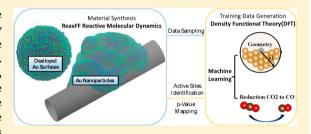


# Identifying Active Sites for CO<sub>2</sub> Reduction on Dealloyed Gold Surfaces by Combining Machine Learning with Multiscale Simulations

Yalu Chen, To Yufeng Huang, To Cheng, And William A. Goddard, III\*

Supporting Information

ABSTRACT: Gold nanoparticles (AuNPs) and dealloyed Au<sub>3</sub>Fe core-shell NP surfaces have been shown to have dramatically improved performance in reducing CO<sub>2</sub> to CO (CO2RR), but the surface features responsible for these improvements are not known. The active sites cannot be identified with surface science experiments, and quantum mechanics (QM) is not practical for the 10 000 surface sites of a 10 nm NP (200 000 bulk atoms). Here, we combine machine learning, multiscale simulations, and QM to predict the performance (a-value) of all 5000-10 000 surface sites on AuNPs



and dealloyed Au surfaces. We then identify the optimal active sites for CO2RR on dealloyed gold surfaces with dramatically reduced computational effort. This approach provides a powerful tool to visualize the catalytic activity of the whole surface. Comparing the a-value with descriptors from experiment, computation, or theory should provide new ways to guide the design of high-performance electrocatalysts for applications in clean energy conversion.

# 1. INTRODUCTION

The development of dramatically improved electrocatalysts is essential for economical renewable energy generation, energy storage, and utilization. 1,2 The search for new catalysts has motivated extraordinary efforts in experimental3-8 and computational 9-12 combinatorial assays to discover new alloys or composites with some successes. Complementing this are new generations of quantum mechanics (QM) that have provided remarkable accuracy for simple low-index surfaces: 0.05 eV for free energy reaction barriers at 298 K for oxygen reduction reaction (ORR) on Pt(111)<sup>13</sup> and CO to ethane and ethanol on Cu(100), 14,15 and onset potentials to 0.05 V at 10 meV/cm<sup>2</sup> for ORR on Cu(111)<sup>16,17</sup> and oxygen evolution reaction (OER) on Fe-doped  $\gamma$ -NiOOOH<sup>18</sup> and IrO<sub>2</sub>(110).<sup>19</sup> However, with remarkable improvement of computational tools for low-index surfaces, experiments have shown dramatically improved performance for nanoparticles (NPs),<sup>20-25</sup> and especially for dealloyed NPs.<sup>26-28</sup>

Unfortunately, there are no available experimental techniques to identify the specific active sites for nanoparticles. We have been able to predict the structure of 10 nm NPs (200 000 atoms) computationally using the ReaxFF reactive force field<sup>29</sup> in chemical vapor deposition simulations.<sup>30</sup> However, the accuracy of ReaxFF is not better than 0.25 eV, whereas we need the 0.05 eV accuracy of QM to properly identify the active sites. Additionally, it would be quite impractical to use QM with cluster models of the 5000-10 000 surface sites.

To solve this conundrum, we propose here a machine learning strategy that uses QM calculations with cluster models for hundreds of the 5000-10 000 surface sites of the ReaxFF grown NP, which we train iteratively to attain 0.05 eV accuracy. We then use this ML predictive model to predict the performance of all 5000-10 000 surface sites. We subsequently analyze the sites with the highest predicted performance (avalue) to discover the specific sites responsible for the superior performance. These results can then be used as a guide for both experimental and computational attempts to maximize the number of sites having this performance. In addition, we can compare the predicted performance of all 5000-10 000 sites with other properties that might be measured experimentally (XPS, electronic density of states, local character of s,p,d orbitals, work function) in order to identify descriptors that might be used in monitoring attempts to synthesize materials with a high density of high-performing sites.

We illustrate this new methodology here for the case of AuNPs. Our previous experimental studies of the dealloyed Au<sub>3</sub>Fe core-shell nanoparticles (CSNPs) led to spectacular

Received: May 8, 2019 Published: June 18, 2019

<sup>&</sup>lt;sup>†</sup>Materials and Process Simulation Center (MSC) and Joint Center for Artificial Photosynthesis (JCAP), California Institute of Technology, Pasadena, California 91125, United States

<sup>§</sup>Institute of Functional Nano & Soft Materials (FUNSOM), Jiangsu Key Laboratory for Carbon-Based Functional Materials & Devices, Joint International Research Laboratory of Carbon-Based Functional Materials and Devices, Soochow University, 199 Renai Road, Suzhou, 215123 Jiangsu, P.R. China

Journal of the American Chemical Society

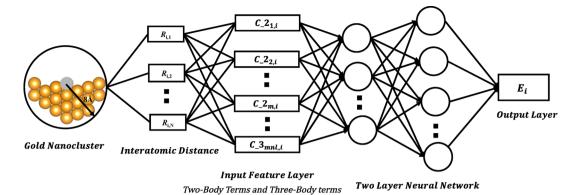


Figure 1. Overall structure of neural network. We use two-body terms (C 2) and three-body terms (C 3) to represent the local geometric feature of the selected surface atom (gray atom) extracted from 8 Å nanocluster model. A fully connected two-layer neural network with 30 nodes in the first layer and 50 nodes in the second layer fits the physical descriptor, with the total number of 2801 model parameters. The training targets  $E_i$  for two neural networks in this work are  $\Delta E_{\rm CO}$  and  $\Delta E_{\rm HOCO}$ , with all training data generated from DFT calculations.

