

# How Realistic are Idealized Copper Surfaces?

## A Machine Learning Study of Rough Copper-Water Interfaces

Linus C. Erhard,<sup>\*,†,‡</sup> Johannes Schörghuber,<sup>†,‡</sup> Aleix Comas-Vives,<sup>†</sup> and Georg K.  
H. Madsen<sup>\*,†</sup>

<sup>†</sup>Institute of Materials Chemistry, TU Wien, A-1060 Vienna, Austria

<sup>‡</sup>These authors contributed equally

E-mail: linus.erhard@tuwien.ac.at; georg.madsen@tuwien.ac.at

### Abstract

Copper is a highly promising catalyst for the electrochemical CO<sub>2</sub> reduction reaction (CO<sub>2</sub>RR) since it is the only pure metal that can form highly added-value products such as ethylene and ethanol. Since the CO<sub>2</sub>RR takes place in aqueous solution, the detailed atomic structure of the water-copper interface is essential for unraveling the key reaction mechanisms. In this study, we investigate copper-water interfaces exhibiting nanometer-scale roughnesses. We introduce two molecular dynamics protocols to create rough copper surfaces, which are subsequently brought into contact with water. From these interfaces, we sample additional training configurations from machine-learning-interatomic-potential-driven molecular dynamics simulations containing hundreds of thousands of atoms. An active learning workflow is developed to identify regions with high spatially resolved uncertainty and convert them into DFT-feasible cells through a modified amorphous matrix embedding approach. Finally, we analyze the local environments at the interface using unsupervised machine-learning techniques. Unique

environments emerge on the rough copper surfaces absent from model systems, including stacking-fault-induced configurations and undercoordinated corner atoms. Notably, corner atoms consistently feature chemisorbed water molecules in our simulations, indicating their potential importance in catalytic processes.

## Introduction

The increasing urgency of the climate crisis necessitates a rapid transition away from fossil fuels and towards renewable energy sources.<sup>1</sup> Copper is the only pure metal catalyst capable of producing substantial quantities of C<sub>2</sub> fuels such as ethylene and ethanol via the electrochemical CO<sub>2</sub> reduction reaction.<sup>2,3</sup> Given that the interface between water and copper plays a critical role in the individual reaction steps, significant research efforts have been invested in the experimental characterization of these interfaces, utilizing techniques such as X-ray photo spectroscopy,<sup>4</sup> X-ray diffraction,<sup>5</sup> or scanning tunneling microscopy.<sup>6–8</sup>

In addition to these experimental studies, a range of theoretical investigations have employed molecular dynamics (MD) simulations to study the copper-water interface, utilizing forces derived from density functional theory (DFT) or machine-learning interatomic potentials (MLIPs).<sup>9–14</sup> These simulations have provided valuable insights into the arrangement of water molecules at the interface, as well as the adsorption behavior and density fluctuations on low-index facets. Double peaks in water density profiles in the interface layer at pristine Cu(100) and Cu(111) confirm chemisorption of water molecules at ontop sites.<sup>11,12,14</sup> At stepped interfaces, the structure is dominated by the undercoordinated ridge sites, which serve as the primary sites for chemisorption.<sup>14</sup>

A significant limitation of these simulations is that they have primarily focused on ideal low-index copper surfaces, neglecting the complexity of rough surfaces, which can substantially influence the reaction mechanisms. *In situ* experiments have observed the roughening of the surface, with characteristics ranging from the Ångstrom scale<sup>15</sup> to height profiles with nanometer-scale variations,<sup>16,17</sup> and island sizes between 5 and 15 Å.<sup>16</sup> Notably, recent stud-

ies have shown that, under CO<sub>2</sub> reduction conditions, planar Cu(111) and Cu(100) surfaces are less stable than kinked surfaces,<sup>18</sup> significantly increasing the step density under reaction conditions by a factor of 4.1 and 5.2 for the Cu(111) and Cu(100) surfaces. Furthermore, several experimental studies have observed restructuring of copper surfaces *in situ*.<sup>19–23</sup> Restructuring and corresponding roughening are particularly significant, as recent studies have shown that the CO<sub>2</sub> reduction reaction tends to occur on steps and kinks, rather than on planar surfaces.<sup>18,24</sup> Moreover, other studies have shown that a rougher surface significantly increases the selectivity towards the production of C<sub>2+</sub> products.<sup>25,26</sup>

These findings emphasize the need for atomistic simulations of rough copper surfaces of nanometric size. However, such simulations pose a significant challenge, as they are beyond the reach of traditional DFT methods, which are limited to a few hundred atoms. Recent advances in MLIPs have enabled the simulation of much larger systems, with efficient feature-based techniques<sup>27,28</sup> facilitating simulations of up to a billion atoms<sup>29</sup> or long simulations of up to tens of nanoseconds.<sup>30</sup> At the same time, equivariant graph representations<sup>31–33</sup> have substantially improved the accuracy of machine-learning interatomic potentials.<sup>34</sup>

Nevertheless, significant challenges persist in large-scale simulations based on MLIPs. Developing effective active learning workflows requires accurate locally resolved uncertainties that correlate directly with the true error. Recent work demonstrates that spatially averaged committee uncertainties, unlike standard committee uncertainties, provide accurate error estimation.<sup>35</sup> While this enables per-atom error quantification, the necessary length scales are computationally prohibitive for direct DFT assessment. Instead, high-uncertainty environments need to be extracted into smaller representative models, which introduces boundary condition challenges. Existing mitigation strategies include embedding environments in amorphous matrices,<sup>36</sup> minimizing boundary uncertainties,<sup>37,38</sup> or reconstructing periodic crystalline arrangements.<sup>39</sup>

