

pubs.acs.org/JACS

Article

Machine-Learning-Guided Discovery and Optimization of Additives in Preparing Cu Catalysts for CO₂ Reduction

Ying Guo,[#] Xinru He,[#] Yuming Su, Yiheng Dai, Mingcan Xie, Shuangli Yang, Jiawei Chen, Kun Wang, Da Zhou,* and Cheng Wang*



ABSTRACT: Discovery and optimization of new catalysts can be potentially accelerated by efficient data analysis using machinelearning (ML). In this paper, we record the process of searching for additives in the electrochemical deposition of Cu catalysts for CO₂ reduction (CO₂RR) using ML, which includes three iterative cycles: "experimental test; ML analysis; prediction and redesign". Cu catalysts are known for CO₂RR to obtain a range of products including C₁ (CO, HCOOH, CH₄, CH₃OH) and C₂₊ (C₂H₄, C₂H₆,



 C_2H_5OH , C_3H_7OH). Subtle changes in morphology and surface structure of the catalysts caused by additives in catalyst preparation can lead to dramatic shifts in CO_2RR selectivity. After several ML cycles, we obtained catalysts selective for CO, HCOOH, and C_{2+} products. This catalyst discovery process highlights the potential of ML to accelerate material development by efficiently extracting information from a limited number of experimental data.

INTRODUCTION

Catalysts are essential for several chemical transformations. However, finding a new catalyst for a target reaction is still challenging due to the complicated catalytical process.¹⁻⁵ Machine-learning (ML) has the potential to accelerate catalyst development by efficient data analysis. Recently, ML has been coupled with quantum chemistry calculations for in silico material screening;⁶⁻¹² however, examples using ML analysis of experimental data to find new catalysts are still limited.^{1,13-16} Here we report the process of using ML to discover and optimize additives in preparing Cu catalysts for electrochemical CO₂ reduction (CO₂RR). Cu catalysts are known for CO₂RR to obtain a range of products including carbon monoxide (CO), formic acid (HCOOH), methane (CH_4) , methanol (CH_3OH) , ethylene (C_2H_4) , ethane (C_2H_6) , ethanol (C₂H₅OH), propanol (C₃H₇OH), etc.¹⁷⁻²² Subtle changes in morphology and surface structure of the catalysts lead to dramatic shifts in CO_2RR selectivity,^{22–27} which can be fine-tuned with additives in catalyst preparation.²⁸⁻³⁰ A clear structure-property relationship is still lacking.

We chose electrochemical deposition as the method to prepare Cu catalysts from Cu salts and added different metal salts and organic molecules as additives. After three iterative cycles of "experimental test–ML analysis–prediction and redesign", this discovery pipeline identified Sn salt as an important additive to obtain CO and HCOOH, and aliphatic alcohols as important additives to promote C_{2+} production.

Further characterization of the catalysts prepared with different additives showed that the aliphatic alcohols possibly promote formation of cubes of Cu_2O in the electrodeposition

step. The Cu₂O cubes were then reduced to metallic Cu under experimental conditions with high selectivity toward C₂₊ products, which is consistent with reported performance of oxide-derived copper (OD-Cu).³¹⁻³⁵

RESULTS AND DISCUSSION

Initial Data Collection. In the beginning, we prepared an additive library including 12 metal salts and 200 water-soluble organic molecules (Tables S1 and S2). The combinations of one metal salt plus one organic additive can give more than 2000 potential recipes (Figure 1), but we used a uniform design algorithm³⁶ to choose only 112 combinations from 12 metal additives and 50 organic additives to perform the experiments in the first round of learning. The uniform design can ensure a most efficient search of a high-dimensional variable space. We present more details about the algorithm in the Supporting Information.

The CO₂RR performance was evaluated by measuring faradaic efficiency and current density (in units of mA cm⁻²). Carbon cloth was used as the support. The potential for electrochemical deposition was set to -1.0 V versus Ag/AgCl, while the CO₂RR electrolysis potential was set to -1.6 V versus Ag/AgCl. The electrodeposition was performed in 0.1

Received: January 11, 2021 Published: April 12, 2021





Figure 1. (a) Preparation of Cu electrocatalysts via electrochemical deposition followed by evaluation of catalytic performance. (b) The learning loop for accelerated discovery based on three iterative cycles of "experimental test–ML analysis–prediction and redesign" to guide the search for high-performance CO_2RR catalysts.

M KCl aqueous solution, after which the working electrode was washed by water and put in a $KHCO_3 (0.1 \text{ M})/CO_2 (1 \text{ bar, sat.})$ buffer in an H-type electrolyzer for CO_2RR test.

At the beginning of the first round of the search, we tested several sets of experimental operational conditions together with the catalyst preparation condition, including the metal salt concentration, ligand concentration, electrolyte concentration, electrodeposition potential, and electrolysis potential. However, in a later search, to save experimental resources, we limit our search under a reasonable fixed operational condition.

First Round of Learning. In the first round of learning, the additives were used as inputs and faradaic efficiencies of H₂ (FE-H₂), CO (FE-CO), HCOOH (FE-HCOOH), the sum of C₂H₄, C₂H₅OH, and C₃H₇OH (FE-C₂₊), and current densities of H₂ (*j*_{H2}), CO (*j*_{CO}), HCOOH (*j*_{HCOOH}), the sum of C₂H₄, C₂H₅OH, and C₃H₇OH (*j*_{C2+}) were the outputs. To analyze this data set, we need to convert chemical structure to variables that can be used in ML, a process known as feature engineering in the ML community.