results: a 100-fold increase in mass activity compared with AuNPs and a 500 meV improvement in the overpotential, for >80% faradaic efficiency (FE) of CO<sub>2</sub> to CO<sup>27</sup> compared with Au foils. Similar dramatic improvements from dealloying have also been observed in PtNi alloy catalysts that 2 nm pure Pt nanowires (NWs) formed by dealloying from 5 nm Ni<sub>7</sub>Pt<sub>3</sub> NWs showing ORR performance 50 times better than that of Pt.<sup>28</sup> We expect that this must be related to the nature of the dealloyed surface. In this Article, we predict the performance of the sites achieved by dealloying.

### 2. RESULTS AND DISCUSSION

2.1. Synthesis and Structure Analysis of Gold Nanoparticles. We synthesized the 10 nm AuNPs on the carbon nanotube support using ReaxFF reactive molecular dynamics as described in the Supporting Information, section S1. Surface atoms and bulk atoms are distinguished on the basis of the surface vector method using an 8 Å nanocluster model. We showed that, in our earlier work, 30,31 such 8 Å clusters already include the geometric features that determine the adsorption energy (to an accuracy of 0.02 eV). More details are included in the Supporting Information, section S2. The whole NP system has 211 619 atoms, with 43 200 carbon atoms and 168 419 gold atoms, from which we extracted 16 919 surface atoms using our surface vector method. These 16 919 surfaces sites are partitioned into six groups on the basis of their coordination number, which range from 6 to 11. Here, first neighbors are defined as those lying within a distance of 3.3 Å from center atom. We choose the cutoff distance according to the radial distribution functions in Figure S3a, where the first peak appears at a distance of 2.80 Å and the second peak at 4.20 Å. The surface sites distribution toward coordination number is also shown in Figure S3b.

From the reaction mechanism revealed by our previous work<sup>32</sup> and works from others,<sup>33,34</sup> we choose two physical descriptors for evaluating the activity of a given site:

• CO adsorption energy ( $\Delta E_{CO}$ )

$$\Delta E_{\rm CO} = E_{*\rm CO} - E_* - E_{\rm CO}$$

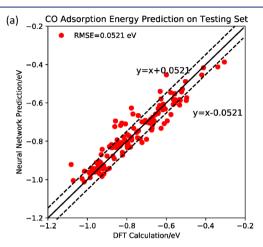
• HOCO transition state formation energy ( $\Delta E_{\text{HOCO}}$ ):

$$\Delta E_{\text{HOCO}} = E_{\text{*HOCO}} - E_{\text{*}} - E_{\text{CO}_{3}} - 0.5E_{\text{H}_{3}}$$

These two descriptors have been proposed and used in other works<sup>30</sup> as well. We expect that sites with high coordination (10 and 11) should be inactive for CO2RR since there is insufficient space to bind reactants, intermediates, or products. Therefore, we investigated these high coordination sites first by randomly selecting 50 sites from the group of coordination 10 and 50 sites from the group of coordination 11. We then carried out DFT calculations (Supporting Information, section S3) for  $\Delta E_{\text{HOCO}}$  on these 100 sites. Among these 100 sites, 96 sites cannot even adsorb the important intermediate of the reaction (HOCO) and the other 4 sites show very high  $\Delta E_{\text{HOCO}}$ . This shows that such high coordination sites are not active. Thus, we narrowed our search to sites with coordination smaller than 10. Even so, there remain 11 537 surface sites to be explored.

2.2. Training Data Generation and Neural Network Model Implementation. We trained two neural networks separately: CO adsorption energy ( $\Delta E_{\rm CO}$ ) and HOCO transition state formation energy ( $\Delta E_{\rm OCOH}$ ). The overall model topology is shown schematically in Figure 1. This type of feature representation method is derived from the work by Behler and Parrinello in 2007.<sup>35</sup> Instead of using Gaussian functions, we choose another type of symmetry function: localized piecewise cosine function. This representative method has earlier been proposed and used by others.<sup>36,37</sup> We used an 8 Å nanocluster from the surface atom to extract local geometric features of selected surface sites as the neural network input through symmetry functions. In our model, we use 12 symmetry functions for two-body terms and 3 symmetry functions for three-body terms, leading to 39 input features. We consider this to give the best balance between training data set size and model complexity. Having defined a set of features, we used a fully connected two-layer neural network with 30 nodes in the first layer and 50 nodes in the second layer for independently fitting two selected physical descriptors:  $\Delta E_{CO}$  and  $\Delta E_{HOCO}$ . The total number of parameters in the model is 2801. Mathematical representations and more details about neural networks are included in the Supporting Information, section S5.