A further distinct challenge exists in the analysis of these large-scale simulations, where surfaces are often ill-defined and environments are highly diverse. Here, unsupervised machine-

learning methods are particularly valuable, as they do not rely on prior assumptions. Such techniques have previously successfully identified structural motifs on model copper-water interfaces,<sup>14</sup> detected defects in crystals,<sup>40</sup> and tracked structural changes at surfaces.<sup>41</sup>

In this work, we leverage recent advances in MLIPs to perform atomistic simulations containing more than 200,000 atoms, allowing for a realistic representation of the diversity of rough copper surfaces. To this end, we design an active learning workflow that samples training data from the interfaces and maps it to DFT-feasible cells. This allows us to sample training data for the local environments at the rough interfaces, thereby developing suitable MLIP for copper-water interfaces that were previously out of reach. Combining this with unsupervised machine-learning techniques, we gain new insights into the structure and behavior of water at rough copper surfaces, and directly compare these findings with idealized model surfaces and the training data selected during the active learning workflow.

## Computational Details

This section describes our computational methodology, including DFT settings, MLIP fitting, and MD simulations. Our computational workflow relied on **ASE**<sup>42</sup> and **OVITO**<sup>43</sup> for data processing and visualization.

## Molecular dynamics

We utilized the **LAMMPS**<sup>44</sup> code for all molecular dynamics simulations in this study. We used a time step of 2 fs for all simulations including only copper, and 0.5 fs for all simulations including copper and water. We used temperature and pressure damping parameters of 100 and 1000 times the time step, respectively. Rough copper surfaces were generated with an existing Cu-Zr MLIP,<sup>45</sup> while copper-water interface simulations utilized our custom-fitted potentials described in Section “Potential fitting”.

## Density functional theory

We employed VASP 6.4.2<sup>46–48</sup> with the RPBE functional<sup>49</sup> for our DFT calculations. Moreover, we applied the D3 correction<sup>50</sup> with a zero damping scheme to all atoms. We used an energy cutoff of 850 eV and Gaussian smearing with a width of 0.05 eV. The k-point density was set to correspond to a  $11 \times 11 \times 11$  grid for a one-atom primitive unit cell of fcc-Cu in the slab direction. In the direction perpendicular to the slab, a single k-point was used. We employed hard projector augmented wave pseudopotentials<sup>51</sup> for hydrogen and oxygen, and the standard version for copper. Our DFT parameters were largely based on those used in Ref. 14. However, we applied the D3 correction to all atoms, rather than just the surface copper atoms and water molecules, due to the complexity of defining surface atoms in our rough surfaces.

## Potential fitting

We used atomic cluster expansion (ACE) and graph atomic cluster expansion (GRACE) type potentials in this study. For fitting the ACE potentials, we used the pacemaker code.<sup>27,37,52,53</sup> Specifically, we utilized nonlinear ACE with 9 embeddings, which has been previously demonstrated to be highly effective.<sup>36,45</sup> We also explored the effects of varying the cutoff radius (5 Å, 6 Å, and 7 Å) and the number of basis functions (900, 1200, and 1500). A detailed discussion of the parameter effects can be found in the results section. We used the gracemaker code<sup>32</sup> to fit the GRACE potentials. For all GRACE models, including GRACE 1-Layer and GRACE 2-Layer, we employed the small model complexity. Additionally, for GRACE 1-Layer, we tested cutoff radii of 5 Å, 6 Å, and 7 Å. We used a fixed cutoff radius of 5 Å for GRACE 2-Layer. Throughout the active learning process, we consistently used a GRACE 1-Layer potential with a 5 Å cutoff radius.

