

Research Article

Open Access



Machine-learning prediction of facet-dependent CO coverage on Cu electrocatalysts

Shanglin Wu¹, Shisheng Zheng^{2,*}, Wentao Zhang¹, Mingzheng Zhang¹, Shunning Li^{1,*}, Feng Pan^{1,*} 

¹School of Advanced Materials, Peking University, Shenzhen Graduate School, Shenzhen 518055, Guangdong, China.

²College of Energy, Xiamen University, Xiamen 361000, Fujian, China.

*Correspondence to: Prof. Shunning Li, Prof. Feng Pan, School of Advanced Materials, Peking University, Shenzhen Graduate School, No. 2199 Lishui Road, Shenzhen 518055, Guangdong, China. E-mail: lisen@pku.edu.cn, panfeng@pkusz.edu.cn; Prof. Shisheng Zheng, College of Energy, Xiamen University, No. 4221, Xiang'an South Road, Xiamen 361000, Fujian, China. E-mail: zhengss@pku.edu.cn

How to cite this article: Wu, S.; Zheng, S.; Zhang, W.; Zhang, M.; Li, S.; Pan, F. Machine-learning prediction of facet-dependent CO coverage on Cu electrocatalysts. *J. Mater. Inf.* 2025, 5, 14. <https://dx.doi.org/10.20517/jmi.2024.77>

Received: 22 Nov 2024 **First Decision:** 19 Dec 2024 **Revised:** 16 Jan 2025 **Accepted:** 20 Jan 2025 **Published:** 26 Feb 2025

Academic Editors: Lei Shen, Ming Hu **Copy Editor:** Ting-Ting Hu **Production Editor:** Ting-Ting Hu

Abstract

Copper-based electrocatalysts, which hold great promise in selectively reducing CO₂ into multicarbon products, have attracted significant recent interest, both experimentally and theoretically. While many studies have suggested a strong dependence of catalytic selectivity on the concentration of the CO reaction intermediate on the Cu surface, it remains challenging for a direct experimental probe of the CO coverage. This necessitates a reliable computational method that can accurately establish the theoretical coverage-dependent phase diagram of CO adsorbates on the catalyst. Here we propose a scheme composed of density functional theory calculations, machine-learning force fields and graph neural networks as a solution. This method enables a fast screening of 7 million adsorption configurations based on a small set of density functional theory data, with a balance between accuracy and efficiency tuned by the combinatorial use of machine-learning force field and graph neural network models. We have investigated eight different Cu facets and discovered that the high-index facets such as (310), (210) and (322) exhibit a much higher CO coverage than the low-index counterparts such as (111), leading to an increased opportunity for C–C coupling for the former. Our results can provide a new perspective for the understanding of the fundamental role of CO coverage on the Cu surface for electrochemical CO₂ reduction.

Keywords: Machine-learning force fields, density functional theory, graph neural networks, coverage effect, electrochemical CO₂ reduction



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



INTRODUCTION

In the realms of surface science and catalysis, the interactions between surfaces and adsorbates are foundational for understanding catalytic processes and crucial for catalyst design and discovery^[1-6]. With the rapid advancement of computational and theoretical chemistry, particularly the extensive application of density functional theory (DFT), researchers can now calculate the adsorption energy of adsorbates on surfaces with unprecedented accuracy, playing an indispensable role in catalyzing the design and innovation process^[6-12].

The study of lateral interactions among adsorbates and their impact on catalytic activity, selectivity, and surface stability, such as in carbon dioxide reduction reactions (CO₂RR)^[13-17], nitrogen oxide reduction reactions (NO_xRR)^[18-20], and the Fischer-Tropsch synthesis^[21-23], has gained increasing attention. These interactions, especially those modulated by varying adsorbate coverage, are vital for precisely controlling the catalytic reaction process, making a deep understanding of lateral adsorbate interactions and coverage effects crucial for optimizing catalyst design and enhancing performance. However, exhaustively calculating coverage-dependent adsorption energies using DFT alone often proves impractical due to the combinatorial growth of adsorption configuration spaces with coverage and site types, and the substantial computational cost of a comprehensive analysis. Various methods have been proposed to address this, including cluster expansion^[24], multi-order lateral interaction models^[14], graph theory^[18,25,26], and machine-learning approaches^[27]. The combination of graph theory and machine learning has been regarded as one of the most effective means to analyze coverage-dependent adsorption energies: graph theory tools for automating the enumeration of vast adsorption configurations, and machine learning models for predicting adsorption energies across the entire configuration space based on a limited DFT dataset^[18,26]. However, graph-based enumeration algorithms encounter significant computational bottlenecks at the stage of isomorphism comparison of configurations, as graph isomorphism comparison is an extremely time-consuming Non-deterministic Polynomial (NP) problem, exponentially growing with the number of atoms in adsorption configurations^[28]. Since the differences between adsorption configurations mainly depend on the occupation of different adsorbate sites, symmetry-based methods often offer a more efficient evaluation of adsorption configurations with multiple site occupations. Machine-learning approaches, especially deep learning methods based on neural networks (NNs), have been primarily divided into two strategies for accelerating the prediction of adsorption energies in vast configuration spaces: (1) direct prediction of stable state adsorption energies from initial configuration guesses using NNs; and (2) acceleration of the geometry optimization process using machine-learning force fields (MLFF)^[29,30]. While NN methods are highly efficient in prediction with negligible computational cost, their accuracy depends not only on the model quality but also on the size of the training dataset^[31]. In contrast, MLFF methods require significantly less DFT computational data for training and can obtain both stable configurations and more accurate adsorption energy predictions. However, the optimization time needed for MLFF methods far exceeds the prediction time of NN methods, though it still represents a substantial saving in computational cost compared to direct DFT calculations. Each strategy has its strengths, and they contribute from different perspectives to the exploration of coverage-dependent adsorption energies. Yet, all these methods can only consider the optimization of a limited number of configurations, facing insurmountable computational burdens when confronted with potentially hundreds of thousands or millions of configurations.

In this work, we have developed a “structure enumeration + MLFF + NNs” approach for efficiently exploring adsorbate-adsorbate interactions, enabling the rapid exploration of nearly ten million configurations on various copper facets with different CO adsorption coverage. By rapidly enumerating and deduplicating adsorption configurations through a “geometry + graph theory + symmetry” approach, we generated an approximately 7 million configuration guess space. Then, using a designated sampling

method, we selected a very small set of configurations for DFT structural optimization to obtain optimization trajectories for training a MLFF based on deep potential molecular dynamics (DPMD). Subsequently, the MLFF was used to perform structural optimizations on an expanded sampling space (~186,000) to obtain corresponding stable adsorption energies. Finally, we trained a graph embedding network model (GEN) using the configuration-energy data, which considered non-bonded adsorbate interactions in feature construction and efficiently predicted energies across the entire target configuration space. Applied to the Cu-CO system, our method achieved results qualitatively consistent with experiments at three orders of magnitude lower computational cost than pure DFT calculations: the adsorption strength of CO on Cu surfaces exhibited a minimal change in energy at first, followed by a significant increase with coverage; high-index Cu surfaces often exhibited higher catalytic activity due to more low-coordination Cu sites. These findings undoubtedly demonstrate the effectiveness and efficiency of our method and its power in exploring vast configuration spaces.

MATERIALS AND METHODS

At high coverage, the configurations of adsorption not only experience an explosive increase in number due to the enumerated surfaces and the geometric structures and binding modes of the adsorbates but also become almost unpredictable due to the complex interactions between the adsorbates. This signifies that the enumeration methods relying solely on expert experience fail under these circumstances^[25]. By integrating expert knowledge with deep learning technologies, a programmable scalable agent model can provide interpretable and reliable analysis and predictions for the vast configuration space.

Workflow

Our study focuses on the adsorption configurations of CO on eight different Cu surfaces at varying coverage levels. As illustrated in [Figure 1A](#), taking the (100) surface as an example, as the CO coverage increases, the distance between the adsorbed CO molecules gradually decreases, along with an increase in the interaction between the adsorbates. The number of configurations shows a trend of initially increasing and then decreasing [[Supplementary Table 1](#)]. Ultimately, nearly 7 million adsorption configurations are generated for the eight Cu surfaces, representing an extremely large configuration space [[Supplementary Table 2](#)]. We will introduce the detailed enumeration process in the following sections.