We first trained a neural network for  $\Delta E_{CO}$  using 1384 data from DFT calculations. These data were partitioned into 1104 in the training set, 140 in the validation set, and 140 in the testing set. We constrained the ratio of sites from different coordination groups to be equal within each of the three data sets, but all three sets are totally independent. The validation set is used to prevent overfitting (early stop). Distributions and partitions of the data are listed in Table S1, and the training trajectory is shown in Figure S4. The root mean squared error (RMSE) of the validation set reached the minimum at the training epoch of 19000, with the training set RMSE at 0.0563 eV and the validation set RMSE at 0.0591 eV. The prediction results for the independent testing set are shown in Figure 2a



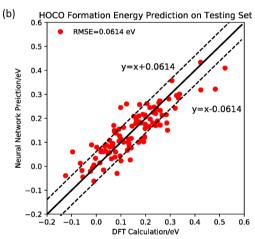
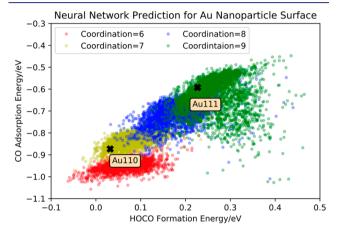


Figure 2. (a) Neural network predictions on the testing set for CO adsorption energy ( $\Delta E_{\rm CO}$ ) shows an RMSE = 0.0521 eV. (b) Neural network predictions on the testing set for HOCO formation energy  $\Delta E_{\rm HOCO}$  shows an RMSE = 0.0614 eV.

leading to RMSE = 0.0521 eV. Using the same approach, we trained our second neural network for  $\Delta E_{\rm HOCO}$  using 1059 data points from DFT calculations. Distribution and partition of the data are listed in Table S2, and the training trajectory is shown in Figure S4b. The validation set RMSE reached the minimum at the training epoch of 11000, with the RMSE of the training set at 0.0616 eV and the validation set RMSE = 0.0591 eV. Figure 2b shows the prediction results on the independent testing set of  $\Delta E_{\rm HOCO}$ , showing RMSE = 0.0614 eV. Our two machine learning models based on the neural network algorithm show RMSE = 0.0521 eV for CO adsorption energy and 0.0614 eV for HOCO formation energy on testing sets. To place our model accuracy in a more straightforward context, we compared our errors to a similar work in predicting CO adsorption energy in thiolated Agalloyed Au nanoclusters, <sup>38</sup> which finds a much higher RMSE  $\approx$ 

0.17 eV using over 2000 data points for training. Another work using machine learning for predicting adsorption energies of  $CH_4$ -related species ( $CH_3$ ,  $CH_2$ , CH, C, and H) on the Cubased alloys<sup>39</sup> reported the best performance with RMSE  $\approx 0.3$  eV after an extra tree regression algorithm. Our model complexity (determined by feature representation and neural network structure) and data set size have the best balance, giving much smaller errors compared to previous works.

**2.3. Model Application.** 2.3.1. Identification of Active Sites on the AuNP Surface. We used the two neural networks trained above to predict  $\Delta E_{\rm CO}$  and  $\Delta E_{\rm HOCO}$  for all 11 537 surface sites obtained as described in section 2.1. The prediction results are shown in Figure 3, where  $\Delta E_{\rm HOCO}$  is



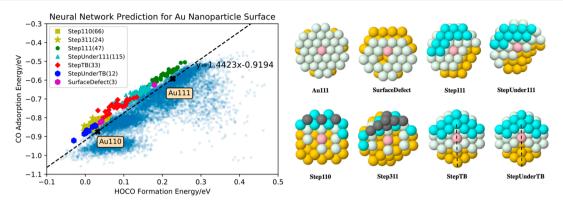
**Figure 3.** Neural network predictions for all 11 537 lower coordination surface sites of AuNPs.  $\Delta E_{\rm CO}$  is placed along the *y*-axis and  $\Delta E_{\rm HOCO}$  along the *x*-axis. All sites are colored on the basis of their coordination number, with Au(111) and Au(110) marked for reference. The general linear relationship between  $\Delta E_{\rm CO}$  and  $\Delta E_{\rm HOCO}$  remains, and we observe a correlation between coordination and catalytic activity.

along the x axis and  $\Delta E_{\rm CO}$  along the y axis. We show with different colors surface sites having different coordination. We observe a correlation between catalytic behavior and coordination. The group with coordination six shows both strongest CO binding and lowest HOCO formation energy while the group with coordination nine shows the weakest CO binding and highest HOCO formation energy. We see a general linear relationship between  $\Delta E_{\rm CO}$  and  $\Delta E_{\rm HOCO}$ , which is consistent with earlier studies.  $^{30,40}$ 

We mark the characteristics for two low-index surfaces (Au(111) and Au(110)) in Figure 3. The (111) surface is generally the most stable for Au, while (110) edge sites were proposed as the most active sites in earlier studies. In general, the predicted performances of the other surface sites on AuNPs surround those of the low-index surfaces since the NP surfaces are irregular and disordered.