# Results and Discussion

## Rough copper surfaces

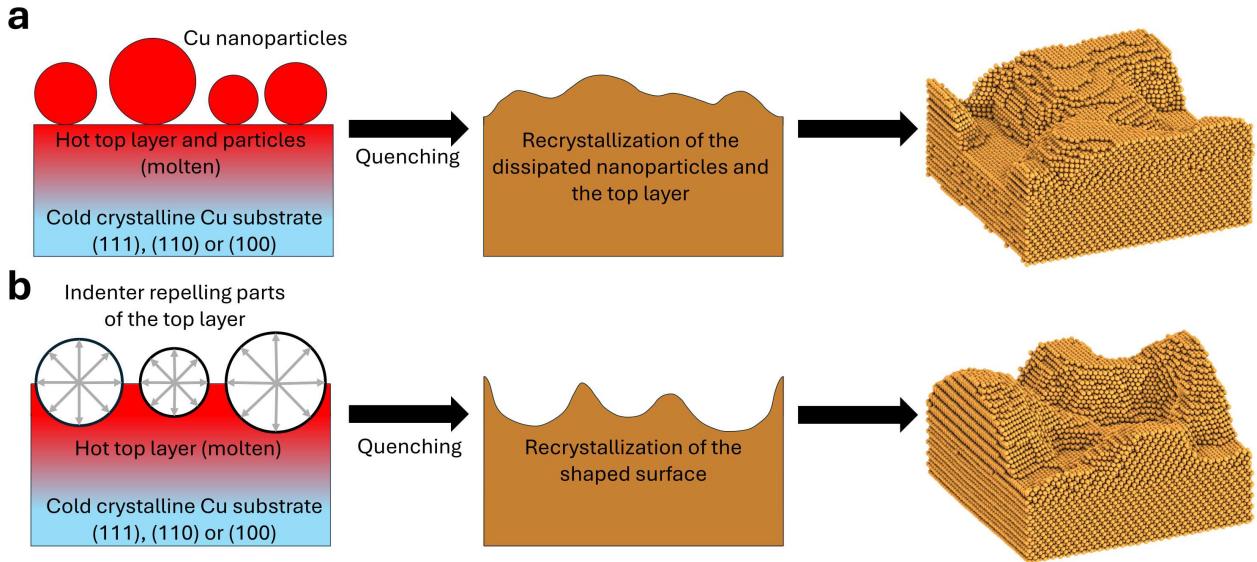


Figure 1: **Protocols to generate rough copper surfaces.** (a) In this protocol, we place copper nanoparticles on top of a copper slab. The upper part of the slab and the nanoparticles are then molten in a molecular dynamics simulation, while the lower part is kept at 500 K. In a subsequent step, the upper part is quenched to 500 K, initiating the recrystallization of the entire upper part. (b) In this protocol, we melt the upper part of the slab at 1500 K while keeping the lower part at 500 K and crystalline. Then, we press into the top layer using indenters, which push away atoms at the surface. Again, we quench the upper layer, leading to recrystallization of this part of the structure.

Figure 1 illustrates two different methods for creating rough copper surfaces. In the first method, shown in Figure 1a, we placed copper nanoparticles on top of a copper surface slab. The radius of the nanoparticles was determined randomly with an average size ranging from 10 to 30 Å. Copper slabs with Miller indices of (100), (110), and (111) were used as the substrate. We used supercells of  $50 \times 58 \times 15$  for the (111) surface,  $50 \times 50 \times 17$  for the (100) surface, and  $35 \times 50 \times 24$  for the (110) surface. Molecular dynamics simulations with simulations times between 1 ns and 4.6 ns (for details see Table S1) of the copper slab with nanoparticles were performed, where the upper part, including the nanoparticles, was annealed (temperatures see Table S1) while the lower part was maintained at 500 K. This

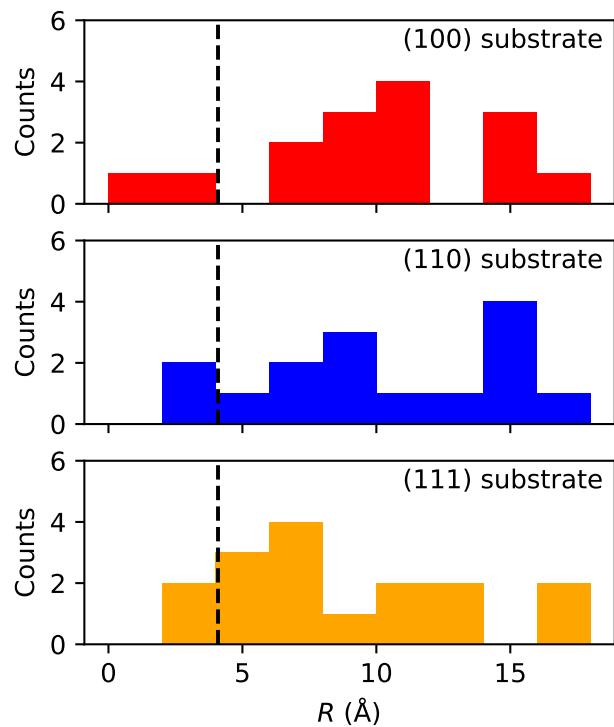
enabled the upper part to melt and reorganize. Following annealing, the upper portion was quenched to a temperature well below the melting point, inducing recrystallization. This approach enabled the generation of rough surfaces with relatively low surface energies and fully crystalline structures. Notably, the surface roughness can be controlled by varying three key factors: the initial nanoparticle size, the maximum temperature, and the quench rate. Two distinct MD protocols, as outlined in Table S1, were used in combination with variations in nanoparticle size to generate a range of surfaces. This resulted in a total of 30 surfaces, with ten surfaces generated for each of the (100), (110), and (111) slab types.

The second method, illustrated in Figure 1b, begins with a pristine Cu (111), (110), or (100) surface. The upper part of the surface is then heated to 1500 K, while the lower part is maintained at 500 K. In this top layer, indenters are inserted, which are represented by dummy particles that interact with the copper via a purely repulsive Lennard-Jones potential. These indenters repel the copper atoms at the surface, inducing roughness. The indenters are then held at a constant position, and the top layer is quenched to 500 K, inducing recrystallization. The exact simulation protocol is outlined in Table S2. Varying the distribution of indenters, intrusion depth, and radius allows for the generation of various degrees of roughness. By randomly varying these parameters, a total of 16 surfaces were produced, comprising six (111) surfaces, and five (100) and (110) surfaces each.