We propose a simple and efficient framework capable of rapidly and accurately predicting the CO adsorption energies of all stable configurations. The workflow is illustrated in [Figure 1B](#), which depicts our comprehensive computational exploration of the configuration space. Initially, the entire configuration space is sampled randomly twice to obtain the first and second sampling spaces, respectively, with both sampling steps covering all surface coverage levels. Subsequently, configurations from the more concise second sampling space undergo DFT structural optimization to obtain the CO adsorption energies of stable configurations, along with the trajectory of configuration optimization and the corresponding energies. Using these trajectories and energies as a dataset, a MLFF based on the DPMD deep potential architecture is trained, which effectively fits the interaction between adsorbates at different coverages^[32]. The fitted force field is then used to optimize the structures of the larger first sampling space to obtain the CO adsorption energies of stable configurations. The resulting configuration-adsorption energy data can train a well-performing GEN, capable of accurately predicting the stable CO energies of approximately 7 million adsorption configurations in the target configuration space, despite using very simple feature combinations.

Compared to the total configurational space, the computational framework requires a significantly lower volume of initial data from DFT calculations, differing by three orders of magnitude, even though DFT calculations are generally considered to be highly resource-intensive. By introducing a machine-learned

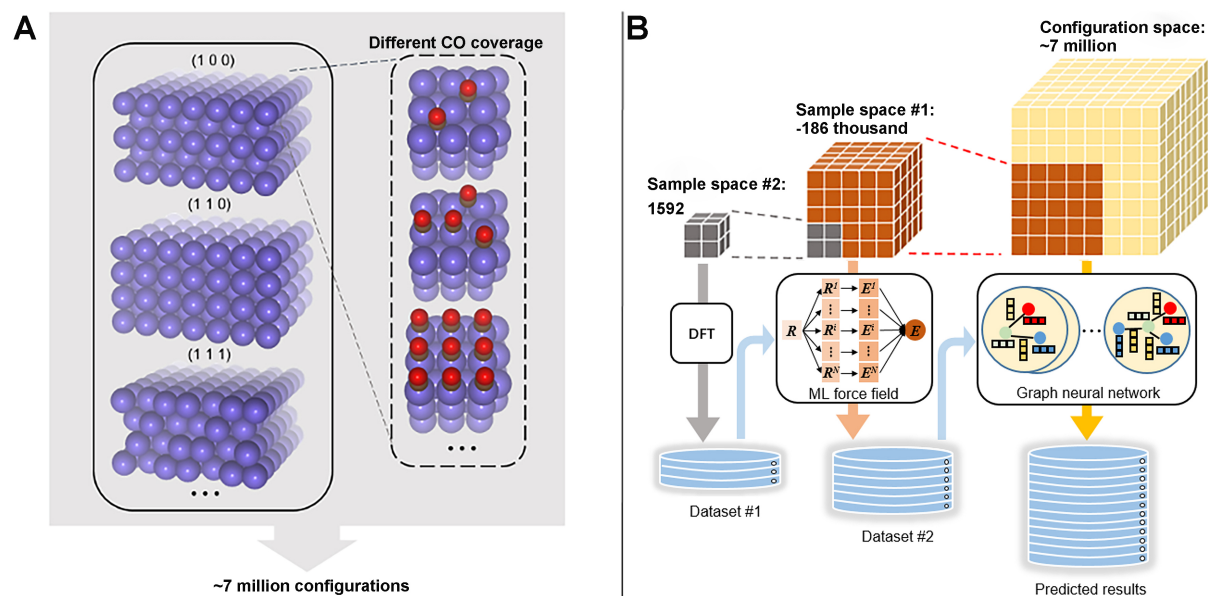


Figure 1. Configuration Space and Workflow for CO Adsorption Energy Prediction on Cu Surfaces. (A) Schematic representation of CO adsorption configurations on Cu(100), Cu(110), and Cu(111) surfaces illustrating different CO coverages. The configurations exhibit varying distances between adsorbed CO molecules in relation to the coverage density, culminating in a broad spectrum of nearly 7 million configurations across eight Cu surfaces, as indicated by the expansive configuration space; (B) Depiction of the workflow: starting with a random selection of 186,000 potential configurations (Sample space 1), it narrows down to 1,592 configurations (Sample space 2) for DFT optimization. These optimized configurations train a DPMD-based ML force field, which is then used to predict adsorption energies for the initial sample space, allowing the graph embedding network to estimate stable-state energies for the extensive configuration space. DFT: Density functional theory; DPMD: deep potential molecular dynamics; ML: machine learning.

force field (MLFF) model with DFT accuracy, we have substantially enriched the training dataset for the adsorption energy prediction model. This approach cleverly circumvents the costly active learning process while enabling the assessment and exploration of the complete, complex configurational space in a cost-measurable manner - a capability not present in previous studies^[33,34]. Moreover, this scalable framework for high-coverage configurational exploration can be flexibly defined and upgraded according to the user's needs.

Independent adsorption configuration enumeration

To acquire the global stable structures of adsorption configurations, researchers typically enumerate and construct sets of configuration guesses that closely approximate local stable structures based on prior knowledge before DFT geometric optimization^[35,36]. These initial guesses often share similar or identical atomic bonding relationships with their corresponding local stable structures. After geometric optimization, the local stable structures with the lowest energy are usually considered the global stable structures of the adsorption configuration; in other words, the global stable structures are a subset of the collection of local stable structures. However, this manual enumeration method for guessing adsorption configurations is not suitable for the high-coverage adsorption configuration systems in our study, due to the richness and combinatorial nature of adsorbable sites on the modeling surface^[25,35]. In this case, we must introduce an automated enumeration process for configuration guesses based on prior experience and conditional constraints. Since our research focuses on the impact of interactions between adsorbates on CO adsorption energy, the variability of adsorption sites is a key consideration in our enumeration of adsorption configuration guesses. Additionally, the global stable structures we focus on do not involve complex changes such as interface reconstruction or adsorbate desorption, at most only changes in the CO adsorption sites on the surface compared to the initial configuration guesses after geometric optimization,

although studies have shown that high coverage of adsorbates can trigger surface reconstruction and significantly change the surface's catalytic properties^[29,34,37]. For potential adsorbate desorption issues after geometric optimization, we can avoid them by introducing an empirical distance threshold in the initial guess modeling before optimization, effectively reducing the computational cost of unreasonable configuration guesses. In our studied systems, the differences in adsorption energy between adsorption configurations primarily depend on the adsorption sites, so enumerating the combination space of adsorption sites on the surface conveniently yields our target configuration guess space, which can be quickly achieved for a finite-sized slab.

The enumeration of adsorption configuration guess space is conducted on eight Cu surfaces, including (100), (110), (111), (210), (221), (310), (311), and (322), which are considered to be worth exploring by researchers^[38-42]. As shown in [Figure 2](#), the enumeration process is divided into three steps: surface site search, adsorption configuration enumeration, and deduplication of equivalent configuration guesses. In the site search part, we define the collection of Cu atoms on the surface that can bond with adsorbates and their surrounding environment as a site. Therefore, the types of sites depend on the size of the Cu atom collection and its local environment. Here, we abstract sites and their local environments into graphs using graph theory methods and determine the uniqueness of sites by judging whether the site graphs are isomorphic^[25]. Using graph theory methods, we conveniently identified 68 unique sites from the eight Cu surfaces and used a combination letter naming method to distinguish different site types, for example, the "Bb" site type, where "B" indicates the site has a coordination number of $n = 2$, and "b" indicates it is the second graph structure among sites with the same coordination number, and so on. This purely geometric definition of sites does not restrict the morphology of the sites, reducing the damage that biased understanding may cause in constructing a complete configuration guess space^[25,36]. Its generality and scalability make it convenient to apply to the construction of configuration guesses in other high coverage systems^[43-45]. The graph-theoretical site comparison method allows for a simple and rapid quantitative description of differences between sites.