To provide a quantitative measure of performance, we used the straight line connecting Au(111) to Au(110), which is  $\Delta E_{\rm CO} = 1.4423.\Delta E_{\rm HOCO} - 0.9194$ . By moving this line upward and downward, we get a family of lines with the same slope (1.4423) but different intercepts:  $\Delta E_{\rm CO} = 1.4423\Delta E_{\rm HOCO} + a$ . For each single surface site, with the neural network predicted  $\Delta E_{\rm CO}$  and  $\Delta E_{\rm HOCO}$ , the a-value can be calculated using  $a = \Delta E_{\rm CO} - 1.4423\Delta E_{\rm HOCO}$ . According to the reaction mechanism of CO2RR on Cu(100) surface, <sup>32</sup> the active sites we are looking for should have low  $\Delta E_{\rm HOCO}$ , as this is the potential-

Journal of the American Chemical Society



**Figure 4.** Active sites identification for AuNPs surfaces. We ranked the activity of all 11 537 surface sites on the basis of their *a*-values and identified the top 300 sites. We classified these 300 sites into seven groups, as marked in the figure. The number of sites in each group is indicated in brackets next to the marker. A representative structure for each of the seven active groups is shown. The center atom is pink, while atoms at the same layer are white. Atoms in the layer below white atoms are gold, while atoms one layer above the center atom are cyan. Atoms above cyan atoms are gray, and twin boundaries are shown with dashed lines.

determining step (PDS), lower  $\Delta E_{\rm HOCO}$  indicates higher CO2RR; high  $\Delta E_{\rm CO}$ , as higher  $\Delta E_{\rm CO}$  indicates that more sites will be released after the reaction. Thus, we expect that the a-value (defined by  $\Delta E_{\rm CO}$  and  $\Delta E_{\rm HOCO}$ ) will indicate the overall performance of a given surface site. Larger a-value means higher CO2RR while smaller a-value indicates poor performance. The top 300 sites ranked by their a-value are marked in the Figure 4. We found that the top 300 sites (out of 11 537) can be classified into seven groups, each with distinctive characteristics. We show in Figure 4 one representative structure of each group (for more structures see Figures S9–S15). We could not find a simple quantitative definition of each group, but we summarize the structure of each group as follows:

- Step111: Center atom has Au(111) features but with steps around the center atom.
- **StepUnder111**: Center atom has under-coordinated Au(111) features but with steps around the center atom.
- Step110: Center atom has 110 features but with steps around the center atom.
- Step311: Center atom has 311 features but with steps around the center atom.
- StepTB: Center atom has twin boundary features but with steps around the center atom.
- StepUnderTB: Center atom has the under-coordinated twin boundary feature but with steps around the center
- SurfaceDefect: Au(111) surface but with one or two missing atoms around the center atom.

For better comparison, we randomly selected 300 sites from whole surface with the identification results showing in Figure S6. The comparison between the top 300 sites (Figure 4) and randomly selected 300 sites is shown in Table S3. Among 300 randomly selected sites, the majority do not belong to the seven active groups as defined above. Thus, only the seven active groups are concentrated above and around the straight line with *a*-value = 0.9194.

2.3.2. Dealloyed Gold Surface Modeling and Active Sites Identification. Compared to AuNPs and Au foils, dealloyed Au surfaces have been shown experimentally to have dramatically improved performance in reducing CO<sub>2</sub> to CO. Thus, starting with a Au<sub>3</sub>Fe NP and exposing to electrochemical conditions suitable for CO<sub>2</sub> reduction, Wang, Goddard, and co-workers

showed that a CSNP is formed in which Fe is depleted from the top layers. Remarkably this  $Au_3Fe$  CSNP showed a 100-fold increase in mass activity compared with AuNP. Moreover it leads to >80% FE at -0.2 V, which is 500 meV lower than for Au foils!<sup>27</sup> Using QM studies on the (111) surface of  $Au_3M$  for 20 different M, Fe had been selected as best. But these calculations did not suggest such dramatically improved performance. It must be that the nature of this dealloyed surface somehow dramatically improves the reduction of  $CO_2$  to  $CO_2$ , but the type of site that can do this remains completely mysterious.

However, similar dramatic improvements from dealloying have been observed previously in PtNi alloy catalysts for ORR. Debe and co-workers discovered that Ni<sub>7</sub>Pt<sub>3</sub> NP leads to 4 times improved activity over Pt, <sup>26</sup> but their XPS studies could not find that any Ni remained, suggesting that it was fully dealloyed. Fortunelli and Goddard followed up using ReaxFF to show that the dealloyed 5 nm Ni<sub>7</sub>Pt<sub>3</sub> NP leads to a porous surface with a preponderance of low coordination sites. <sup>42</sup> They found that the surface area was doubled over a NP (in agreement with experiment) and that one important step in the reaction, O\* + H<sub>2</sub>O\*  $\rightarrow$  2OH\*, is dramatically accelerated on these sites. Later Huang, Duan, Goddard, and co-workers showed that 5 nm Ni<sub>7</sub>Pt<sub>3</sub> NWs after dealloying under ORR conditions led to 2 nm pure Pt NWs with ORR performance 50 times better than that of Pt!<sup>28</sup> The sites responsible for this dramatic improvement have not yet been identified.

