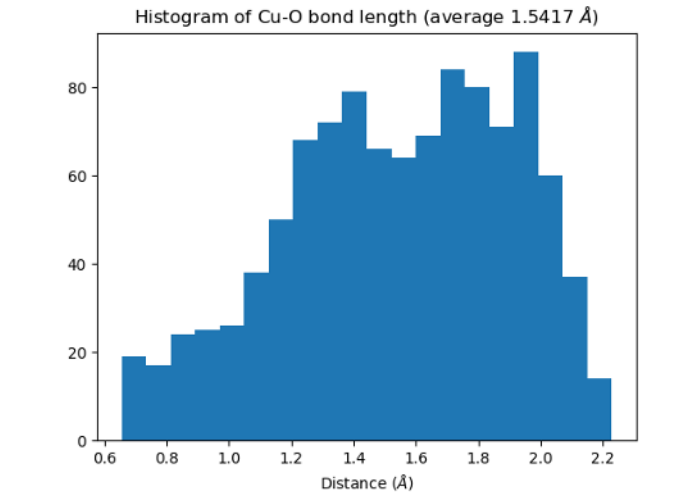


Figure B-61: The distribution of Cu-O bond length in the Figure B-60.

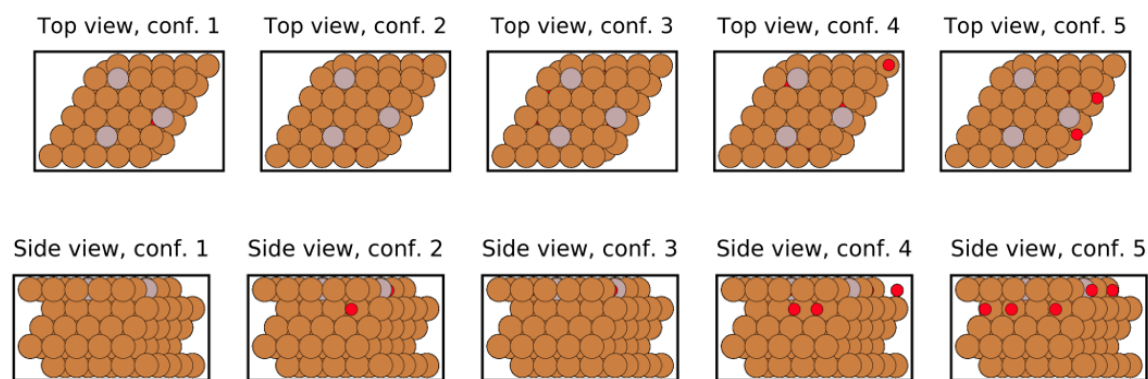


Figure B-62: Models used to calculate the reaction energies in the rate-determining step of C-C bond making in the CO_2 reduction. 12% Al incorporated Cu (111) surface with different amounts of Cu-O bonds as shown in configurations 1–5.

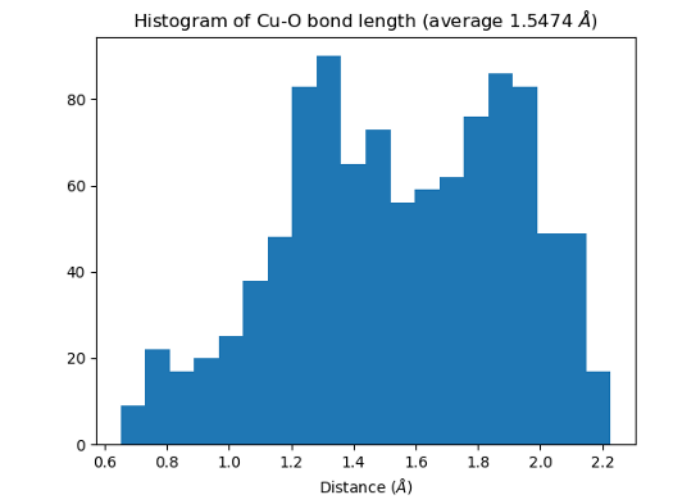


Figure B-63: The distribution of Cu-O bond length in the Figure B-62.