In Figure 2, we provide the surface roughness distribution of all the 46 surfaces created. We calculate the root-mean-square roughness by,

$$R = \sqrt{\frac{1}{N} \sum_j^N (z_j - \bar{z})^2}, \quad (1)$$

with the number of atoms at the surface  $N$ , the  $z_j$  coordinate of atom  $j$  and the mean  $z$  coordinate  $\bar{z}$ . The values of the root-mean-square roughness can be as high as 18 Å and are nearly independent of the substrate surface we used. The smallest value of the surface roughness is around 2 Å. This is in good agreement with the roughest surfaces from Ref. 15,



**Figure 2: Roughness of the generated copper surface for different substrates.** We show the count of the root-mean-square roughness value for all rough copper interfaces we have generated. The results are shown separately for different types of substrate surfaces. The dashed black line marks the roughness of the surface we used for our final analysis.

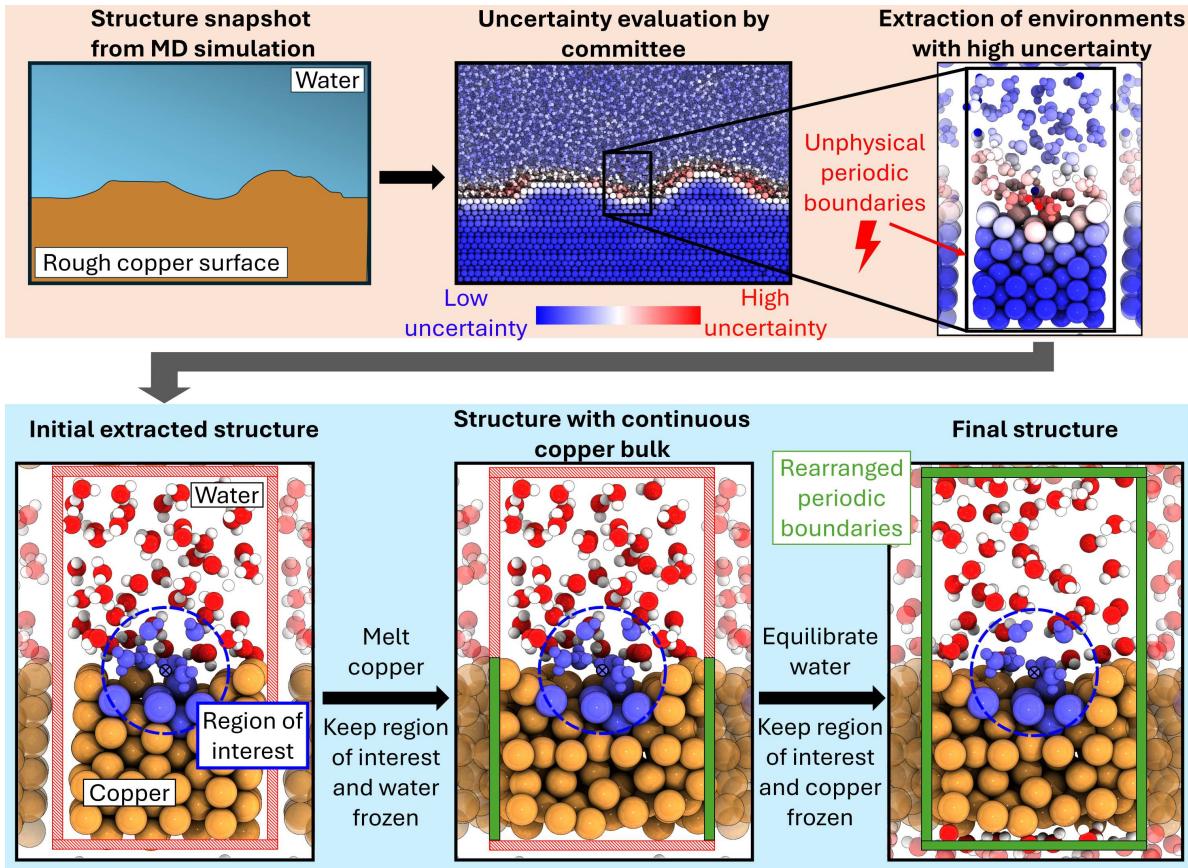
which are slightly below 2 Å. The studies with a stronger difference in the height profile<sup>16,17</sup> do not provide a value for the roughness. Since these would presumably be higher, with our structures, we are able to cover a wide range of possible surface roughness.

## Active learning and training data

Accurate modelling of rough copper-water interfaces with MLIPs requires training data for these more complex interface arrangements. In this section, we will illustrate a way to sample training data for these rough copper-water interfaces using an active learning workflow. As a starting point for our study, we utilized a training database for the copper-water system developed earlier in our group.<sup>14</sup> This database focused primarily on model surfaces like (100), (110), and (111), with additional data for more stepped surfaces like (211), (322), and (433). However, the database was not designed to accommodate rough copper surfaces, consisting not only of various types of edge and corner atoms, but also potentially different types of defects like vacancies or stacking faults.

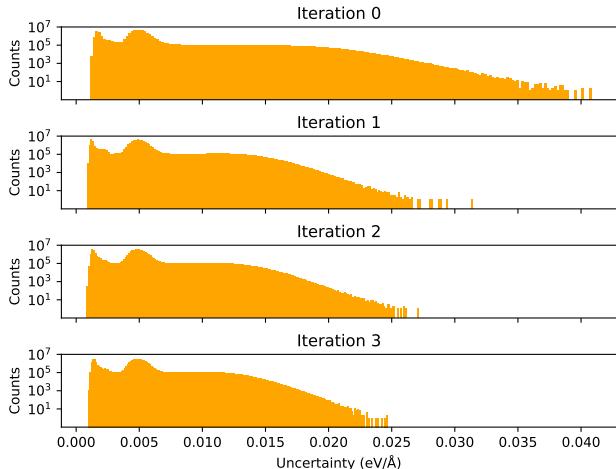
To address this limitation, we designed an active learning workflow to extract environments from large-scale simulations with high spatially resolved uncertainties<sup>35</sup> and convert them into DFT-feasible cells using a modified amorphous matrix embedding approach.<sup>36</sup> The workflow is illustrated in Figure 3 and used as a starting point for the surfaces we generated in Fig. 1.