After detecting all adsorption sites on the Cu surfaces, we used a method of filling CO molecules on the clean Cu surface with distance limitations to achieve the enumeration of configuration guesses, with the main requirements being: CO adsorbs monodentately on the Cu surface with only C contacting Cu atoms; the orientation of CO molecules is set to the vector sum of the direction vector from its coordinating Cu atom to the C atom; the filling distance limitation requires that the interatomic distance between different CO molecules must not be less than 2.3 Å. Consequently, we obtained approximately 44 million initial adsorption configuration guesses, with the number of configurations enumerated on different index faces increasing with CO coverage before decreasing [[Supplementary Table 1](#)]. Based on the site type naming method, we named the adsorption configuration according to the combination types of sites they belong to, for example, "Aa2Ab1Da1Dc2" site combination type, indicating 2, 1, 1, 2 CO molecules adsorb on "Aa", "Ab", "Da", "Dc" types of sites, respectively. Moreover, we defined a simplified site combination type, such as the simplified type for "Aa2Ab1Da1Dc2" being "AD". Subsequent configuration data sampling calculations will be based on these two types of site combinations. Notably, the previous enumeration process did not consider the intrinsic symmetry of different Cu surfaces; for example, the Cu(111) surface corresponds to the $p3m1$ plane group [[Supplementary Tables 3 and 4](#)]^[46-48]. The existence of such planar symmetry results in a large number of duplicate structures in the enumerated configuration guesses. As shown in [Supplementary Figure 1](#), two adsorption configurations A and B on Cu(111) with two CO molecules that do not correspond to the same atomic coordinates may actually be equivalent structures. Through symmetry operations, we obtained approximately 7 million independent initial adsorption configuration guesses from the 44 million, significantly narrowing the target configuration space range

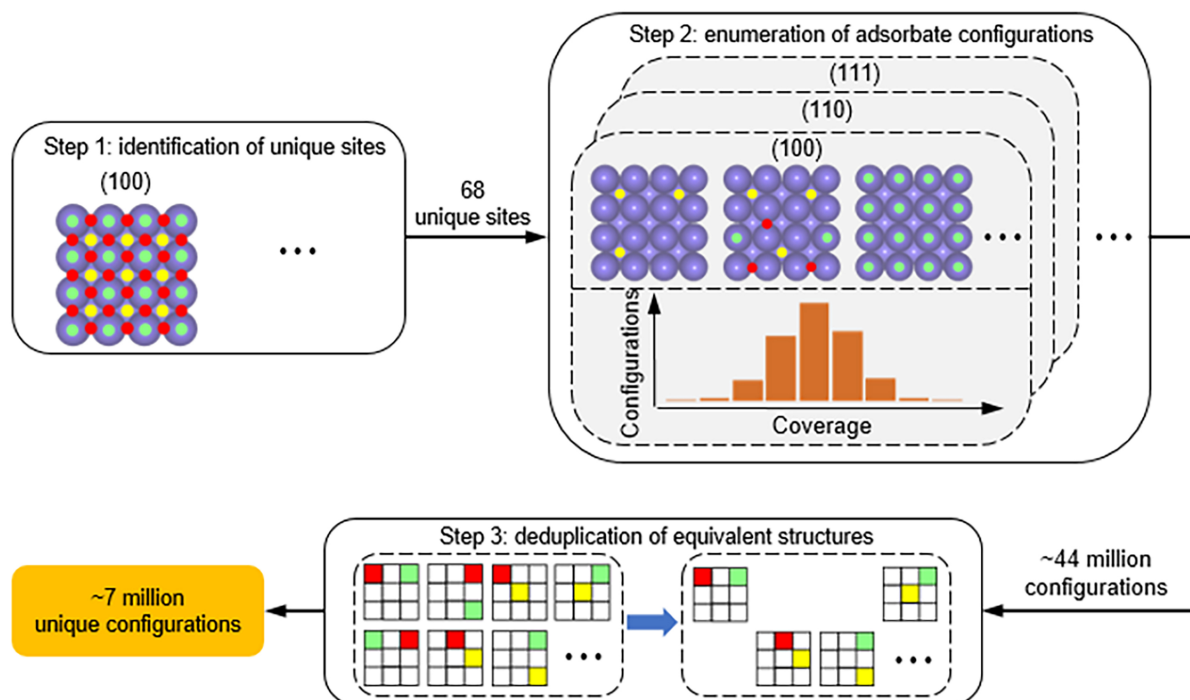


Figure 2. Enumeration of unique CO Adsorption Configurations. A three-step method for defining unique CO adsorption configurations on Cu surfaces. Initially, Step 1 identifies 68 distinct adsorption sites using graph theory to capture surface atom arrangements. Step 2 demonstrates the systematic filling of CO on these sites while adhering to spatial constraints, yielding approximately 44 million preliminary configurations. Step 3 applies symmetry operations to distill these down to around 7 million unique configurations, streamlining the dataset for further computational exploration.

[Supplementary Table 2]. Compared to the graph-theoretical deduplication algorithm, this method not only has a crushing advantage in comparison speed but also can theoretically overcome the limitations of the graph isomorphism algorithm in comparing periodic crystal structures^[25]. Overall, this enumeration process of the configuration space allows for a more comprehensive consideration of the adsorbate sequence space, thereby reducing errors associated with energetics-based configuration sampling^[25,29].

Machine-learning force field

Given the vast adsorption configuration space spanning eight different Cu surfaces, it is imperative to leverage artificial intelligence to navigate this extensive configuration space. Two research approaches are considered viable: The first involves constructing MLFFs through the rich trajectory data obtained during the first-principles geometric optimization of a small set of adsorption configurations, followed by using the trained force field model to optimize the remaining adsorption configurations^[30]. The second approach entails developing a deep learning model capable of directly predicting the steady-state adsorption energy from initial configuration guesses^[18,49]. Given that the construction of force fields can significantly utilize data from the structural optimization process, the initial DFT calculations required for the first strategy are considerably less. However, using the force field model to optimize the remaining configurations is also a time-consuming task, especially considering the configuration space volume is close to seven million. Additionally, selecting a small number of representative configurations for DFT calculations from a vast and unevenly distributed sample space, particularly among high-index surface configurations, poses a challenging problem.

Therefore, after considering both the predictive accuracy of MLFFs and the prediction speed of deep learning models, we designed an integrated solution that uses MLFF as a data augments for the adsorption energy prediction model. Initially, for low-index surfaces such as Cu(100), Cu(110), and Cu(111) depicted in Figure 1B, where the enumerated configuration count is low, we randomly select 20% of the unoptimized adsorption configurations for DFT calculations. For higher-index surfaces, we perform a primary sampling based on adsorption site combination types, followed by a secondary sampling based on simplified adsorption site combinations. The configurations obtained from secondary sampling undergo DFT structural optimization. This sampling process yielded a total of 1,592 structures. The resulting structure-energy pairs through structural optimization were obtained for training the MLFF [Supplementary Table 5]. Subsequently, we train the MLFF for the Cu-CO system with accuracy close to DFT calculations using the open-source and user-friendly deep learning architecture DPMD, which has been proven to have high simulation accuracy and efficiency in vast Si atomic systems^[32,50]. This force field model is then applied to optimize the remaining low-index surface configurations and those from the primary sampling. By expanding the training dataset with adsorption configurations using MLFFs, we can significantly address the issue of data scarcity faced when constructing deep learning models, thereby substantially improving model quality^[31].

Our trained force field exhibits high accuracy, with root mean square error (RMSE) values for energy and force being < 1 meV/atom [Figure 3A] and < 0.03 eV/Å [Figure 3B], respectively. The optimization of initial configuration guesses in the secondary sampling space using the well-fitted force field further assesses the force field quality: the steady-state adsorption configurations obtained from force field optimization almost perfectly match those obtained from DFT optimization [Figure 3C]. Moreover, analyzing the distribution of Cu-C bond lengths in the steady-state adsorption configuration set reveals that the distribution from MLFF optimization closely aligns with that from DFT optimization [Figure 3D], with an average Cu-C bond length discrepancy of < 0.01 Å. This demonstrates that the MLFF can achieve near-DFT calculation accuracy for adsorption configurations across eight Cu surfaces. The high-quality force field is attributed not only to the superior MLFF architecture but also to the effective sampling method - ensuring the sampled configuration guesses, i.e., the starting points for DFT geometric optimization, are as diverse as possible, facilitating a rich optimization trajectory that benefits model learning of more complex multi-body interactions. We used the MFL to optimize $\sim 186,000$ structures for the following graph neural network (GNN) model and the distribution of structures was presented in Supplementary Table 6. The specific training parameters of MLFF are shown in Supplementary Table 7.