Here we used functional groups on the molecule as the features, representing each additive by a one-hot vector that gives information about the presence/absence of a certain metal and whether the molecular additive contains a certain functional group or not (Table S4).

We first assigned the FEs to be in the high or low region by comparing to a predefined threshold value (Table S4) and used five classification algorithms, "decision tree" (DT), "random forest" (RF), gradient boosting classification tree (GBCT), logistic regression (LR), and XGboost, to analyze the data (Figures S3–S6, Table S5).^{37–41} Different models have different strengths, which are listed in the Supporting Information. Although the regression model seems to be more attractive for the task, considering significant errors in the experimental measurements, the classification model can give a much higher signal-to-noise ratio than the regression model, as it effectively integrates samples within a range of different FEs. The accuracy for FE-CO prediction was up to 96% (XGboost) (Table S5), and the accuracy for predicting the FE-C₂₊ was up to 76% (LR) (Table S5).

pubs.acs.org/JACS

For FE-CO, feature importance analysis from RF and GBCT showed the metal additive Sn to be the most significant variable (Figures S3 and S4). This is not surprising, as Sn itself is known as an active metal for reducing CO₂ to CO. However, the CO current density of a codeposition of Cu and Sn (2.5 mA cm⁻²) is much higher than a deposition of Sn (0.5 mA cm⁻²) alone. For FE-C₂₊, feature importance analysis from RF and GBCT showed that the aliphatic OH group on the organic additive is most important (Figures S5 and S6).

We also quantitatively analyzed the FE and *j* using regression models including linear regression, least absolute shrinkage and selection operator (LASSO), gradient boost decision tree regressor (GBDTR), and multilayer perceptron (MLP).^{42–44} Each regression model has its own strength. We thus tested all of them to find the best performing one. All the models are just mathematical fittings without any preset physical assumptions. In this step, we try to find a reliable mathematical correlation, which may lead to chemical findings in further analysis of variables in the mathematical expression. These regression models confirmed the significant and positive roles of Sn for FE-CO (Figure 2a, Figure S7) and the aliphatic OH group for FE-C₂₊ (Figure 2b, Figure S8) and also suggested Ag additive as a promoter for j_{CO} and j_{C2+} (Figure S9, Figure S10).



Figure 2. Feature importance of (a) FE-CO and (b) FE-C_{2+} obtained by gradient boost decision tree regressor analysis for the first round machine-learning. (c) High-performance catalysts selected after the second round of machine learning and (d) corresponding additives for catalysts in c.

Second Round of Learning. To construct the second round of the data set, we zoom-in the regions of high FE-CO, FE-HCOOH, and FE-C₂₊ to add more data points. We prepared another 28 samples with Sn additives and different organic additives. We also added another 47 samples using different organic additives with aliphatic OH groups. Combining the first round and the second round experiments, we obtained the second data set including 187 data points, among which 94 catalysts contained Cu as the only metal and 42 catalysts contained Sn/Cu.

To better represent local structure of the organic additives in the second round, besides the functional group-based featurization, we used molecular fragment featurization (MFF) modified from the extended-connectivity fingerprint (ECFP) method⁴⁵ to extract fragments of a molecule (Figure 3a, Table S6). We obtained 137 fragments from 94 molecules as input variables for Cu catalysts and 70 fragments from 42

pubs.acs.org/JACS



Figure 3. (a) Converting molecular structure to MFF representation. (b) The third round machine learning to extract interaction-based feature combinations using random intersection tree (RIT). (c) Designing new molecular additives based on results from RIT.

molecules as input variables for Sn/Cu catalysts. These MFF features were used for classification, but one-hot functional group features were still used in regression due to their simplicity.

We first analyzed the sum of partial current densities for generating HCOOH and CO ($j_{\text{HCOOH+CO}}$). Regression analyses suggested that besides Sn additive, the carboxyl group on the molecular additive has a positive effect (Figures S16).

We then proceeded to analyze the ratio of FEs between generating HCOOH and generating CO (FE-HCOOH/FE-CO). We set FE-HCOOH/FE-CO = 1 as a borderline for 0/1 assignment in the classification. RF and GBCT analyses both showed that the nitrogen-containing groups on the organic additives are important (Figures S13 and S14). Regression analyses confirmed that the aliphatic amino group favors generating HCOOH while nitrogen heteroaromatic ring favors generating CO (Figures S15).

On the basis of these analyses, we used ligands containing a nitrogen heteroaromatic ring and a carboxyl group to achieve high FE-CO (sample 215: Sn + tryptophan as additives, FE-CO = 91%, j_{CO} = 10 mA cm⁻²) and used ligands containing aliphatic amino and carboxyl groups to obtain high FE-

HCOOH (sample 216: Sn + L-aspartic acid as additives, FE-HCOOH = 65%, j_{HCOOH} = 20.5 mA cm⁻²) (Figure 2c,d, Table S2).