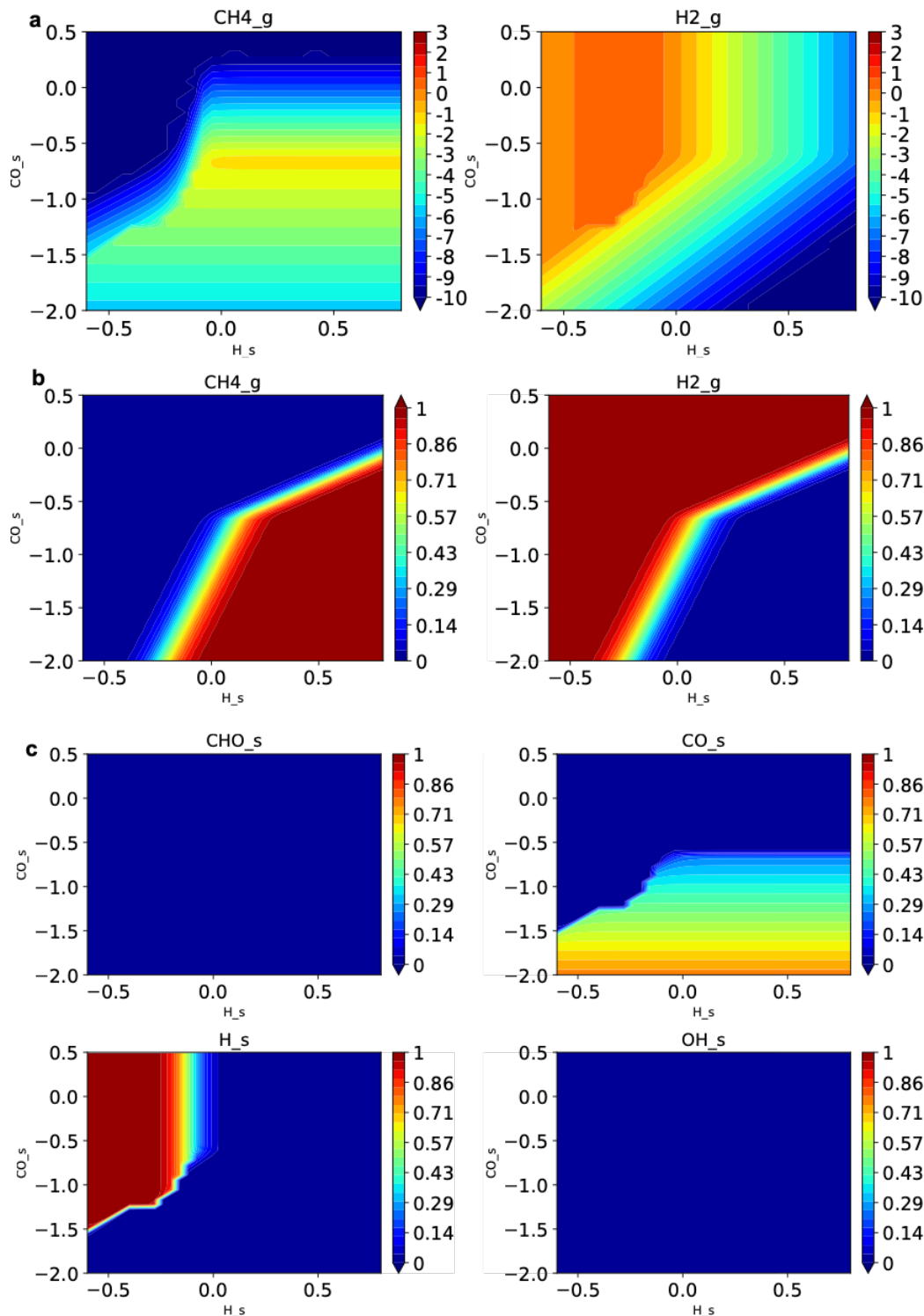


Figure B-64: Activity, selectivity, and coverage maps at -0.5 V applied potential, based on (211) metal scaling relations and microkinetic models. a, Activity of CH_4 and H_2 production (in units of $\log(\text{TOF [1/s]})$) as a function of ΔE_{CO} vs. ΔE_{H} . b, Selectivity towards CH_4 and H_2 production as a function of ΔE_{CO} vs. ΔE_{H} . c, Surface coverages of CHO , CO , H , and OH adsorbates as a function of ΔE_{CO} vs. ΔE_{H} .