Using a committee of potentials, we evaluated the spatially resolved uncertainties within a cutoff of 4 Å around each atom. We found that atoms with high uncertainty were predominantly located at the interface, in good agreement with earlier observations.<sup>14</sup> To calculate the forces and energy of a configuration with high uncertainty, we extracted the configuration and its surrounding area into a DFT-feasible box, as illustrated in Figure 3. This initial simulation box had unreasonable boundaries due to a small vacuum layer. To resolve this issue, we employed a two-stage annealing procedure. First, we annealed the copper slab to 600 K while keeping the water and region of interest fixed. The region of interest



**Figure 3: Small-scale extraction of interface structures.** We sample structures from large-scale molecular dynamics simulations of rough copper-water interfaces. Using spatially resolved uncertainties based on committee errors,<sup>35</sup> we determine the uncertainty for each atom. Within our simulations, the uncertainty at the interface was always the largest. Since we cannot calculate forces and energies for the entire interface, we need to extract small-scale representations. For this, we use a variation of amorphous matrix embedding.<sup>36</sup> Within this approach, we extract a small cell from our simulation box. The atoms with the highest uncertainty are then placed centrally in the box. This central region, within a certain cutoff around the atom with high uncertainty, is the region of interest (marked blue), as training data is missing for these configurations. As we impose periodic boundary conditions in our DFT calculations, we must address the boundary artifacts that arise from inserting the extracted atoms into an arbitrary box. To fix this, we use a two-step protocol. First, we anneal the copper at high temperatures, while keeping the water and the entire area of interest fixed. This allows the copper to rearrange at the copper-copper interface, marked by the green interface area. In a second step, we equilibrate the water at room temperature, while keeping the copper and the region of interest fixed. This allows for rearrangement at the other interfaces, while the region of interest remains the same as in the beginning.

was defined as all atoms within a certain cutoff around the central atom, which was the one with high uncertainty. This annealing step allowed us to eliminate the vacuum layer, resulting in a reasonably well-arranged, albeit amorphous, copper structure. In the second step, we repeated the process with water, letting it equilibrate at room temperature while keeping the region of interest and copper fixed. This enabled us to fill the vacuum region and re-relax the boundaries of all parts in a reasonable manner. The resulting structure was suitable for DFT convergence without significant issues, while the environment around the atom with high uncertainty remained unchanged. This approach is a modification of the recently proposed amorphous matrix embedding,<sup>36</sup> with the key difference being the use of a two-stage annealing procedure due to the large difference in melting points between water and copper.



**Figure 4: Change of uncertainty over several iterations.** We evaluated the uncertainty of the committees from several iterations on the same snapshots from the last iteration’s molecular dynamics run.

We performed a total of three active learning iterations. In the first iteration, we conducted large-scale molecular dynamics simulations on a single rough copper surface exhibiting a root-mean-square roughness of  $4 \text{ \AA}$  for an extended period. We then sampled 129 local environments using an uncertainty threshold of  $0.032 \text{ eV \AA}^{-1}$ . Figure 4 shows the reduction in uncertainty of the final trajectory over AL iterations. The data generated in the first batch significantly reduced the uncertainty of the potentials. In the subsequent iterations,

we sampled training data for shorter time scales for all 46 rough interfaces generated using uncertainty thresholds of  $0.023 \text{ eV } \text{\AA}^{-1}$  in the second iteration, and  $0.022 \text{ eV } \text{\AA}^{-1}$  in the third iteration to determine which configurations to add to the database. As can be seen in Figure 4, later generations resulted in smaller improvements than the first. Using this approach, we sampled in total 238 configuration, with 129 configurations sampled in the first iteration, 44 in the second iteration and 65 in the third iteration. For the test set, we sampled configurations from environments with slightly lower uncertainties than our threshold, resulting in 38 additional configurations. This test set was then combined with the test set from Ref. 14.