Local pH should be an important parameter to dictate catalytic selectivity. We added the local pH as an additional variable in the machine learning of FE-HCOOH/FE-CO. The local pH can be calculated using current density according to

$$pH_{local} = pH_{bulk} + log\left(1 + \frac{j_{OH^-}}{j_d}\right)$$
(1)

where

$$j_{\rm d} = F \frac{D_{\rm CO_3^2} - [\rm CO_3^{2-}]_{\rm bulk}}{l}, j_{\rm OH^-} = j - 0.5 \times j_{\rm HCOOH}$$
 (2)

which has been detailed in our previous work.⁴⁶ GBDTR showed that the most important factor is the local pH (Figure 4a, Figure S17). At a higher local pH, most catalysts exhibit a higher ratio of FE-HCOOH/FE-CO (Figure 4b). It is consistent with the notion that a basic environment disfavors C–O bond cleavage. This finding on local pH is similar to that of an earlier report of the pH effect of $[Ru(bpy)_2(CO)_2]^{2+}$



Figure 4. (a) Feature importance of FE-HCOOH/FE-CO using GBDTR for Sn/Cu data set. (b) The same catalysts exhibit a higher ratio of FE-HCOOH/FE-CO when the local pH is increased.

catalyst in $\rm CO_2 RR^{47}$ but is found for the first time in a Sn/Cu system.

For FE-C₂₊ and j_{C2+} of the Cu catalysts, ML of the second data set confirmed the importance of the aliphatic OH group (Figures S18–S21, Table S7). Linear regression also showed that an aliphatic amine had a positive effect while N-containing aromatic heterocycles and an aromatic amine had negative effects on FE-C₂₊ (Figure S20).

These findings can be verified by visualizing distribution of FE-C₂₊ of additives with and without these functional groups (Figure S22), the significance of which was further verified by testing null-hypothesis using the Wilcoxon rank test method (Figure S23).

Consistent with these analyses, the catalyst prepared with sorbitol as the additive gave the highest C_{2+} selectivity of 53% and partial current density of 4.0 mA cm⁻² in CO₂ saturated 0.1 M KHCO₃ in H-type electrolyzer for these products (Figure 2c,d). We then loaded this catalyst to a gas diffusion electrode and tested its performance in a flow-cell system. A current density of 250 mA cm⁻² with FE-C₂₊ of 63% was achieved in 1.0 M KOH at -0.97 V_{RHE} (Table S3).

Third Round of Learning. The above feature importance analyses focused on a particular feature of the additive. However, a pair of features may work cooperatively. To search for such a positive or negative synergistic effect, we used a "random intersection tree" (RIT)48 in the third round of learning to examine important variable combinations (Figure 3). The RIT algorithm can quickly extract interaction-based feature combinations even when the number of all possible combinations is large. Key to this algorithm is to find common features of positive samples through an intersection operation to construct tree structures with feature subsets as the nodes. A total of 1495 such subsets was obtained from 1000 trees. An "Iscore" was used to evaluate their importance, and 23 subsets with the highest I-scores were selected, each of which contained 2 to 7 features. We then predicted FE-C2+ of these 23 sets using classification models obtained from the second round ML using LR, RF, GBCT, and XGBoost. Only 11 out of the 23 feature subsets have consistent predictions from the four models. These 11 subsets showed that combinations of the aliphatic hydroxyl group with aliphatic carboxylic acids, an aliphatic amine, or aliphatic ammonium salts tend to enhance FE-C2+, while combinations of aromatic rings, aromatic carboxylic acids and nitrogen-containing heterocycles tend to reduce FE-C2+.

We designed 24 molecules following feature combinations in the 11 subsets (Table S8). A voting regressor combining MLP, LASSO, and GBDTR models from the second round of ML assessed FE-C₂₊ of these 24 additives (Table S9, Figure S24, Figure 5a). Scores of synthetic Bayesian accessibility (SYBA)⁴⁹





Figure 5. (a) Prediction of FE-C_{2+} of designed catalysts in the 3rd round of learning and the corresponding SYBA score. (b) Experimental FE-C_{2+} of corresponding predicted catalysts (circled in part a).

were also calculated to predict whether the designed molecules were easy to synthesize. From the 24 molecules, 3 commercially available ones with quite different FE-C₂₊ predictions were chosen to be experimentally tested in Htype electrolyzer (Figure 5b, Table S2). The experimental result roughly confirmed the prediction (FE-C₂₊ for 223 *meso*erythritol: 28%; 224 4-methylpyridine-2-carboxylic acid: 7%; 225 4-methylpyridine-2-carboxylic acid: 0%). The *meso*erythritol additive gave a high C₂₊ selectivity and was then tested in a flow cell using a gas-diffusion electrode in 1.0 M KOH to give FE-C₂₊ of 64% with a current density of 300 mA cm⁻² as shown in Table S3.

Searching for the Function of the Additives. The organic additives played an important role in Cu catalyst preparation, but how do they affect the catalytic process? We first performed an infrared (IR) measurement of several of the catalysts to see whether the organic molecules adhere to the metal surface. We did not observe signals of the organic molecules, although we cannot rule out small amounts of surface-attached molecules that are below the detection limit of IR spectroscopy (Figure S25). Furthermore, adding the promoting organic additives to CO_2RR solution does not lead to enhanced performance (Figure S26). The additive organic molecules are thus unlikely to directly participate in the CO_2RR process.

We then analyzed powder X-ray diffraction (PXRD) patterns of the Cu depositions before catalysis (Figures S27-S32). We found Cu₂O phases (peaks at 36.5° , 42.5° , 61.5° , JCPDS no. 65-3288) in several Cu preparations especially those with aliphatic OH groups on the additives, while the Cu₂O phase was not found in many other preparations (e.g., with pyridine as the additive). We also searched for signs of CuO or Cu(OH)₂ by PXRD and XPS but could not find them in the samples.