Table B.8: The reaction energies in the rate-determining step of C-C bond making on different configurations in the Figure B-62.

No.	1	2	3	4	5	Pure Cu
RD step barrier (eV)	0.87	0.92	0.90	0.94	1.19	1.47

Table B.9: Zero-point energy and specific heat of different reactants, intermediates and products on the Al incorporated Cu (111) surface. It corresponds to the data used in Figure B-32.

Structure	E_{ZPE} (eV)	$C_{v,trans}(0 \rightarrow T)$ (eV)	$C_{v,rot}(0 \rightarrow T)$ (eV)	$C_{v,vib}(0 \rightarrow T)$ (eV)	H (eV)
Initial	2.17	0.0390	0.0390	0.330	2.61
Final _a	2.62	0.0390	0.0390	0.430	3.16
Final _b	2.15	0.0390	0.0390	0.377	2.63

Table B.10: Entropy and Gibbs free energy of different reactants, intermediates and products on the Al incorporated Cu (111) surface. It corresponds to the data used in Figure B-32.

Structure	S_{trans} (eV/K)	S_{rot} (eV/K)	S_{vib} (eV/K)	S (eV/K)	TS (eV)
Initial	0.00219	0.00204	0.00208	0.00632	1.88
Final _a	0.00219	0.00200	0.00278	0.00697	2.08
Final _b	0.00219	0.00203	0.00668	0.00668	1.99

Table B.11: Zero-point energy and specific heat of different reactants, intermediates and products on the Al incorporated Cu (100) surface. It corresponds to the data used in Figure B-32.

Structure	E_{ZPE} (eV)	$C_{v,trans}(0 \rightarrow T)$ (eV)	$C_{v,rot}(0 \rightarrow T)$ (eV)	$C_{v,vib}(0 \rightarrow T)$ (eV)	H (eV)
Initial	2.15	0.0390	0.0390	0.188	2.44
Final _a	2.43	0.0390	0.0390	0.182	2.71
Final _b	1.97	0.0390	0.0390	0.176	2.25

Appendix C

Supplementary Information for Chapter 5

This work originally appeared as the Supplementary Information for: Kevin Tran, Willie Neiswanger,* Junwoong Yoon, Qingyang Zhang, Eric Xing, and Zachary W Ulissi. Methods for comparing uncertainty quantifications for material property predictions. Machine Learning: Science and Technology, 1(025006), 2020.*

Figure C-1 shows how the residuals of each model are correlated with each other. All pairs of models show a positive correlation between each other. This suggests that poor predictions made by one model were also made by most other models, qualitatively speaking. This observation is consistent with the finding that the accuracy of all models in this study were comparable. No single model was substantially better at predicting the outlying points than any other model.

Figure C-2 shows how the predicted uncertainties of each model are correlated with each other. The only pattern we could discern was the correlation between the GP and GP_{NN- μ} methods. This correlation likely due to the fact that both methods used the same exact feature space for their GPs. The only difference between the two were their mean functions.