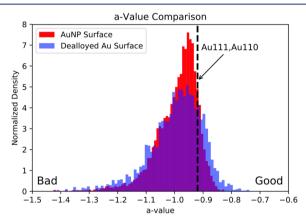
To search for sites on the dealloyed AuNP surfaces responsible for the dramatic improvement in activity, we modeled the dealloyed Au surface by cutting a 10 nm Au sphere from the single crystal, randomly removing 25% of the surface Au, and then allowing the structure to relax using ReaxFF. A more detailed description of the model and structure equilibration is included in the Supporting Information, section S4. We assume in our calculations the extreme hypothesis that the catalytic activity is not affected by the Fe in the core, nor by electronic effects or strain effect caused by lattice mismatch. This assumption may be valid since the Fe may be too far from the surface in the experiments. Thus, we leached out Au just at the top layer of the system. We show below that this simplifying assumption may be sufficient because we do predict dramatically improved CO2RR on the basis this hypothesis. If follow-up studies using QM on the structures we identify here turn out not to predict

the dramatic 500 mV decrease in the onset potential for >80% FE, we will consider dealloying multiple layers using ReaxFF to relax the structures. If this does not explain the improvements, then we will add in <sup>1</sup>/<sub>2</sub>Fe in the core. We expect that this sequence of studies combining QM, ReaxFF, and machine learning will eventually identify the mechanism. We show below that this simple model provides insights and reasonable hypotheses to explain the dramatically improved performance.

Using the same methodology as in section 2.3.1, we predicted the performance of this dealloyed Au surface using our neural networks. To have a more quantitative comparison, we define the active ratio (AR) as follows:

$$AR = \frac{\text{number of sites with } a > 0.9194}{\text{total number of surface sites}}$$

where 0.9194 is the *a*-value of Au(111) and Au(110). Figure 5 shows a normalized distribution of *a*-value comparisons, where



**Figure 5.** Normalized *a*-value distribution. The total area of each histogram is normalized to 1. As defined in the text, AR (active ratio) is the area to the right of the dashed line for each plot. Both AuNPs and dealloyed surface have a number of active sites with *a*-value larger than 0.9194, which is the *a*-value of Au(111) and Au(110). The dealloyed surface shows even better performance, with AR = 0.1307 compared to the AuNPs with AR = 0.0836.

the dashed line is the a-value for the low-index Au surfaces, Au(111) and Au(110). The red histogram is the distribution for AuNPs, while the blue histogram is the distribution for dealloyed Au surfaces. AR is then the area lying to the right of the dashed line, since the total area of each plot is normalized to 1. Comparing to low-index Au surfaces (dashed line), AuNPs and dealloyed Au surfaces both have increased numbers of active sites, with AR = 0.0836 for AuNPs and AR = 0.1307 for the dealloyed surface. This is consistent with experimental studies showing that the dealloyed surface exhibits much better performance than AuNPs, which in turn are better than Au foils.

Active site identification for the dealloyed Au surface is shown in Figure 6. Among the top 100 sites out of 3095 ranked by *a*-value,

- 43 of them arise from the StepUnder111 group
- 23 of them are from the SurfaceDefect group

which we consider as the most promising. There are also 23 sites from Step111 group and small numbers of sites from the Step110 and Step311 groups, but we consider them as less likely contributors to the dramatically improved performance, partly because these kinds of sites may already be present

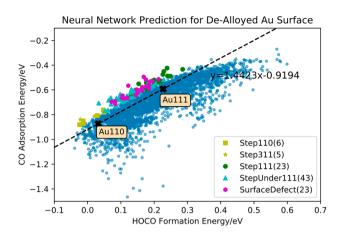


Figure 6. Active sites identification for dealloyed Au surfaces. The top 100 sites out of 3095 are marked with the number in each group shown in brackets next to the marker. As Step110, Step311, and Step111 may already be present before dealloying, we expect that the creation of increased StepUnder111 and SurfaceDefect sites by the dealloying process is likely the main contributor to the dramatically improved performance of the dealloyed Au<sub>3</sub>Fe CSNP.

before dealloying. We expect dealloying will naturally create such structure defects as under-coordinated sites. Therefore, we expect the increased fraction of **SurfaceDefect** and **StepUnder111** groups to be responsible for the improved performance.

In this example, we illustrate how to apply a machine learning model to analyze and understand extremely irregular and disordered systems. This leads to new structures that can be tested with realistic QM-based studies to validate the predictions; the ultimate goal is to validate the success for our ML-QM-ReaxFF strategy. Here we eschew the exact modeling of the dealloying process, which requires complicated and expensive calculations. To further verify neural network prediction results, we randomly selected five sites from each of the seven groups and carried out DFT cluster calculations (no solvation). Figures S7 and S8 show that our predicted energy values for both  $\Delta E_{\rm CO}$  and for  $\Delta E_{\rm HOCO}$  are mostly within the RMSE of the neural network predictions (0.0521 eV for  $\Delta E_{\rm CO}$  and 0.0614 eV for  $\Delta E_{\rm HOCO}$ ). This validates the accuracy of our ML model. It also validates that these seven groups are those sites with higher CO2RR.

2.3.3. a-Value Mapping and Catalytic Activity Visualization. One important application of our machine learning model is to evaluate the performance of each single site even on a highly disordered and irregular surface. The catalytic activity can be visualized by mapping the a-value back on the particle. As shown in Figure 7, all the sites are colored on the basis of their predicted a-value. The red sites are poor sites with small a-values, and the blue sites are active sites showing larger a-values. Three common surfaces of Au(111), Au(110), and Au(100) are marked in Figure 7a, showing that Au(111) and Au(110) with the color blue have better performance at CO2RR than Au(100) with the color green, which is consistent with earlier experimental and computational studies.<sup>23,41</sup> Two main contributors for the dramatically improved performance of dealloyed Au surfaces, SurfaceDefect and StepUnder111 sites, are also marked in Figure 7b, colored blue to indicate they have high performance. Our avalue mapping provides a powerful tool for catalytic activity visualization.