We thus hypothesized that one function of the organic additive is to control the phase of the deposits. Figure 6a shows the relationship between the presence/absence of the Cu₂O phase in the deposits and the FE-CO₂RR, revealing that the FE-CO₂RR is correlated with the Cu₂O phase in the deposits. For Sn/Cu catalysts, we found that addition of Sn induces formation of Cu₂O (Figure S34). This finding is consistent with many literature reports of high activity of oxide-derived Cu (OD-Cu).^{31–35} X-ray photoelectron spectroscopy and Auger electron spectroscopy of Cu showed that the sample prepared with an organic additive with an aliphatic OH group had more Cu¹ before catalysis (Figures S35–S37), although Cu in all samples was reduced into Cu⁰ after 1 h catalysis (Figure S33). The amount of Cu deposited does not have a significant effect on product selectivity (Figure S38).



Figure 6. (a) Relationship between FE-CO₂RR, FE-H₂, and the XRD phase of the samples. Samples in the red area contain the Cu₂O phase while samples in the blue area do not. (b) Relationship between FE-C₂₊ and the morphology of the samples. Samples in the red area contain the Cu₂O cube while samples in the blue area do not. (c) SEM images of sample 167 without the Cu₂O cube. (d) SEM images of sample 170 with the Cu₂O cube.

We also analyzed electron microscopy (EM) images of the catalysts (Figures S39–S43). We found that catalysts with higher FE-C₂₊ showed more regular Cu₂O cubes before catalysis (Figure 6b–d), although shapes of these cubes were lost after electrolysis of 20 min (Figure S41). Recent literature reports also showed that Cu₂O cubes as a precursor can enhance the selectivity of C₂₊ products.⁵⁰ Besides the Cu₂O cubes, we investigated Cu₂O nanoparticles enclosed by different crystal facets (nos. 170, 150, 131) for C₂₊ products. High-resolution transmission electron microscopy (HRTEM) and selected area electron diffraction (SAED) (Figure S42) showed that sample 170 has exposed {001} facets, and sample 150 has some {11–2} facets, while sample 131 has some {1-1-1} facets besides the {001} facets. Although their exposed crystal faces are different, they all show similarly high FE-C₂₊ (170:52.9%, 150:46.2%, 131:51.7%).

By contrast, sample 167 (Figure S43) contained fragmented polycrystalline Cu₂O nanoparticles and showed low FE-C₂₊ (167:0%). We found that catalysts with higher FE-C₂₊ showed more regular Cu₂O crystals in the precursor before catalysis while it seems that there is no direct correlation between the exposed crystal face and FE-C₂₊. This finding is different from the reported correlation between the exposed crystal planes of Cu₂O and FE-C₂₊. ^{51,52}

EXPERIMENTAL METHODS

Materials and Apparatus. Reagents were commercially available and used without further purification unless otherwise noted. Powder X-ray diffraction (PXRD) was carried out on a Rigaku DMax- γ A rotation anode X-ray diffractometer equipped with graphite monochromatized Cu K α radiation ($\gamma = 1.54$ Å). IR spectra were obtained on a Nicolet iS50 FTIR spectrometer. Scanning electron microscopy images were obtained on a Zeiss sigma. ¹H NMR spectra were recorded on a Bruker NMR 500 DRX spectrometer at 500 MHz. Transmission electron microscopy (TEM) images were acquired on a JEOL 2100 high resolution transmission electron microscope. The XPS measurements were performed using a PHI Quantum 2000 instrument.

Synthetic Procedures. Synthetic Conditions for Cu-Mixed Precursor. Cu-mixed precursors were made from 0.1 M $CuSO_4$: SH_2O with or without a 10 mM concentration of another metal salt

(SnCl₂·2H₂O, CoCl₂·6H₂O, NiCl₂·6H₂O, ZnCl₂, FeSO₄·7H₂O, AgNO₃, In(NO₃)₃·xH₂O, Pd(NO₃)₂, Bi(NO₃)₃, Ce(NO₃)₃, HAuCl₄) and/or a 10 mM concentration of organic additive in 0.1 M KCl aqueous solution.

Synthetic Conditions for Cu-Mixed Catalyst. Cu-mixed samples were electrodeposited in a plating bath made from the above Cu-mixed precursor. Carbon cloth was used as a substrate for electrodeposition. Cu-mixed catalyst was in situ electrodeposited at a constant potential -1.0 V vs Ag/AgCl until a final deposition charge of 1.9 C/cm² was reached on the carbon cloth. A graphitic sheet was used as the counter electrode, and a Ag/AgCl electrode was placed near the working electrode as the reference electrode.

Electrochemistry. Electrochemical measurements were performed using a three-electrode system and an electrochemical workstation (CHI660E). Electrolysis was performed at room temperature in a H type cell with a Ag/AgCl reference electrode and a graphitic sheet counter electrode. The cathode and anode compartments were separated by a proton exchange membrane (Nafion 117). The potentials on the working electrodes were converted to voltages with respect to the RHE reference electrode by E (vs the RHE) = E (vs the Ag/AgCl) + 0.20 V + 0.0591 × pH. Catalytic results with a H-type cell are shown in Tables S1 and S2. Electrolysis was also performed at room temperature in a flow cell with an Ag/AgCl reference electrode and a nickel foam counter electrode. Catalytic results with the flow cell are shown in Table S3.