Graph-based model for adsorption energy prediction

In recent years, GNN models have shone brightly in material design and performance prediction, leveraging graph embeddings of materials for various predictive tasks^[51-54]. One of their key advantages over other deep learning models lies in the natural suitability of graph structures for describing chemical and material structures, with nodes and edges representing atoms and their interactions, respectively. Hence, extracting high-quality graph data from materials is crucial. In our study, the interactions among CO molecules affect their adsorption energy on Cu surfaces, especially at high coverages. Previous research often overlooked non-bonding interactions beyond hydrogen bonds when constructing graph data, failing to accurately characterize adsorption structures^[18]. We propose an adsorption configuration graph data extraction method that effectively captures the impact of CO interactions on their adsorption energy, as shown in Figure 4A and Supplementary Figure 2. The first step involves searching for atoms in contact with a CO molecule on the Cu surface, considering it as the central CO, within the van der Waals radius to form a neighbor atom set, including first-order neighbor COs in contact. The second step continues the search for atoms in contact with these first-order neighbor COs and builds local subgraphs for multiple neighbor sets centered on CO. The third step merges the central CO and its neighboring COs' local subgraphs into a

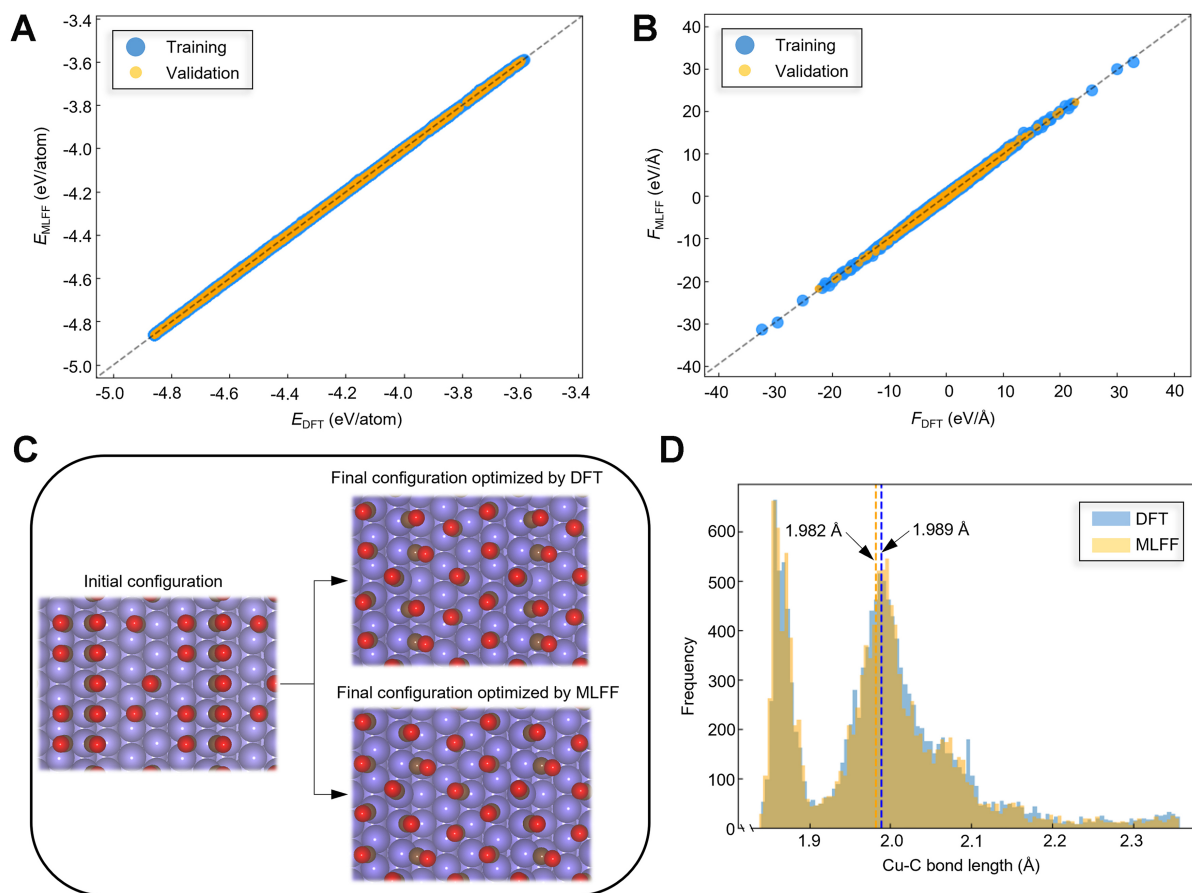


Figure 3. Evaluation of MLFF accuracy. (A) Scatter plot of the energy predictions, demonstrating the MLFF's high accuracy, with an RMSE for energy within 1 meV/atom, as validated against DFT calculations; (B) Corresponding force predictions plot, where the MLFF achieves an RMSE of less than 0.03 eV/Å, indicating the model's precision in capturing the forces in the Cu-CO system across a vast configuration space; (C) Comparing the initial CO adsorption configuration with those optimized by DFT and MLFF, showing the MLFF's ability to replicate DFT-level structural accuracy on Cu surfaces; (D) Histogram of Cu-C bond lengths from stable-state configurations, highlighting the close match between MLFF and DFT results, with bond length discrepancies averaging less than 0.01 Å. MLFF: Machine-learning force field; DFT: density functional theory; RMSE: root mean square error.

feature graph representing the adsorption environment of the central CO. Different interaction types in the feature graph are assigned distinct indicator vectors during encoding, including non-bonding interactions between CO molecules, akin to the encoding of hydrogen bonds in previous research^[26].

Despite using only a few very simple attribute features [Supplementary Table 8] and conventional training steps [Supplementary Table 9], the GNN model demonstrated excellent predictive accuracy and generalization ability across eight index surfaces at varying coverages [Supplementary Figure 3]. The model's superior performance stems partly from more accurate graph descriptors. Unlike models that ignored CO non-bonding interactions when constructing graph data, our model significantly improved performance [Supplementary Figure 4]. Additionally, it benefited from the richness of training graph data, thanks to the expansion of training data via MLFF, presenting an advantage over models trained solely on data from DFT calculations. The accuracy obtained by training solely on DFT data [Figure 4B] is significantly lower than that achieved through training on an expanded dataset (about 186,000) under the same model architecture [Figure 4C]. The latter exhibits improvements in accuracy by 16% and 28% in terms of the coefficient of determination (R^2) and root mean squared error (RMSE), respectively