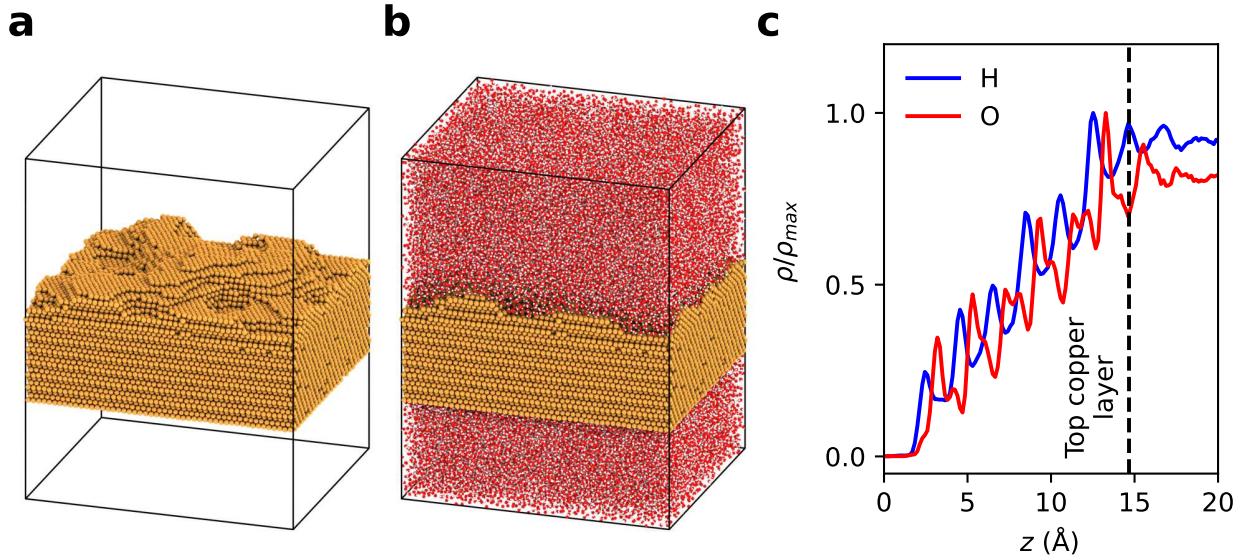
## Selection of the MLIP model

Having successfully generated rough copper-water interfaces and created a comprehensive database that can describe these interfaces, we required a reliable machine learning interatomic potential (MLIP) for our final simulations. To this end, we compared the performance of GRACE 1-Layer, GRACE 2-Layer, and ACE models. Figure S1 presents a comparison of the computational efficiency and accuracy of these models, in terms of root-mean-square errors. While GRACE 2-Layer offers the highest accuracy, its limitation to systems with fewer than 10,000 atoms makes it unsuitable for our purposes, given that our systems comprise around 200,000 atoms. Therefore, we considered the other two models. As shown in Figure S1, ACE models exhibit significantly lower computational costs, albeit with slightly higher errors compared to GRACE 1-Layer models. Details on the parameter settings can be found in Section “Potential fitting”. Although the high computational efficiency of ACE models might make them an attractive choice, we evaluated the models based not only on their errors on the test set but also on their structural predictions at the copper-water interface. We investigated the model surfaces (100), (110), (111), (211), (322), and (433) and the arrangement of water at these surfaces, using the GRACE 2-Layer model as a reference, which is suitable for the system sizes required for the model systems. The results for the

density curves of water as a function of distance to the surfaces are presented in Figure S2-S4, for oxygen and hydrogen separately and for all atoms together. While all models are in reasonable agreement with the GRACE 2-Layer reference, except the ACE potential with a 5 Å cutoff, a detailed analysis reveals some smaller discrepancies. Notably, the first peak corresponding to chemisorbed water is underestimated for the (100) and (111) surfaces, and the hydrogen distribution deviates from the reference for the (111) surface, failing to reproduce the shape of the broad first peak. Smaller deviations are observed for the (322) and (433) surfaces, particularly in the shape of the first peak in the oxygen density function. Given that these surfaces are stepped and our interest lies in rough surfaces with many steps, these distribution functions are particularly important. Ultimately, the choice of surrogate model is challenging, but we selected the GRACE 1-Layer potential with a 6 Å cutoff for our final simulation, as it accurately describes the interface between water and stepped Cu surfaces.

## Interface structure

Figure 5a shows the rough copper surface selected for detailed analysis of the interface. The structure was generated via the indenter method and exhibits a root-mean-square roughness of 4 Å, which represents a balanced case with significant roughness consistent with experimental measurements<sup>15-17</sup> while avoiding extreme values. Figure 5b shows the full simulation box containing both the copper surface and water molecules. In Figure 5c hydrogen and oxygen density profiles obtained from the MD simulations of Cu-H<sub>2</sub>O interfaces are shown. Due to the surface roughness, we observe increased water densities between the lowest and highest copper surface layers. The first oxygen peak appears at approximately 2.9 Å, in agreement with earlier work on model surfaces.<sup>11,14</sup> This is followed by periodic peaks arising from the copper layers between the lowest and highest copper surface layers. These oxygen peaks coincide with hydrogen peaks at slightly lower heights, consistent with the density profiles observed for the model stepped surfaces in the previous work and can