Product Analysis. Liquid-phase products were analyzed by proton nuclear magnetic resonance (¹H NMR) spectroscopy (Bruker AVANCE AV III 500), in which 0.5 mL of the electrolysis solution was mixed with 0.1 mL of deuterated water (D₂O) for field locking and 0.02 μ L of dimethyl sulfoxide (DMSO) as an internal standard. The ¹H NMR spectrum was measured with water suppression using a presaturation sequence. Gas-phase products were detected online by using a gas chromatograph (GC) connected to the headspace of the electrolysis cell. A thermal conductivity detector (TCD) was used to quantify hydrogen, and a flame ionization detector (FID) equipped with a methanizer was used to quantify carbon monoxide, methane, ethane, and ethylene.

Machine Learning Methods. Feature Selection Using a Random Intersection Tree. A "random intersection tree" (RIT) can quickly extract interaction-based feature combinations.⁴⁸ The target property (denoted y) for the RIT analysis is labeled either '1' (positive sample) or '0' (negative sample). Here we chose Faradaic efficiency as the target property. We set a criteria for Faradaic efficiency, above which γ is labeled '1' while below it is labeled '0' to convert the data into binary combinations that are suitable for RIT analysis. A positive sample was then randomly chosen as the root node to construct a tree (RIT). A child node of the root node was constructed by randomly choosing another positive sample and calculating the common features of the root sample and the newly chosen one, which is called an intersection operation. These common features constitute one child node. Several other children nodes were obtained by randomly choosing other positive samples to undergo the intersection operation with the common root node to complete a second layer of the tree. The third layer of the tree was constructed by repeating the intersection operation between the second layer and randomly chosen positive samples. The algorithm will stop if the selected features of a node do not show up significantly more often in positive samples than in negative samples. A leaf node thus contains a feature subset that is important in determining the target property. Negative samples are also used for the intersection operation but with a modified rule. Features appearing in a negative sample will be given penalties to decrease their likelihood to constitute positive subsets.

In building the RITs, the number of branches from each nonleaf node is set to 3, the maximum depth is set to 9, and the maximum number of features in a returned subset is set to 15. Positive samples are randomly collected as "root nodes" to produce 1000 RITs. All these trees acted like a forest to yield 1495 possible subsets. An "I-score" is used in the RIT analysis,⁴⁸ which can be computed

An "I-score" is used in the RIT analysis,⁴⁸ which can be computed by the formula

Article

$$I = \sum_{j \in P_k} \left[n_1(j) - \overline{n}_1(j) \right]^2$$

where P_k is a partition consisting of k features (a feature subset or the leaf node of the size of k), j is a partition element among 2^k possibilities, $\overline{n}_1(j)$ is the number of positive samples whose feature values are corresponding to j, and $\overline{n}_1(j)$ is the expected number of positive samples according to the occurrence frequency of *j* and positive ratio observed from the whole data set. The I-score can be normalized by dividing $n\sigma^2$, where *n* is the number of samples and σ^2 is the variance of label values y of all samples. The higher the I-score, the more significant the selected features. A greedy algorithm named "backward dropping algorithm" is thus implemented to drop features until all the subsets generated by RIT reach their highest I-score. The final return sets after the entire process can be considered as the most important subsets extracted from the original data set. Further tests are needed to tell whether they are positively or negatively related to the target label. The I-scores of all 1495 sets were calculated, and the 23 highest were put into the backward dropping algorithm for further refinement to obtain I-scores between 1.5675 and 3.8937. Only 2 to 7 features were contained in each one of these 23 sets. For example, one such set with an I-score of 2.5406 contains 'With 2 CO', 'With 1 CC', 'With 2 CCO', and 'With 1 CC(C)(C)C', suggesting that these feature combinations might be influential for the label values.

CONCLUSIONS

In summary, we accelerate the discovery of electrodeposition additives to prepare Cu catalysts for CO_2RR using machinelearning. We found that molecules with an aliphatic OH group as the additive promotes faradaic efficiency in generating C_{2+} products, possibly via controlling formation of Cu₂O cubes in the catalyst precursor. This Cu₂O phase is also important in the Sn/Cu bimetallic system to generate CO/HCOOH in CO₂RR. Currently, ML-assisted catalyst development is still limited by the number of experimental data. High throughput screening methods may combine with ML strategy to open new horizons in catalyst discovery.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c00339.

Characterization of catalysts, electrochemical measurements, and machine learning procedure (PDF)

AUTHOR INFORMATION

Corresponding Authors

- Cheng Wang State Key Laboratory of Physical Chemistry of Solid Surfaces, iChem, Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China;
 orcid.org/0000-0002-7906-8061; Email: wangchengxmu@xmu.edu.cn
- Da Zhou School of Mathematical Sciences and Fujian Provincial Key Laboratory of Mathematical Modeling and High-Performance Scientific Computation, Xiamen University, Xiamen 361005, P. R. China; Email: zhouda@ xmu.edu.cn

Authors

Ying Guo – State Key Laboratory of Physical Chemistry of Solid Surfaces, iChem, Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China; School of Chemistry and Chemical Engineering, Northwestern Polytechnical University, Xi'an 710129, P. R. China