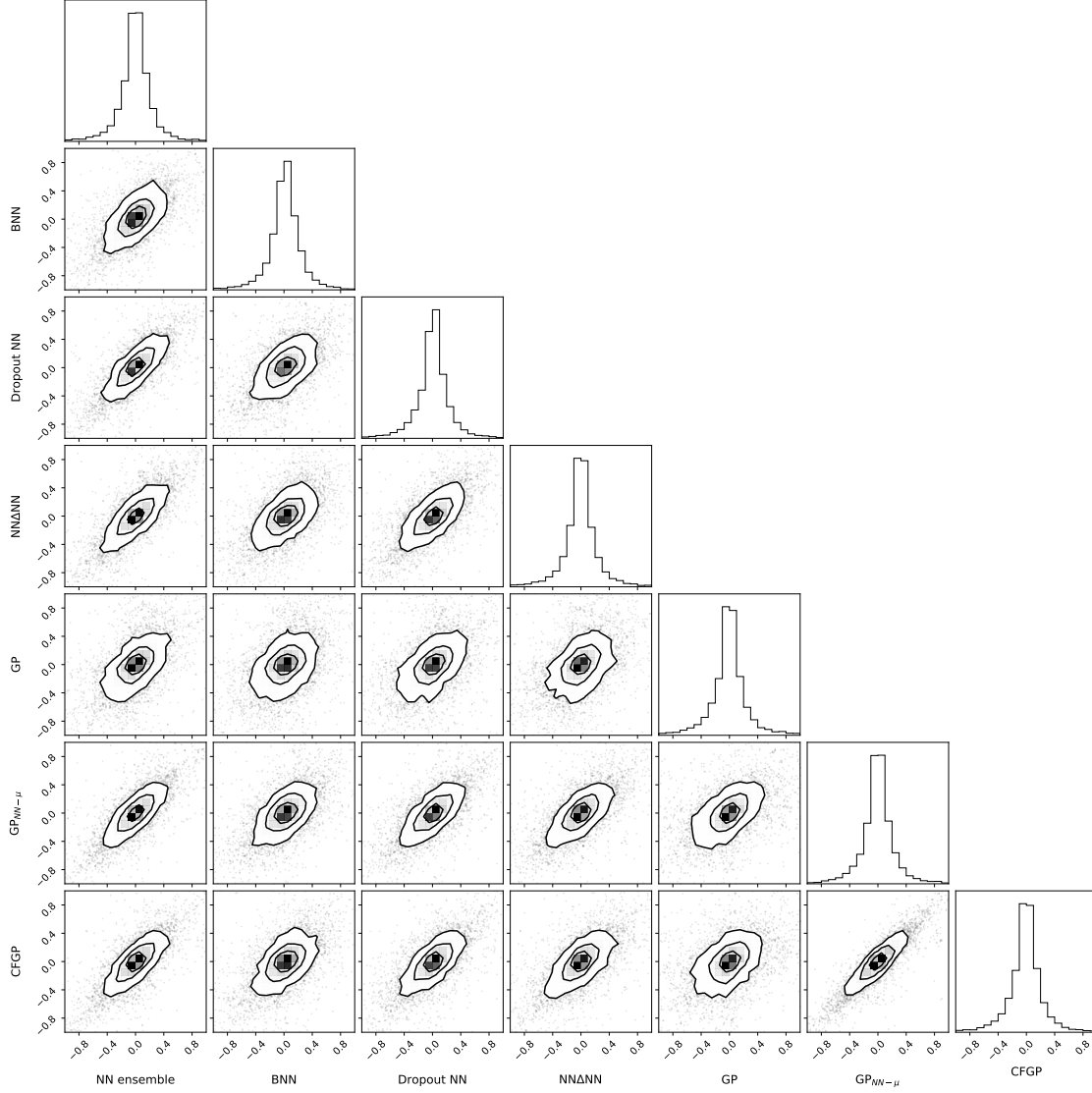


Figure C-1: Corner plot of the residuals of all models. Each subfigure shows the parity between the residuals of pairs of models. Solid contour lines delineate quartiles of the point distribution. Single, faded points indicate parity points in the fourth, least dense quartile of points. Shaded pixels indicate the highest density of points with darker shading indicating a higher density. The figures along the diagonal show histogram distributions of the residuals for each model. All units are in eV.