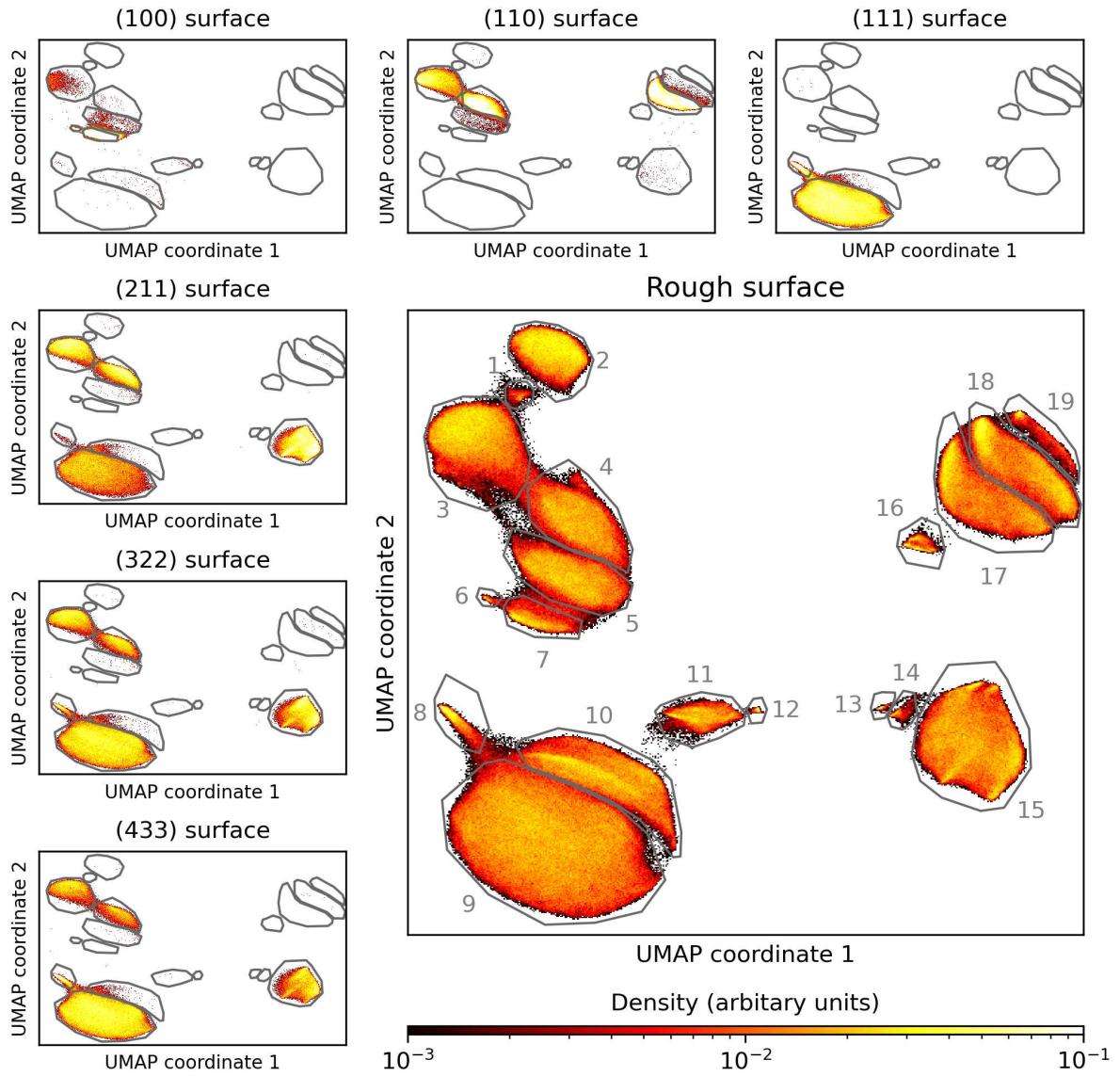


**Figure 5: Selected structure, corresponding simulation box and water distribution on the rough surface.** (a) Rough copper structure selected for detailed analysis. (b) Simulation box containing the rough copper surface and water bulk phase. (c) Hydrogen and oxygen density profiles as a function of the distance to the first surface layer, with the dotted black line indicating the highest copper layer position. The curves were normalized with respect to their global maximum.

be attributed to the water chemisorbed at step edges inducing a H-down orientation of the adjacent water molecules in order to facilitate hydrogen bonding.<sup>14</sup>

Global descriptors such as water density profiles are commonly employed to characterize the structure of metal-water interfaces in model systems defined by Miller indices. However, on rough surfaces, the absence of a well-defined surface onset complicates interpretation, motivating the use of alternative methods. To capture the complexity of the interface structure, we instead employ unsupervised learning to classify local atomic environments. Similar approaches have been used to study copper-surface dynamics at elevated temperatures<sup>41</sup> and identify crystal defects,<sup>40</sup> but we focus specifically on the interfacial copper atoms, following the methodology outlined in Ref. 14. Local environments are encoded invariantly using spherical Bessel descriptors<sup>54,55</sup> and embedded via UMAP<sup>56</sup> for both model and the rough interface.

Figure 6 visualizes and enumerates 19 distinct clusters representing diverse local environ-