- Xinru He State Key Laboratory of Physical Chemistry of Solid Surfaces, iChem, Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China
- Yuming Su State Key Laboratory of Physical Chemistry of Solid Surfaces, iChem, Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China
- Yiheng Dai State Key Laboratory of Physical Chemistry of Solid Surfaces, iChem, Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China
- Mingcan Xie State Key Laboratory of Physical Chemistry of Solid Surfaces, iChem, Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China
- Shuangli Yang State Key Laboratory of Physical Chemistry of Solid Surfaces, iChem, Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China
- Jiawei Chen State Key Laboratory of Physical Chemistry of Solid Surfaces, iChem, Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China
- Kun Wang School of Mathematical Sciences and Fujian Provincial Key Laboratory of Mathematical Modeling and High-Performance Scientific Computation, Xiamen University, Xiamen 361005, P. R. China

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c00339

Author Contributions

[#]These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge funding support from the Ministry of Science and Technology of China (2016YFA0200702), the National Natural Science Foundation of China (no. 22071207, no. 21721001, no. 11971405), NFFTBS (no. J1310024)

REFERENCES

(1) Ross, M. B.; De Luna, P.; Li, Y.; Dinh, C.-T.; Kim, D.; Yang, P.; Sargent, E. H. Designing materials for electrochemical carbon dioxide recycling. *Nat. Catal* **2019**, *2* (8), 648–658.

(2) Zheng, Y.; Vasileff, A.; Zhou, X.; Jiao, Y.; Jaroniec, M.; Qiao, S.-Z. Understanding the Roadmap for Electrochemical Reduction of CO₂ to Multi-Carbon Oxygenates and Hydrocarbons on Copper-Based Catalysts. J. Am. Chem. Soc. **2019**, 141 (19), 7646–7659.

(3) Zhang, S.; Fan, Q.; Xia, R.; Meyer, T. J. CO_2 Reduction: From Homogeneous to Heterogeneous Electrocatalysis. *Acc. Chem. Res.* **2020**, 53 (1), 255–264.

pubs.acs.org/JACS

(4) Diercks, C. S.; Liu, Y.; Cordova, K. E.; Yaghi, O. M. The role of reticular chemistry in the design of CO_2 reduction catalysts. *Nat. Mater.* **2018**, *17* (4), 301–307.

(5) Jiao, Y.; Zheng, Y.; Chen, P.; Jaroniec, M.; Qiao, S.-Z. Molecular Scaffolding Strategy with Synergistic Active Centers To Facilitate Electrocatalytic CO_2 Reduction to Hydrocarbon/Alcohol. *J. Am. Chem. Soc.* **2017**, *139* (49), 18093–18100.

(6) Tran, K.; Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal* **2018**, 1 (9), 696–703.

(7) Ma, X.; Li, Z.; Achenie, L. E.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO_2 Electroreduction Catalyst Screening. J. Phys. Chem. Lett. **2015**, 6 (18), 3528–33.

(8) Huang, Y.; Chen, Y.; Cheng, T.; Wang, L.-W.; Goddard, W. A. Identification of the Selective Sites for Electrochemical Reduction of CO to C_{2+} Products on Copper Nanoparticles by Combining Reactive Force Fields, Density Functional Theory, and Machine Learning. ACS Energy Lett. **2018**, 3 (12), 2983–2988.

(9) Wang, X.; Ye, S.; Hu, W.; Sharman, E.; Liu, R.; Liu, Y.; Luo, Y.; Jiang, J. Electric Dipole Descriptor for Machine Learning Prediction of Catalyst Surface–Molecular Adsorbate Interactions. *J. Am. Chem. Soc.* **2020**, *142* (17), 7737–7743.

(10) Sun, M.; Dougherty, A. W.; Huang, B.; Li, Y.; Yan, C.-H. Accelerating Atomic Catalyst Discovery by Theoretical Calculations-Machine Learning Strategy. *Adv. Energy Mater.* **2020**, *10* (12), 1903949.

(11) Artrith, N.; Lin, Z.; Chen, J. G. Predicting the Activity and Selectivity of Bimetallic Metal Catalysts for Ethanol Reforming using Machine Learning. *ACS Catal.* **2020**, *10* (16), 9438–9444.

(12) Meyer, B.; Sawatlon, B.; Heinen, S.; von Lilienfeld, O. A.; Corminboeuf, C. Machine learning meets volcano plots: computational discovery of cross-coupling catalysts. *Chem. Sci.* **2018**, *9* (35), 7069–7077.

(13) Moosavi, S. M.; Jablonka, K. M.; Smit, B. The Role of Machine Learning in the Understanding and Design of Materials. *J. Am. Chem. Soc.* **2020**, *142* (48), 20273–20287.

(14) Zhong, M.; Tran, K.; Min, Y.; Wang, C.; Wang, Z.; Dinh, C. T.; De Luna, P.; Yu, Z.; Rasouli, A. S.; Brodersen, P.; Sun, S.; Voznyy, O.; Tan, C. S.; Askerka, M.; Che, F.; Liu, M.; Seifitokaldani, A.; Pang, Y.; Lo, S. C.; Ip, A.; Ulissi, Z.; Sargent, E. H. Accelerated discovery of CO_2 electrocatalysts using active machine learning. *Nature* **2020**, *581* (7807), 178–183.

(15) Zhan, C.-H.; Zheng, Q.; Long, D.-L.; Vilà-Nadal, L.; Cronin, L. Controlling the Reactivity of the $[P_8W_{48}O_{184}]^{40-}$ Inorganic Ring and Its Assembly into POMZite Inorganic Frameworks with Silver Ions. *Angew. Chem., Int. Ed.* **2019**, *58* (48), 17282–17286.