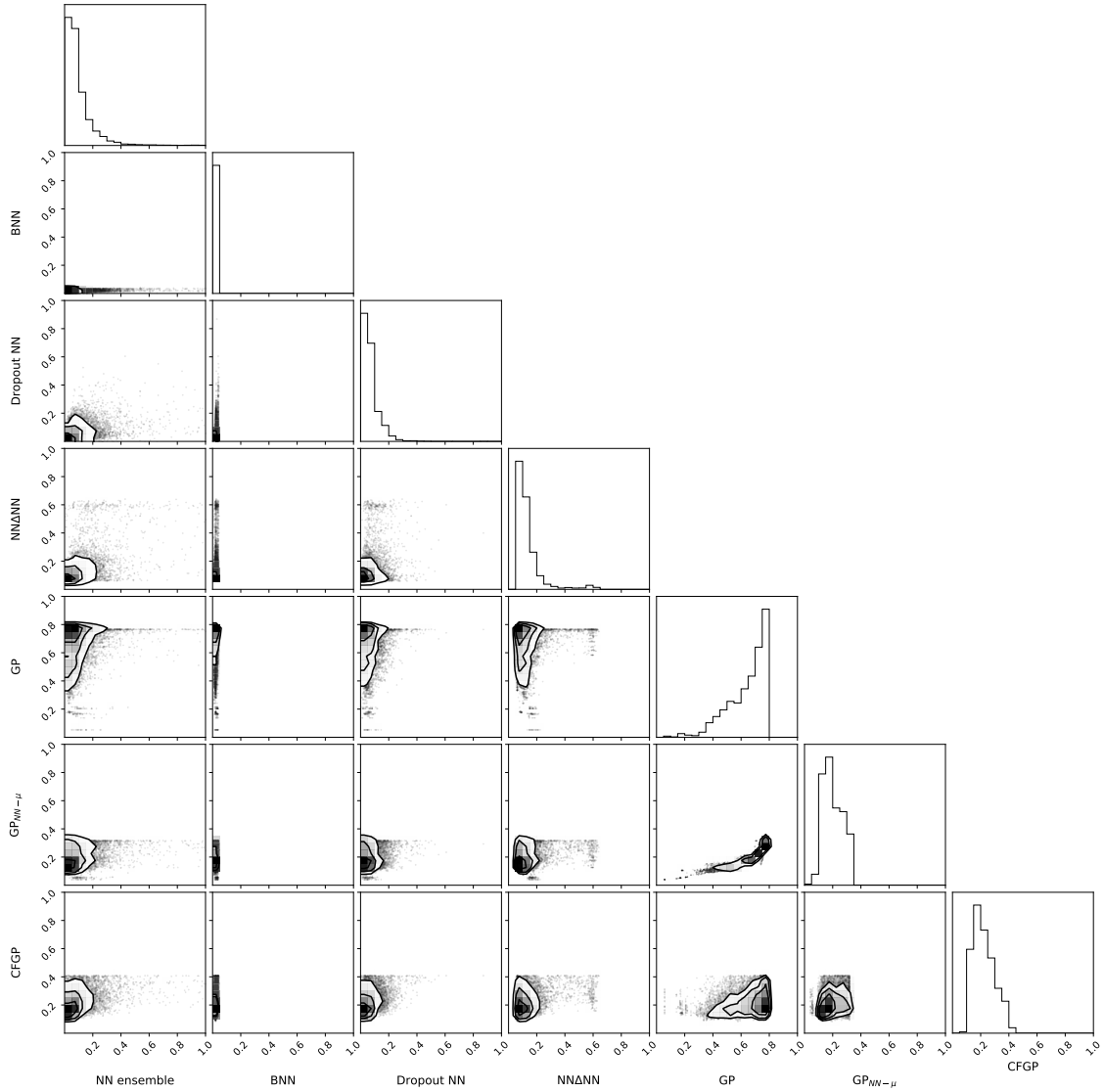


Figure C-2: Corner plot of the residuals of all models. Each subfigure shows the parity between the estimated standard deviations of pairs of models. Solid contour lines delineate quartiles of the point distribution. Single, faded points indicate parity points in the fourth, least dense quartile of points. Shaded pixels indicate the highest density of points with darker shading indicating a higher density. The figures along the diagonal show histogram distributions of the predicted standard deviations for each model. All units are in eV.