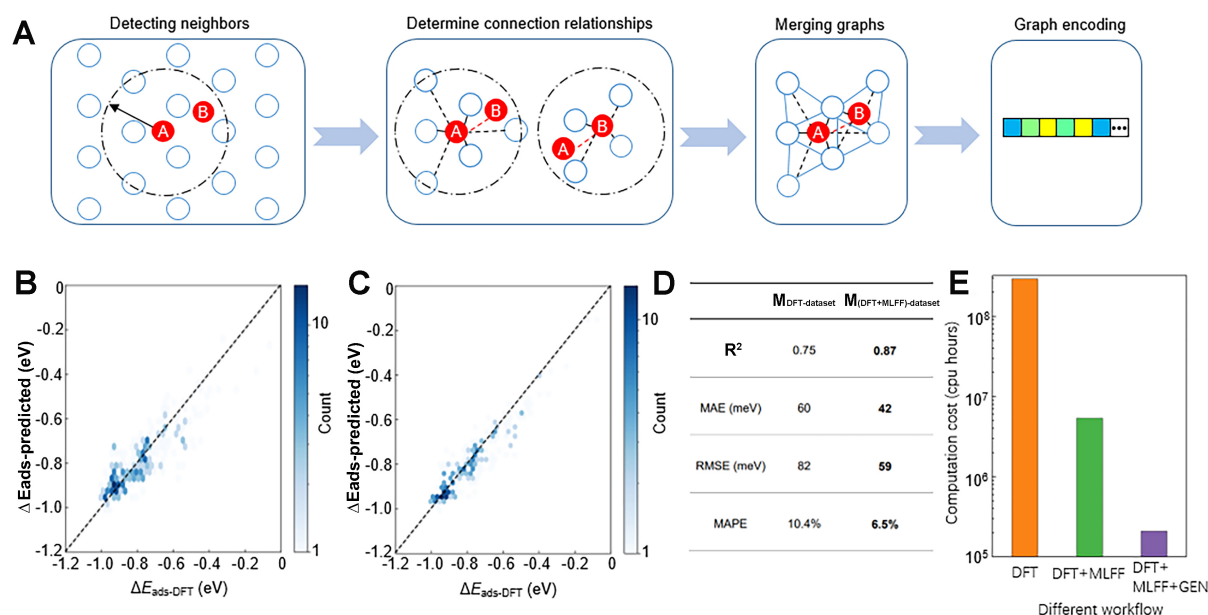


Figure 4. GNN model for CO Adsorption Prediction. (A) The graph data extraction method for CO adsorption configurations, detailing the steps from detecting neighboring atoms under van der Waals conditions to merging local subgraphs into a feature graph that accurately represents the adsorption environment of CO molecules on Cu surfaces. (D) The comparison of the GNN model's predictive performance between using only the DFT calculation dataset corresponding to (B) and using the DFT + MLFF calculation dataset corresponding to (C). (E) The computational efficiency gains of the DFT + MLFF + GNN workflow compared to traditional DFT and DFT + MLFF methods, showcasing a significant reduction in computational cost and time, thus enabling the study of vast adsorption configuration spaces with enhanced efficiency. DFT: Density functional theory; MLFF: machine-learning force field; GNN: graph neural network; RMSE: root mean square error; GEN: graph embedding network model; MAE: mean absolute error; MAPE: mean absolute percentage error.

[Figure 4D]. Moreover, compared to direct DFT calculations or combined DFT + MLFF approaches for exploring target configuration spaces, our DFT + MLFF + GNN methodology significantly reduces computational costs by three and one orders of magnitude, respectively, greatly enhancing research efficiency in vast adsorption configuration spaces [Figure 4E]. This acceleration allows for exploring extensive configuration spaces within a foreseeable short period, a capability previously unattainable^[25,29]. This dual-speed framework, integrating high-precision MLFFs with advanced graph representation learning, is also applicable to other catalytic systems with large search spaces, such as catalytic reaction path searches, stable adsorbate motif determination and protein-ligand structure prediction^[55-57]. The primary objective is to identify the global minimum adsorption configurations and their corresponding energies across varying surface coverages. Such critical information is unlikely to be fully captured within the smaller, randomly sampled set of $\sim 186,000$ configurations. In contrast, the comprehensive dataset of ~ 7 million configurations exhaustively enumerates nearly all possible adsorption states, thereby ensuring that the most stable adsorption configurations are accurately identified.

The generality of the proposed framework primarily manifests in two stages. Firstly, in the process of independent adsorption configuration enumeration, the method is applicable to all high-coverage adsorption configurations involving monodentate intermediates, such as H^{*}, N^{*}, O^{*}, and OH^{*}. For specific adsorbates, only minor adjustments are required. However, for configurations involving multidentate intermediates, our method still needs further improvement. Secondly, in the process of adsorption energy prediction, our approach is expected to be easily extended to metal surfaces with a fcc lattice. However, for metal surfaces with other lattice structures, some adjustments may be necessary based on the corresponding system.

RESULTS AND DISCUSSION

Coverage-dependent adsorption energy on different indexed surfaces

Leveraging our trained model for predicting adsorption energies, we calculated the steady-state CO adsorption energies for approximately 7 million configurations within the target configuration space, plotting CO adsorption energy versus coverage spectra for eight Cu crystallographic surfaces [Figure 5]. It was observed that each of the eight Cu surfaces corresponds to a unique CO coverage threshold, within which the adsorption energy of the most stable adsorption configurations changes minimally. Beyond this coverage threshold, the adsorption energy of the most stable configurations decreases significantly with increasing coverage. This trend aligns with physical intuition: at low coverages, the CO molecules in steady-state adsorption configurations are widely spaced, making non-bonding interactions between CO molecules negligible; however, at medium to high coverages, the surface arrangement of CO molecules in steady-state configurations becomes more compact, increasing molecular repulsion and thus increasing the system's potential energy, leading to decreased adsorption energies for CO molecules. Furthermore, aside from the (100) and (111) surfaces (where all surface atoms have the same coordination number), CO molecules preferentially occupy lower-coordination Cu sites that are energetically less favorable, and, with increasing coverage, gradually cover higher-coordination Cu sites [Supplementary Figure 5]. At the same time, the average coordination number of Cu atoms occupied by CO at each coverage level remains lower than the average coordination number of surface Cu atoms, indicating a clear lowest-energy orientation for CO adsorption. These computational results are consistent with previous research findings^[58-60]. Understanding the ease of C–C coupling on different Cu surfaces is crucial for developing Cu-based nanometal catalysts with high selectivity for C₂₊ products in CO₂RR reactions, as easier C–C coupling between CO molecules leads to higher selectivity for C₂₊ products^[61-65].

Here, we employed two straightforward metrics to evaluate the ease of C–C coupling on Cu surfaces. The first metric is the mean minimum C–C distance (MMCD) within the set of most stable adsorption configurations at various coverages, which serves as an indicator of the probability of C–C coupling. A smaller MMCD suggests a higher likelihood of coupling. The second metric is the characteristic coverage of the surface: taking the densely packed (111) surface as a reference and using its maximum adsorption energy as a threshold, the maximum coverage achievable by other surfaces without falling below this adsorption energy threshold is deemed their characteristic coverage. To facilitate successful C–C coupling, CO must exhibit sufficient adsorption strength on Cu surfaces to prevent the reactants from desorbing, which would interrupt the coupling reaction. Moreover, compared to the MMCD, the surface's characteristic coverage offers a more macroscopic dimension for representing the probability of C–C coupling, providing a holistic view. As illustrated in Figure 6, we analyzed these two metrics across eight Cu surfaces and found that the (310) surface exhibits the best C–C coupling performance, while the (111) surface performs the worst. The performance ranking of (310) > (210) > (311) > (100) > (111) is largely in agreement with experimental observations^[40,60]. High-index surfaces tend to have smaller MMCDs and greater characteristic coverages, indicating a higher probability of C–C coupling and better selectivity for C₂₊ products. This finding aligns with our current theoretical and experimental research, which shows that high-index surfaces offer a greater variety of surface sites and an abundance of low-coordination surface Cu atoms. These features provide more stable adsorption sites and a superior surface electronic environment conducive to C–C coupling^[66,67]. Additionally, the (322) surface emerged as a potential candidate due to its MMCD and characteristic coverage closely approaching those of the (310) surface, which has been experimentally proven to exhibit excellent selectivity for C₂₊ products^[40].