Journal of the American Chemical Society

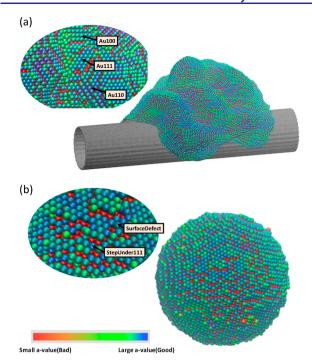


Figure 7. a-value mapping and catalytic activity visualization. Each single site is given an a-value based on neural network prediction. These a-values are then mapped back on the particle to visualize the catalytic activity of the whole surface. As indicated in the color bar, the red sites are inactive sites, and the blue sites are active ones. (a) Catalytic visualization for AuNPs. Au(111), Au(110), and Au(100) are marked. (b) Catalytic visualization for dealloyed Au surface. SurfaceDefect and StepUnder111 sites are shown in blue, indicating they have high performance at CO2RR.

We can also map descriptors from experiment, computation, or theory onto a companion diagram in order to provide guidance toward physical descriptors to design high-performance electrocatalysts for applications in electrocatalysis.

# 3. CONCLUSION

In summary, we developed two neural network based machine learning models for accurate prediction of CO adsorption energy and HOCO formation energy on extremely irregular and disordered Au surfaces, showing the accuracy of 0.05 eV needed for catalyst development. Applications of the model to AuNPs and dealloyed Au surfaces allow us to identify the feature of active sites responsible for the dramatically improved CO2RR performance of such disordered and irregular surfaces. We also develop an a-value mapping methodology to construct and visualize the catalytic activity of whole surface. We expect that this can help guide the design of high-performance electrocatalysts for CO2RR to enable storing of solargenerated energy and for CO<sub>2</sub>-free transportation. This provides an example of how to combine multiscale simulations, QM, and artificial intelligence to drive the discovery of new catalysts for clean energy conversion.

# ASSOCIATED CONTENT

# **S** Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.9b04956.

Simulation details, machine learning model description and training, more structures of active groups, DFT

verification of active groups, and additional calculations, including Figures S1-S15 and Tables S1-S3 (PDF)

#### AUTHOR INFORMATION

# **Corresponding Author**

\*wag@caltech.edu

#### ORCID ®

Yalu Chen: 0000-0002-0589-845X Yufeng Huang: 0000-0002-0373-2210

William A. Goddard, III: 0000-0003-0097-5716

The authors declare no competing financial interest.

# ACKNOWLEDGMENTS

This work was supported by the Joint Center for Artificial Photosynthesis, a DOE Energy Innovation Hub, supported through the Office of Science of the U.S. Department of Energy under Award Number DE-SC0004993. This work uses the computational resources of Caltech High Performance Computing Center (HPC).

#### REFERENCES

- (1) Goeppert, A.; Czaun, M.; Jones, J.-P.; Surya Prakash, G. K.; Olah, G. A. Recycling of Carbon Dioxide to Methanol and Derived Products - Closing the Loop. Chem. Soc. Rev. 2014, 43 (23), 7995-
- (2) Whipple, D. T.; Kenis, P. J. A. Prospects of CO2 Utilization via Direct Heterogeneous Electrochemical Reduction. J. Phys. Chem. Lett. 2010, 1 (24), 3451-3458.
- (3) Greeley, J.; Stephens, I. E. L.; Bondarenko, A. S.; Johansson, T. P.; Hansen, H. A.; Jaramillo, T. F.; Rossmeisl, J.; Chorkendorff, I.; Nørskov, J. K. Alloys of Platinum and Early Transition Metals as Oxygen Reduction Electrocatalysts. Nat. Chem. 2009, 1 (7), 552-
- (4) Hatsukade, T.; Kuhl, K. P.; Cave, E. R.; Abram, D. N.; Feaster, J. T.; Jongerius, A. L.; Hahn, C.; Jaramillo, T. F. Carbon Dioxide Electroreduction Using a Silver-Zinc Alloy. Energy Technol. 2017, 5 (6), 955-961.
- (5) Clark, E. L.; Hahn, C.; Jaramillo, T. F.; Bell, A. T. Electrochemical CO 2 Reduction over Compressively Strained CuAg Surface Alloys with Enhanced Multi-Carbon Oxygenate Selectivity. J. Am. Chem. Soc. 2017, 139 (44), 15848-15857.
- (6) Tang, M. H.; Hahn, C.; Klobuchar, A. J.; Ng, J. W. D.; Wellendorff, J.; Bligaard, T.; Jaramillo, T. F. Nickel-Silver Alloy Electrocatalysts for Hydrogen Evolution and Oxidation in an Alkaline Electrolyte. Phys. Chem. Chem. Phys. 2014, 16 (36), 19250.
- (7) Mendes, L. V. P.; Snider, J. L.; Fleischman, S. D.; Kibsgaard, J.; McEnaney, J. M.; Aranda, D. A. G.; Jaramillo, T. F. Polyol Synthesis of Cobalt-Copper Alloy Catalysts for Higher Alcohol Synthesis from Syngas. Catal. Lett. 2017, 147 (9), 2352-2359.
- (8) Hahn, C.; Abram, D. N.; Hansen, H. A.; Hatsukade, T.; Jackson, A.; Johnson, N. C.; Hellstern, T. R.; Kuhl, K. P.; Cave, E. R.; Feaster, J. T.; et al. Synthesis of Thin Film AuPd Alloys and Their Investigation for Electrocatalytic CO 2 Reduction. J. Mater. Chem. A 2015, 3 (40), 20185-20194.
- (9) Zheng, X.; Ji, Y.; Tang, J.; Wang, J.; Liu, B.; Steinrück, H.-G.; Lim, K.; Li, Y.; Toney, M. F.; Chan, K.; et al. Theory-Guided Sn/Cu Alloying for Efficient CO2 Electroreduction at Low Overpotentials. Nat. Catal 2019, 2 (1), 55-61.
- (10) Hansen, H. A.; Shi, C.; Lausche, A. C.; Peterson, A. A.; Nørskov, J. K. Bifunctional Alloys for the Electroreduction of CO 2 and CO. Phys. Chem. Chem. Phys. 2016, 18 (13), 9194-9201.
- (11) Qian, J.; Fortunelli, A.; Goddard, W. A. Effect of Co Doping on Mechanism and Kinetics of Ammonia Synthesis on Fe(1 1 1) Surface. J. Catal. 2019, 370, 364-371.