(16) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of higher-selectivity catalysts by computerdriven workflow and machine learning. *Science* **2019**, *363* (6424), eaau5631.

(17) Ross, M. B.; Dinh, C. T.; Li, Y.; Kim, D.; De Luna, P.; Sargent, E. H.; Yang, P. Tunable Cu Enrichment Enables Designer Syngas Electrosynthesis from CO₂. J. Am. Chem. Soc. **2017**, 139 (27), 9359–9363.

(18) Wang, Y.; Chen, Z.; Han, P.; Du, Y.; Gu, Z.; Xu, X.; Zheng, G. Single-Atomic Cu with Multiple Oxygen Vacancies on Ceria for Electrocatalytic CO_2 Reduction to CH_4 . ACS Catal. **2018**, 8 (8), 7113–7119.

(19) Kim, C.; Weng, L.-C.; Bell, A. T. Impact of Pulsed Electrochemical Reduction of CO_2 on the Formation of C_{2+} Products over Cu. ACS Catal. 2020, 10 (21), 12403–12413.

(20) Chou, T.-C.; Chang, C.-C.; Yu, H.-L.; Yu, W.-Y.; Dong, C.-L.; Velasco-Vélez, J.-J.; Chuang, C.-H.; Chen, L.-C.; Lee, J.-F.; Chen, J.-M.; Wu, H.-L. Controlling the Oxidation State of the Cu Electrode and Reaction Intermediates for Electrochemical CO₂ Reduction to Ethylene. J. Am. Chem. Soc. **2020**, 142 (6), 2857–2867.

(21) Li, J.; Kuang, Y.; Meng, Y.; Tian, X.; Hung, W.-H.; Zhang, X.; Li, A.; Xu, M.; Zhou, W.; Ku, C.-S.; Chiang, C.-Y.; Zhu, G.; Guo, J.; Sun, X.; Dai, H. Electroreduction of CO₂ to Formate on a CopperBased Electrocatalyst at High Pressures with High Energy Conversion Efficiency. J. Am. Chem. Soc. 2020, 142 (16), 7276–7282.

(22) Yang, K. D.; Ko, W. R.; Lee, J. H.; Kim, S. J.; Lee, H.; Lee, M. H.; Nam, K. T. Morphology-Directed Selective Production of Ethylene or Ethane from CO_2 on a Cu Mesopore Electrode. *Angew. Chem., Int. Ed.* **2017**, *56* (3), 796–800.

(23) Dutta, A.; Rahaman, M.; Luedi, N. C.; Mohos, M.; Broekmann, P. Morphology Matters: Tuning the Product Distribution of CO_2 Electroreduction on Oxide-Derived Cu Foam Catalysts. *ACS Catal.* **2016**, *6* (6), 3804–3814.

(24) De Luna, P.; Quintero-Bermudez, R.; Dinh, C.-T.; Ross, M. B.; Bushuyev, O. S.; Todorović, P.; Regier, T.; Kelley, S. O.; Yang, P.; Sargent, E. H. Catalyst electro-redeposition controls morphology and oxidation state for selective carbon dioxide reduction. *Nat. Catal* **2018**, *1* (2), 103–110.

(25) Scholten, F.; Sinev, I.; Bernal, M.; Roldan Cuenya, B. Plasma-Modified Dendritic Cu Catalyst for CO_2 Electroreduction. *ACS Catal.* **2019**, 9 (6), 5496–5502.

(26) Dutta, A.; Rahaman, M.; Mohos, M.; Zanetti, A.; Broekmann, P. Electrochemical CO_2 Conversion Using Skeleton (Sponge) Type of Cu Catalysts. ACS Catal. **2017**, 7 (8), 5431–5437.

(27) Jiang, K.; Sandberg, R. B.; Akey, A. J.; Liu, X.; Bell, D. C.; Nørskov, J. K.; Chan, K.; Wang, H. Metal ion cycling of Cu foil for selective C–C coupling in electrochemical CO₂ reduction. *Nat. Catal* **2018**, *1* (2), 111–119.

(28) Hoang, T. T. H.; Ma, S.; Gold, J. I.; Kenis, P. J. A.; Gewirth, A. A. Nanoporous Copper Films by Additive-Controlled Electrodeposition: CO_2 Reduction Catalysis. *ACS Catal.* **2017**, 7 (5), 3313–3321.

(29) Popović, S.; Smiljanić, M.; Jovanovič, P.; Vavra, J.; Buonsanti, R.; Hodnik, N. Stability and Degradation Mechanisms of Copper-Based Catalysts for Electrochemical CO_2 Reduction. *Angew. Chem., Int. Ed.* **2020**, *59* (35), 14736–14746.

(30) Han, Z.; Kortlever, R.; Chen, H.-Y.; Peters, J. C.; Agapie, T. CO_2 Reduction Selective for $C \ge 2$ Products on Polycrystalline Copper with N-Substituted Pyridinium Additives. *ACS Cent. Sci.* **2017**, 3 (8), 853–859.

(31) Arán-Ais, R. M.; Scholten, F.; Kunze, S.; Rizo, R.; Roldan Cuenya, B. The role of in situ generated morphological motifs and Cu(i) species in C_{2+} product selectivity during CO₂ pulsed electroreduction. *Nat. Energy* **2020**, *5* (4), 317–325.