CONCLUSIONS

In this work, our investigation into the adsorption configurations of CO on Cu surfaces unveils critical

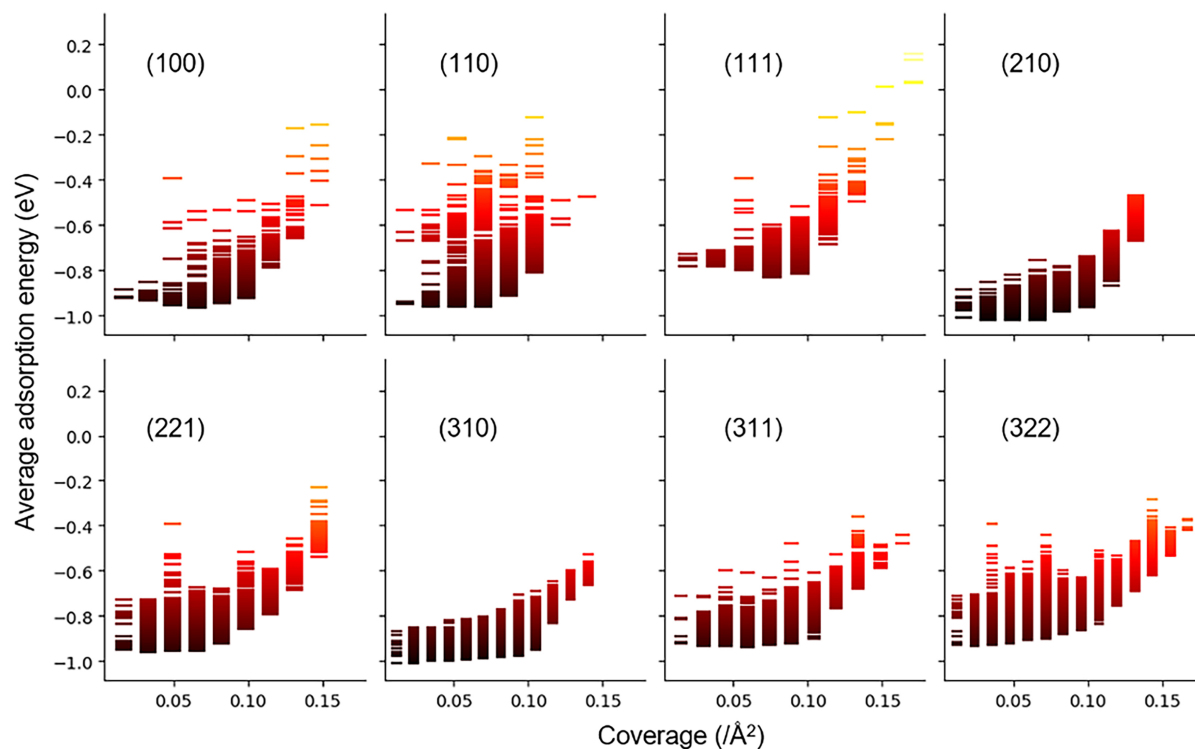


Figure 5. Coverage-Dependent CO Adsorption Energies on Cu Surfaces. The vertical axis corresponds to the average adsorption energy of CO, while the horizontal axis represents the number of CO molecules per unit area. The darker the bar color, the greater the average adsorption energy of CO.

insights into the mechanisms underlying electrocatalytic reactions. By leveraging advanced computational techniques, we have mapped out the energy landscapes of nearly 7 million configurations, identifying key trends in CO adsorption energies across different surface indices and coverages. Our findings underscore the importance of surface coverage in dictating the stability and activity of adsorption configurations, with implications for the efficiency of CO₂RR processes. The development and application of a MLFF, complemented by a graph-based adsorption energy prediction model, have significantly enhanced our ability to predict and understand the complex interactions at play. The discernment of high-index Cu surfaces as favorable for C–C coupling not only aligns with experimental observations but also opens new avenues for the design of catalysts with heightened selectivity for C₂₊ product formation. This work not only advances our theoretical understanding of catalytic mechanisms at high coverage but also sets the stage for future research aimed at optimizing catalyst designs for sustainable energy conversion technologies.

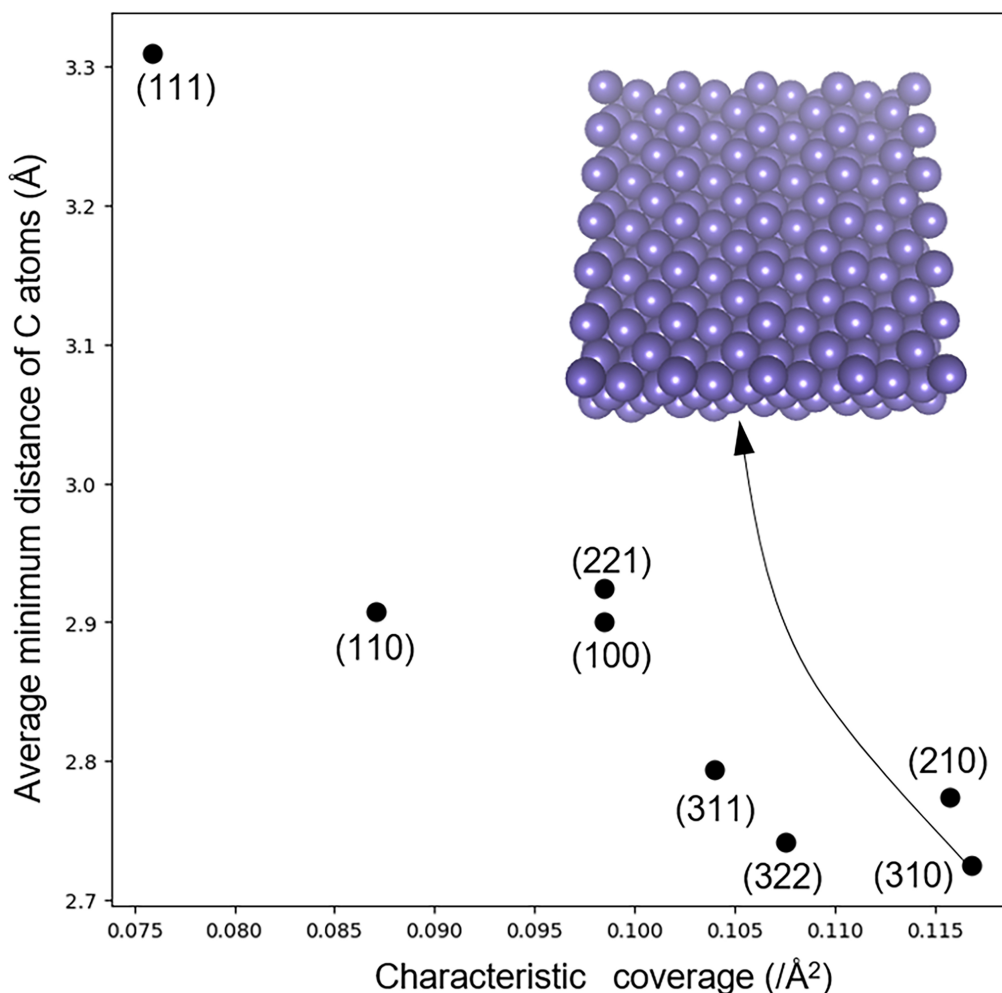


Figure 6. Correlation between C–C Distance and Coverage for Cu-Catalyzed Coupling. Surfaces such as Cu(310) with smaller C–C distances and higher characteristic coverages are identified as optimal for C–C coupling, enhancing selectivity for C_{2+} products in CO_2 reduction reactions. The inset illustrates the surface model of Cu(310).

DECLARATIONS

Authors' contributions

Conceptualization, data curation, investigation, methodology, validation, writing - original draft and revisions: Wu, S.; Li, S.; Zheng, S.

Methodology, discussion, data analysis: Zhang, W.; Zhang, M.

Supervision, methodology, writing - review: Li, S.; Pan, F.; Zheng, S.

Availability of data and materials

The data supporting the findings of this study are available within this Article and its [Supplementary Materials](#). Further data are available from the corresponding authors upon request. All the codes are provided at <https://github.com/PKU-WuSL/facet-dependent-coverage>.

Financial support and sponsorship

This work was supported by the National Natural Science Foundation of China (22402163), the Natural Science Foundation of Xiamen, China (3502Z202472001), the Soft Science Research Project of Guangdong Province (No. 2017B030301013), the Basic and Applied Basic Research Foundation of Guangdong Province

(2021B1515130002 and 2023A1515011391), and the Major Science and Technology Infrastructure Project of Material Genome Big-science Facilities Platform supported by Municipal Development and Reform Commission of Shenzhen.