- (12) Cheng, M.-J.; Clark, E. L.; Pham, H. H.; Bell, A. T.; Head-Gordon, M. Quantum Mechanical Screening of Single-Atom Bimetallic Alloys for the Selective Reduction of CO 2 to C 1 Hydrocarbons. ACS Catal. 2016, 6 (11), 7769–7777.
- (13) Cheng, T.; Goddard, W. A.; An, Q.; Xiao, H.; Merinov, B.; Morozov, S. Mechanism and Kinetics of the Electrocatalytic Reaction Responsible for the High Cost of Hydrogen Fuel Cells. *Phys. Chem. Chem. Phys.* **2017**, *19* (4), 2666–2673.
- (14) Cheng, T.; Xiao, H.; Goddard, W. A. Full Atomistic Reaction Mechanism with Kinetics for CO Reduction on Cu(100) from Ab Initio Molecular Dynamics Free-Energy Calculations at 298 K. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (8), 1795–1800.
- (15) Cheng, T.; Xiao, H.; Goddard, W. A. Free-Energy Barriers and Reaction Mechanisms for the Electrochemical Reduction of CO on the Cu(100) Surface, Including Multiple Layers of Explicit Solvent at PH 0. *J. Phys. Chem. Lett.* **2015**, *6* (23), 4767–4773.
- (16) Xiao, H.; Cheng, T.; Goddard, W. A.; Sundararaman, R. Mechanistic Explanation of the PH Dependence and Onset Potentials for Hydrocarbon Products from Electrochemical Reduction of CO on Cu (111). *J. Am. Chem. Soc.* **2016**, *138* (2), 483–486.
- (17) Xiao, H.; Cheng, T.; Goddard, W. A. Atomistic Mechanisms Underlying Selectivities in C 1 and C 2 Products from Electrochemical Reduction of CO on Cu(111). *J. Am. Chem. Soc.* **2017**, 139 (1), 130–136.
- (18) Shin, H.; Xiao, H.; Goddard, W. A. In Silico Discovery of New Dopants for Fe-Doped Ni Oxyhydroxide (Ni 1- x Fe x OOH) Catalysts for Oxygen Evolution Reaction. *J. Am. Chem. Soc.* **2018**, *140* (22), *6745–6748*.
- (19) Ping, Y.; Nielsen, R. J.; Goddard, W. A. The Reaction Mechanism with Free Energy Barriers at Constant Potentials for the Oxygen Evolution Reaction at the IrO 2 (110) Surface. *J. Am. Chem. Soc.* **2017**, 139 (1), 149–155.
- (20) Kim, D.; Kley, C. S.; Li, Y.; Yang, P. Copper Nanoparticle Ensembles for Selective Electroreduction of CO 2 to C 2 -C 3 Products. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (40), 10560–10565.
- (21) Chen, Y.; Li, C. W.; Kanan, M. W. Aqueous CO2 Reduction at Very Low Overpotential on Oxide-Derived Au Nanoparticles. *J. Am. Chem. Soc.* **2012**, *134* (49), 19969–19972.
- (22) Li, C. W.; Ciston, J.; Kanan, M. W. Electroreduction of Carbon Monoxide to Liquid Fuel on Oxide-Derived Nanocrystalline Copper. *Nature* **2014**, 508 (7497), 504–507.
- (23) Back, S.; Yeom, M. S.; Jung, Y. Active Sites of Au and Ag Nanoparticle Catalysts for CO2 Electroreduction to CO. ACS Catal. **2015**, 5 (9), 5089–5096.
- (24) Zhu, W.; Michalsky, R.; Metin, Ö.; Lv, H.; Guo, S.; Wright, C. J.; Sun, X.; Peterson, A. A.; Sun, S. Monodisperse Au Nanoparticles for Selective Electrocatalytic Reduction of CO 2 to CO. *J. Am. Chem. Soc.* 2013, 135 (45), 16833–16836.
- (25) Zhao, S.; Jin, R.; Jin, R. Opportunities and Challenges in CO2 Reduction by Gold- and Silver-Based Electrocatalysts: From Bulk Metals to Nanoparticles and Atomically Precise Nanoclusters. ACS Energy Letters 2018, 3 (2), 452–462.
- (26) Stevens, D. A.; Mehrotra, R.; Sanderson, R. J.; Vernstrom, G. D.; Atanasoski, R. T.; Debe, M. K.; Dahn, J. R. Dissolution of Ni from High Ni Content Pt1-xNix Alloys. *J. Electrochem. Soc.* **2011**, *158* (8), B905.
- (27) Sun, K.; Cheng, T.; Wu, L.; Hu, Y.; Zhou, J.; Maclennan, A.; Jiang, Z.; Gao, Y.; Goddard, W. A.; Wang, Z. Ultrahigh Mass Activity for Carbon Dioxide Reduction Enabled by Gold-Iron Core-Shell Nanoparticles. *J. Am. Chem. Soc.* **2017**, *139* (44), 15608–15611.
- (28) Li, M.; Zhao, Z.; Cheng, T.; Fortunelli, A.; Chen, C.-Y.; Yu, R.; Zhang, Q.; Gu, L.; Merinov, B. V.; Lin, Z.; et al. Ultrafine Jagged Platinum Nanowires Enable Ultrahigh Mass Activity for the Oxygen Reduction Reaction. *Science* **2016**, *354* (6318), 1414–1419.
- (29) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105* (41), 9396–9409.
- (30) Cheng, T.; Huang, Y.; Xiao, H.; Goddard, W. A. Predicted Structures of the Active Sites Responsible for the Improved Reduction