(32) Yi, J.-D.; Xie, R.; Xie, Z.-L.; Chai, G.-L.; Liu, T.-F.; Chen, R.-P.; Huang, Y.-B.; Cao, R. Highly Selective CO_2 Electroreduction to CH_4 by In Situ Generated Cu_2O Single-Type Sites on a Conductive MOF: Stabilizing Key Intermediates with Hydrogen Bonding. *Angew. Chem., Int. Ed.* **2020**, *59* (52), 23641–23648.

(33) Zhang, W.; Huang, C.; Xiao, Q.; Yu, L.; Shuai, L.; An, P.; Zhang, J.; Qiu, M.; Ren, Z.; Yu, Y. Atypical Oxygen-Bearing Copper Boosts Ethylene Selectivity toward Electrocatalytic CO₂ Reduction. *J. Am. Chem. Soc.* **2020**, *142* (26), 11417–11427.

(34) Lee, S. Y.; Jung, H.; Kim, N.-K.; Oh, H.-S.; Min, B. K.; Hwang, Y. J. Mixed Copper States in Anodized Cu Electrocatalyst for Stable and Selective Ethylene Production from CO₂ Reduction. *J. Am. Chem. Soc.* **2018**, *140* (28), 8681–8689.

(35) Lum, Y.; Ager, J. W. Stability of Residual Oxides in Oxide-Derived Copper Catalysts for Electrochemical CO_2 Reduction Investigated with ¹⁸O Labeling. *Angew. Chem., Int. Ed.* **2018**, 57 (2), 551–554.

(36) Fang, K. T.; Lin, D. K. J.; Winker, P.; Zhang, Y. Uniform design: Theory and application. *Technometrics* **2000**, *42* (3), 237–248.

(37) Quinlan, J. R. Induction of decision trees. *Machine Learning* **1986**, *1* (1), 81–106.

(38) Hosmer, D. W., Jr.; Lemeshow, S.; Sturdivant, R. X. Applied logistic regression; John Wiley & Sons, 2013; Vol. 398.

(39) Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **2001**, *29* (5), 1189–1232.

(40) Breiman, L. Random forests. Mach Learn 2001, 45 (1), 5-32.

(41) Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, 2016; pp 785–794.

(42) Kim, S.-J.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An Interior-Point Method for Large-Scale -Regularized Least Squares. *IEEE Journal of Selected Topics in Signal Processing* **2007**, *1* (4), 606–617.

(43) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature* **1986**, 323 (6088), 533–536.

(44) Ye, J.; Chow, J.-H.; Chen, J.; Zheng, Z. Stochastic gradient boosted distributed decision trees. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*; Association for Computing Machinery: Hong Kong, China, 2009; pp 2061–2064.

(45) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. J. Chem. Inf. Model. 2010, 50 (5), 742-754.

(46) Guo, Y.; Shi, W.; Yang, H.; He, Q.; Zeng, Z.; Ye, J. Y.; He, X.; Huang, R.; Wang, C.; Lin, W. Cooperative Stabilization of the [Pyridinium-CO₂-Co] Adduct on a Metal–Organic Layer Enhances Electrocatalytic CO₂ Reduction. J. Am. Chem. Soc. **2019**, 141 (44), 17875–17883.

(47) Ishida, H.; Tanaka, K.; Tanaka, T. Electrochemical CO₂ reduction catalyzed by ruthenium complexes $[Ru(bpy)_2(CO)_2]^{2+}$ and $[Ru(bpy)_2(CO)Cl]^+$. Effect of pH on the formation of CO and HCOOH. *Organometallics* **1987**, *6* (1), 181–186.

(48) Shah, R. D.; Meinshausen, N. Random Intersection Trees. J. Mach Learn Res. 2014, 15 (1), 629–654.

(49) Voršilák, M.; Kolář, M.; Čmelo, I.; Svozil, D. SYBA: Bayesian estimation of synthetic accessibility of organic compounds. *J. Cheminf.* **2020**, *12*, 35.

(50) Aran-Ais, R. M.; Rizo, R.; Grosse, P.; Algara-Siller, G.; Dembele, K.; Plodinec, M.; Lunkenbein, T.; Chee, S. W.; Cuenya, B. R. Imaging electrochemically synthesized Cu_2O cubes and their morphological evolution under conditions relevant to CO_2 electroreduction. *Nat. Commun.* **2020**, *11* (1), 3489.

(51) Möller, T.; Scholten, F.; Thanh, T. N.; Sinev, I.; Timoshenko, J.; Wang, X.; Jovanov, Z.; Gliech, M.; Roldan Cuenya, B.; Varela, A. S.; Strasser, P. Electrocatalytic CO_2 Reduction on CuO_x Nanocubes: Tracking the Evolution of Chemical State, Geometric Structure, and Catalytic Selectivity using Operando Spectroscopy. *Angew. Chem., Int. Ed.* **2020**, *59* (41), 17974–17983.

(52) Gao, Y.; Wu, Q.; Liang, X.; Wang, Z.; Zheng, Z.; Wang, P.; Liu, Y.; Dai, Y.; Whangbo, M.-H.; Huang, B. Cu2O Nanoparticles with Both $\{100\}$ and $\{111\}$ Facets for Enhancing the Selectivity and Activity of CO₂ Electroreduction to Ethylene. *Advanced Science* **2020**, 7 (6), 1902820.