Conflicts of interest

Pan, F. is an Associate Editor of the journal *Journal of Materials Informatics* but was not involved in any steps of the editorial process, including the selection of reviewers, manuscript handling, or decision-making. The other authors declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Beck, A.; Paunović, V.; van, B. J. A. Identifying and avoiding dead ends in the characterization of heterogeneous catalysts at the gas-solid interface. *Nat. Catal.* **2023**, *6*, 873-84. DOI
2. Li, X.; Mitchell, S.; Fang, Y.; Li, J.; Perez-Ramirez, J.; Lu, J. Advances in heterogeneous single-cluster catalysis. *Nat. Rev. Chem.* **2023**, *7*, 754-67. DOI PubMed
3. Bruix, A.; Margraf, J. T.; Andersen, M.; Reuter, K. First-principles-based multiscale modelling of heterogeneous catalysis. *Nat. Catal.* **2019**, *2*, 659-70. DOI
4. Handoko, A. D.; Wei, F.; Jenndy; Yeo, B. S.; Seh, Z. W. Understanding heterogeneous electrocatalytic carbon dioxide reduction through operando techniques. *Nat. Catal.* **2018**, *1*, 922-34. DOI
5. Tran, K.; Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution. *Nat. Catal.* **2018**, *1*, 696-703. DOI
6. Greeley, J.; Stephens, I. E.; Bondarenko, A. S.; et al. Alloys of platinum and early transition metals as oxygen reduction electrocatalysts. *Nat. Chem.* **2009**, *1*, 552-6. DOI PubMed
7. Wu, Z.; Li, Z.; Li, Y.; Zhang, Y.; Li, J. Improving the DFT computational accuracy for CO activation on Fe surfaces by Bayesian error estimation functional with van der Waals correlation. *Comput. Theor. Chem.* **2023**, *1219*, 113968. DOI
8. Araujo, R. B.; Rodrigues, G. L. S.; Dos, S. E. C.; Pettersson, L. G. M. Adsorption energies on transition metal surfaces: towards an accurate and balanced description. *Nat. Commun.* **2022**, *13*, 6853. DOI PubMed PMC
9. Nørskov, J. K.; Bligaard, T.; Logadottir, A.; et al. Trends in the exchange current for hydrogen evolution. *J. Electrochem. Soc.* **2005**, *152*, J23. DOI
10. Nørskov, J. K.; Rossmeisl, J.; Logadottir, A.; et al. Origin of the overpotential for oxygen reduction at a fuel-cell cathode. *J. Phys. Chem. B.* **2004**, *108*, 17886-92. DOI PubMed
11. Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I. B.; Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.* **2006**, *5*, 909-13. DOI PubMed
12. Liu, X.; Xiao, J.; Peng, H.; Hong, X.; Chan, K.; Nørskov, J. K. Understanding trends in electrochemical carbon dioxide reduction rates. *Nat. Commun.* **2017**, *8*, 15438. DOI PubMed PMC
13. Zhan, C.; Dattila, F.; Rettenmaier, C.; et al. Revealing the CO coverage-driven C-C coupling mechanism for electrochemical CO₂ reduction on Cu₂O nanocubes via operando raman spectroscopy. *ACS. Catal.* **2021**, *11*, 7694-701. DOI PubMed PMC
14. Cave, E. R.; Shi, C.; Kuhl, K. P.; et al. Trends in the catalytic activity of hydrogen evolution during CO₂ electroreduction on transition metals. *ACS. Catal.* **2018**, *8*, 3035-40. DOI
15. Ding, H.; Zheng, S.; Yang, X.; et al. Role of surface hydrogen coverage in C-C coupling process for CO₂ electroreduction on Ni-based catalysts. *ACS. Catal.* **2024**, *14*, 14330-8. DOI
16. Zheng, S.; Liang, X.; Pan, J.; Hu, K.; Li, S.; Pan, F. Multi-center cooperativity enables facile C-C coupling in electrochemical CO₂ reduction on a Ni₂P catalyst. *ACS. Catal.* **2023**, *13*, 2847-56. DOI
17. Zheng, S.; Ding, H.; Yang, X.; Li, S.; Pan, F. Automating discovery of electrochemical urea synthesis reaction paths via active learning and graph theory. <https://www.chinesechemsoc.org/doi/full/10.31635/ccschem.024.202404955> (accessed 2024-02-07).
18. Ghanekar, P. G.; Deshpande, S.; Greeley, J. Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis. *Nat. Commun.* **2022**, *13*, 5788. DOI PubMed PMC

19. Xiong, Y.; Wang, Y.; Zhou, J.; Liu, F.; Hao, F.; Fan, Z. Electrochemical nitrate reduction: ammonia synthesis and the beyond. *Adv. Mater.* **2024**, *36*, e2304021. DOI PubMed
20. Zheng, S.; Yang, X.; Shi, Z. Z.; Ding, H.; Pan, F.; Li, J. F. The loss of interfacial water-adsorbate hydrogen bond connectivity position surface-active hydrogen as a crucial intermediate to enhance nitrate reduction reaction. *J. Am. Chem. Soc.* **2024**, *146*, 26965-74. DOI PubMed
21. Enger, B. C.; Holmen, A. Nickel and Fischer-Tropsch synthesis. *Catal. Rev.* **2012**, *54*, 437-88. DOI
22. Chen, Y.; Wei, J.; Duyar, M. S.; Ordonsky, V. V.; Khodakov, A. Y.; Liu, J. Carbon-based catalysts for Fischer-Tropsch synthesis. *Chem. Soc. Rev.* **2021**, *50*, 2337-66. DOI PubMed
23. Weststrate, C. J. K.; Sharma, D.; Garcia, R. D.; Gleeson, M. A.; Fredriksson, H. O. A.; Niemantsverdriet, J. W. H. Mechanistic insight into carbon-carbon bond formation on cobalt under simulated Fischer-Tropsch synthesis conditions. *Nat. Commun.* **2020**, *11*, 750. DOI PubMed PMC
24. Ai, C.; Chang, J. H.; Tygesen, A. S.; Vegge, T.; Hansen, H. A. Impact of hydrogen concentration for CO₂ reduction on PdHx : a combination study of cluster expansion and kinetics analysis. *J. Catal.* **2023**, *428*, 115188. DOI
25. Deshpande, S.; Maxson, T.; Greeley, J. Graph theory approach to determine configurations of multidentate and high coverage adsorbates for heterogeneous catalysis. *npj. Comput. Mater.* **2020**, *6*, 345. DOI
26. Bang, K.; Hong, D.; Park, Y.; Kim, D.; Han, S. S.; Lee, H. M. Machine learning-enabled exploration of the electrochemical stability of real-scale metallic nanoparticles. *Nat. Commun.* **2023**, *14*, 3004. DOI PubMed PMC
27. Liu, P.; Wang, J.; Avargues, N.; et al. Combining machine learning and many-body calculations: coverage-dependent adsorption of CO on Rh(111). *Phys. Rev. Lett.* **2023**, *130*, 078001. DOI PubMed
28. Jenner, B.; Köbler, J.; McKenzie, P.; Torán, J. Completeness results for graph isomorphism. *J. Comput. Syst. Sci.* **2003**, *66*, 549-66. DOI
29. Sumaria, V.; Sautet, P. CO organization at ambient pressure on stepped Pt surfaces: first principles modeling accelerated by neural networks. *Chem. Sci.* **2021**, *12*, 15543-55. DOI PubMed PMC
30. Yang, Y.; Jiménez-Negrón, O. A.; Kitchin, J. R. Machine-learning accelerated geometry optimization in molecular simulation. *J. Chem. Phys.* **2021**, *154*, 234704. DOI PubMed
31. Jha, D.; Gupta, V.; Ward, L.; et al. Enabling deeper learning on big data for materials informatics applications. *Sci. Rep.* **2021**, *11*, 4244. DOI PubMed PMC
32. Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001. DOI PubMed
33. Ulissi, Z. W.; Tang, M. T.; Xiao, J.; et al. Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO₂ reduction. *ACS. Catal.* **2017**, *7*, 6600-8. DOI
34. Sumaria, V.; Nguyen, L.; Tao, F. F.; Sautet, P. Atomic-scale mechanism of platinum catalyst restructuring under a pressure of reactant gas. *J. Am. Chem. Soc.* **2023**, *145*, 392-401. DOI PubMed
35. Xu, W.; Reuter, K.; Andersen, M. Predicting binding motifs of complex adsorbates using machine learning with a physics-inspired graph representation. *Nat. Comput. Sci.* **2022**, *2*, 443-50. DOI PubMed
36. Boes, J. R.; Mamun, O.; Winther, K.; Bligaard, T. Graph theory approach to high-throughput surface adsorption structure generation. *J. Phys. Chem. A.* **2019**, *123*, 2281-5. DOI PubMed
37. Eren, B.; Zherebetsky, D.; Patera, L. L.; et al. Activation of Cu(111) surface by decomposition into nanoclusters driven by CO adsorption. *Science* **2016**, *351*, 475-8. DOI PubMed
38. Pérez-Gallent, E.; Figueiredo, M. C.; Calle-Vallejo, F.; Koper, M. T. Spectroscopic observation of a hydrogenated CO dimer intermediate during CO reduction on Cu(100) electrodes. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 3621-4. DOI PubMed
39. Nakano, H.; Nakamura, I.; Fujitani, T.; Nakamura, J. Structure-dependent kinetics for synthesis and decomposition of formate species over Cu(111) and Cu(110) model catalysts. *J. Phys. Chem. B.* **2001**, *105*, 1355-65. DOI
40. Hori, Y.; Takahashi, I.; Koga, O.; Hoshi, N. Electrochemical reduction of carbon dioxide at various series of copper single crystal electrodes. *J. Mol. Catal. A. Chem.* **2003**, *199*, 39-47. DOI
41. Wang, S.; Jian, M.; Su, H.; Li, W. First-Principles microkinetic study of methanol synthesis on Cu(221) and ZnCu(221) surfaces. *Chin. J. Chem. Phys.* **2018**, *31*, 284-90. DOI
42. Schouten, K. J. P.; Pérez, G. E.; Koper, M. T. M. Structure sensitivity of the electrochemical reduction of carbon monoxide on copper single crystals. *ACS. Catal.* **2013**, *3*, 1292-5. DOI
43. Guo, W.; Vlachos, D. G. Effect of local metal microstructure on adsorption on bimetallic surfaces: atomic nitrogen on Ni/Pt(111). *J. Chem. Phys.* **2013**, *138*, 174702. DOI PubMed
44. Deshpande, S.; Greeley, J. First-Principles analysis of coverage, ensemble, and solvation effects on selectivity trends in NO electroreduction on Pt₃Sn alloys. *ACS. Catal.* **2020**, *10*, 9320-7. DOI
45. Ojeda, M.; Nabar, R.; Nilekar, A. U.; Ishikawa, A.; Mavrikakis, M.; Iglesia, E. CO activation pathways and the mechanism of Fischer-Tropsch synthesis. *J. Catal.* **2010**, *272*, 287-97. DOI
46. Stróż, K. Plane groups - from basic to advanced crystallographic concepts. *Z. Kristallogr. Cryst. Mater.* **2003**, *218*, 642-9. DOI
47. Hoffmann, F. Symmetry in the plane: about wallpaper patterns, islamic mosaics, drawings from escher, and heterogeneous catalysts. In *Introduction to Crystallography*; Hoffmann, F., Eds.; Springer International Publishing: Cham, 2020; pp 127-50. DOI
48. Hahn, T. The 17 plane groups (two-dimensional space groups). In *International Tables for Crystallography Volume A: Space-group*