- of Carbon Dioxide by Gold Nanoparticles. J. Phys. Chem. Lett. 2017, 8 (14), 3317–3320.
- (31) Cheng, T.; Xiao, H.; Goddard, W. A. Nature of the Active Sites for CO Reduction on Copper Nanoparticles; Suggestions for Optimizing Performance. *J. Am. Chem. Soc.* **2017**, 139 (34), 11642–11645.
- (32) Cheng, T.; Xiao, H.; Goddard, W. A. Reaction Mechanisms for the Electrochemical Reduction of CO 2 to CO and Formate on the Cu(100) Surface at 298 K from Quantum Mechanics Free Energy Calculations with Explicit Water. J. Am. Chem. Soc. 2016, 138 (42), 13802–13805.
- (33) Nie, X.; Esopi, M. R.; Janik, M. J.; Asthagiri, A. Selectivity of CO 2 Reduction on Copper Electrodes: The Role of the Kinetics of Elementary Steps. *Angew. Chem., Int. Ed.* **2013**, *52* (9), 2459–2462.
- (34) Luo, W.; Nie, X.; Janik, M. J.; Asthagiri, A. Facet Dependence of CO2 Reduction Paths on Cu Electrodes. *ACS Catal.* **2016**, *6* (1), 219–229.
- (35) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, 98 (14), 146401.
- (36) 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 Letters 2018, 3, 2983–2988.
- (37) Huang, Y.; Kang, J.; Goddard, W. A.; Wang, L.-W. Density Functional Theory Based Neural Network Force Fields from Energy Decompositions. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2019**, 99 (6), 064103.
- (38) Panapitiya, G.; Avendaño-Franco, G.; Ren, P.; Wen, X.; Li, Y.; Lewis, J. P. Machine-Learning Prediction of CO Adsorption in Thiolated, Ag-Alloyed Au Nanoclusters. *J. Am. Chem. Soc.* **2018**, *140* (50), 17508–17514.
- (39) Chowdhury, A. J.; Yang, W.; Walker, E.; Mamun, O.; Heyden, A.; Terejanu, G. A. Prediction of Adsorption Energies for Chemical Species on Metal Catalyst Surfaces Using Machine Learning. *J. Phys. Chem. C* 2018, 122 (49), 28142–28150.
- (40) Peterson, A. A.; Nørskov, J. K. Activity Descriptors for CO 2 Electroreduction to Methane on Transition-Metal Catalysts. *J. Phys. Chem. Lett.* **2012**, 3 (2), 251–258.
- (41) Zhu, W.; Zhang, Y.-J.; Zhang, H.; Lv, H.; Li, Q.; Michalsky, R.; Peterson, A. A.; Sun, S. Active and Selective Conversion of CO2 to CO on Ultrathin Au Nanowires. *J. Am. Chem. Soc.* **2014**, *136* (46), 16132–16135.
- (42) Fortunelli, A.; Goddard III, W. A.; Sementa, L.; Barcaro, G.; Negreiros, F. R.; Jaramillo-Botero, A. The Atomistic Origin of the Extraordinary Oxygen Reduction Activity of Pt3Ni7 Fuel Cell Catalysts. *Chemical Science* **2015**, *6* (7), 3915–3925.