- symmetry*; Hahn, T., Ed. Springer Netherlands: Dordrecht, 2002; pp 92-109. DOI
49. Chanussot, L.; Das, A.; Goyal, S.; et al. Open catalyst 2020 (OC20) dataset and community challenges. *ACS. Catal.* **2021**, *11*, 6059-72. DOI
 50. Lu, D.; Wang, H.; Chen, M.; et al. 86 PFLOPS deep potential molecular dynamics simulation of 100 million atoms with *ab initio* accuracy. *Comput. Phys. Commun.* **2021**, *259*, 107624. DOI
 51. Weng, M.; Wang, Z.; Qian, G.; et al. Identify crystal structures by a new paradigm based on graph theory for building materials big data. *Sci. China. Chem.* **2019**, *62*, 982-6. DOI
 52. Li, S.; Chen, Z.; Wang, Z.; et al. Graph-based discovery and analysis of atomic-scale one-dimensional materials. *Natl. Sci. Rev.* **2022**, *9*, nwac028. DOI PubMed PMC
 53. Li, S.; Liu, Y.; Chen, D.; Jiang, Y.; Nie, Z.; Pan, F. Encoding the atomic structure for machine learning in materials science. *WIREs. Comput. Mol. Sci.* **2022**, *12*, e1558. DOI
 54. Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301. DOI PubMed
 55. Steiner, M.; Reiher, M. Autonomous reaction network exploration in homogeneous and heterogeneous catalysis. *Top. Catal.* **2022**, *65*, 6-39. DOI PubMed PMC
 56. Gu, G. H.; Lee, M.; Jung, Y.; Vlachos, D. G. Automated exploitation of the big configuration space of large adsorbates on transition metals reveals chemistry feasibility. *Nat. Commun.* **2022**, *13*, 2087. DOI PubMed PMC
 57. Qiao, Z.; Nie, W.; Vahdat, A.; Miller, T. F.; Anandkumar, A. State-specific protein-ligand complex structure prediction with a multiscale deep generative model. *Nat. Mach. Intell.* **2024**, *6*, 195-208. DOI
 58. Grabow, L. C.; Hvolbæk, B.; Nørskov, J. K. Understanding trends in catalytic activity: the effect of adsorbate-adsorbate interactions for CO oxidation over transition metals. *Top. Catal.* **2010**, *53*, 298-310. DOI
 59. Lausche, A. C.; Medford, A. J.; Khan, T. S.; et al. On the effect of coverage-dependent adsorbate-adsorbate interactions for CO methanation on transition metal surfaces. *J. Catal.* **2013**, *307*, 275-82. DOI
 60. Huang, Y.; Handoko, A. D.; Hirunsit, P.; Yeo, B. S. Electrochemical reduction of CO₂ using copper single-crystal surfaces: effects of CO* coverage on the selective formation of ethylene. *ACS. Catal.* **2017**, *7*, 1749-56. DOI
 61. Hou, J.; Chang, X.; Li, J.; Xu, B.; Lu, Q. Correlating CO coverage and CO electroreduction on Cu via high-pressure *in situ* spectroscopic and reactivity investigations. *J. Am. Chem. Soc.* **2022**, *144*, 22202-11. DOI PubMed
 62. Jin, J.; Wicks, J.; Min, Q.; et al. Constrained C₂ adsorbate orientation enables CO-to-acetate electroreduction. *Nature* **2023**, *617*, 724-9. DOI PubMed
 63. Zheng, Y.; Zhang, J.; Ma, Z.; et al. Seeded growth of gold-copper janus nanostructures as a tandem catalyst for efficient electroreduction of CO₂ to C₂₊ products. *Small* **2022**, *18*, e2201695. DOI PubMed
 64. Sun, W.; Wang, P.; Jiang, Y.; et al. V-doped Cu₂Se hierarchical nanotubes enabling flow-cell CO₂ electroreduction to ethanol with high efficiency and selectivity. *Adv. Mater.* **2022**, *34*, e2207691. DOI PubMed
 65. Yan, X.; Chen, C.; Wu, Y.; et al. Boosting CO₂ electroreduction to C₂₊ products on fluorine-doped copper. *Green. Chem.* **2022**, *24*, 1989-94. DOI
 66. Reller, C.; Krause, R.; Volkova, E.; et al. Selective electroreduction of CO₂ toward ethylene on nano dendritic copper catalysts at high current density. *Adv. Energy. Mater.* **2017**, *7*, 1602114. DOI
 67. Xiang, K.; Shen, F.; Fu, Y.; et al. Boosting CO₂ electroreduction towards C₂₊ products via CO* intermediate manipulation on copper-based catalysts. *Environ. Sci. Nano.* **2022**, *9*, 911-53. DOI