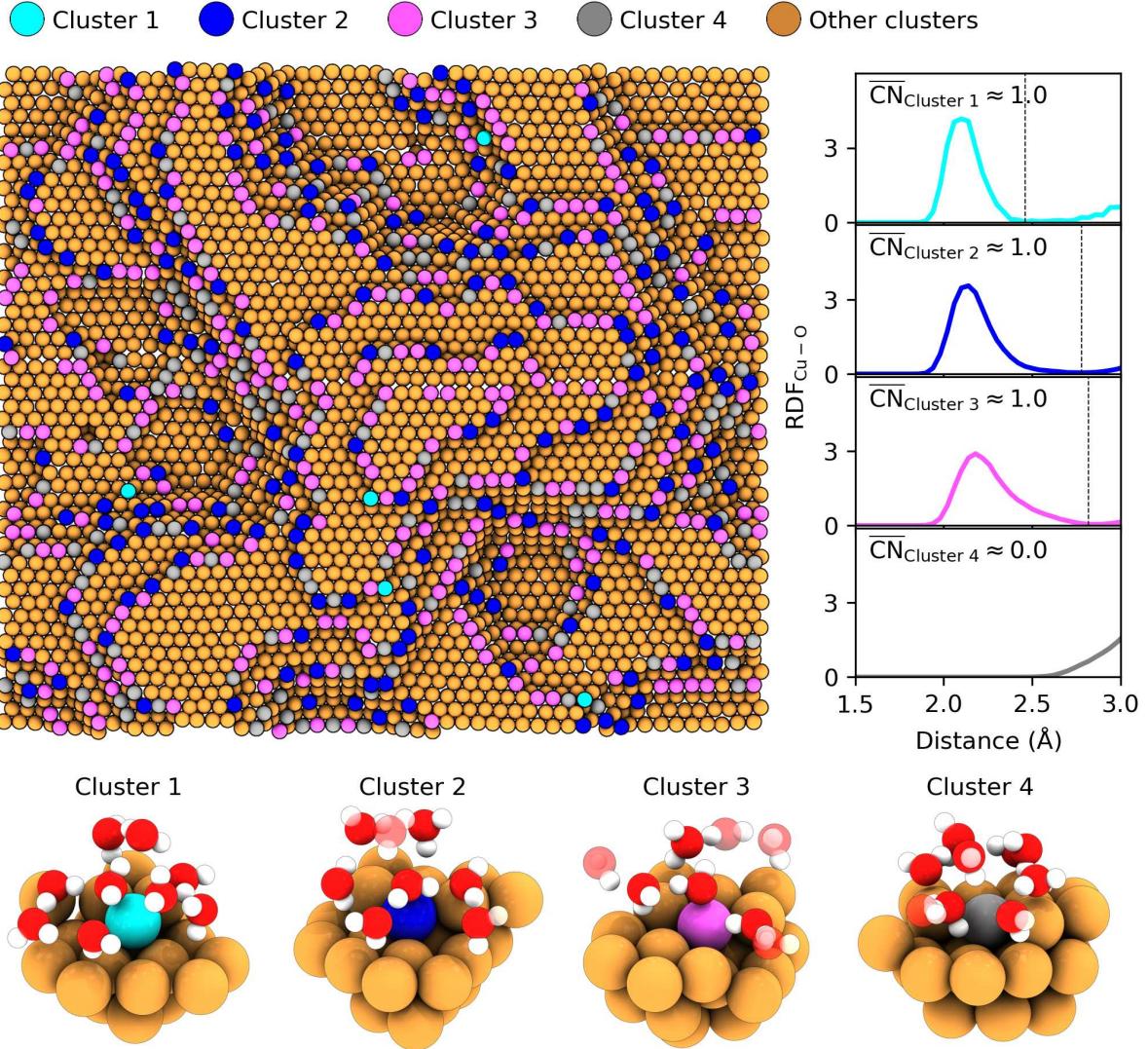
**Figure 6: UMAP embeddings of copper atoms at model surfaces and the rough surface.** The smaller panels show two-dimensional UMAP embeddings of copper atoms at various copper-water model interfaces, in particular (100), (110), (111), (211), (322), and (433). Additionally, we show in the large panel the same plot for copper surface atoms in one of our rough interface models. Moreover, we marked different clusters by gray lines and a corresponding number. These clusters are also shown for the model surfaces. Some clusters only appear on the rough copper-water interface, while all clusters that are part of the model surfaces also appear in the rough interface model. The UMAP embedding was fit on all descriptors obtained for the rough surface with 30 neighbors, a minimum distance of 0.1, and the Euclidean distance metric, and subsequently applied to the descriptors for the model surfaces.

ments of top-layer copper atoms in the rough surface together with the UMAP embeddings of the copper-water model interfaces. The rough interface encompasses all environment types observed on the model surfaces, while also featuring additional clusters absent from any model interfaces. A comprehensive classification, including representative structures, Figure S5-S11, and cluster assignments, Table S3, is provided in the SI.

The primary distinguishing feature is the copper coordination number. Clusters 1-7 all correspond to undercoordinated copper sites, including environments reminiscent of (110) or (100) facets and some step edges of the Cu( $n+1, n, n$ ) surfaces included in the initial database. Strongly undercoordinated copper atoms are found at step-edge intersections, forming clusters 1, 2, and 6, which are not represented in the model structures. While slopes generated by the indenters expose some (110) and (100) facets, the majority are (111), consistent with the surface being generated on a Cu(111) substrate. These environments, clusters 8 and 9, are well represented by the model surfaces. Clusters 10-12 and 15-18 correspond to edge/corner atoms located directly beneath step edges. Clusters 13, 14, and 16 occur exclusively at stacking faults and are therefore also unique for our rough copper surface. Finally, cluster 19 corresponds to atoms at the bottom of surface vacancies. Most of these latter clusters capture defect-like environments inaccessible to idealized surface models. Our unsupervised classification approach thus uncovers structural motifs without prior system knowledge.

In addition to the primary features related to the copper coordination, differences in Cu-H<sub>2</sub>O coordination further distinguish clusters. This effect was already observed for the (111) surface where one cluster represented chemisorbed water while the other lacks chemisorption.<sup>14</sup> The same distinction can be recognized for the clusters 8 and 9, where cluster 8 represents chemisorbed water at (111) facets and cluster 9 (111) facets without chemisorbed water. Similarly, clusters 3 and 4 and 6 and 7, correspond to step edge environments differing by their water coordination.

Figure 7 illustrates the structural resolution achieved through our clustering approach by

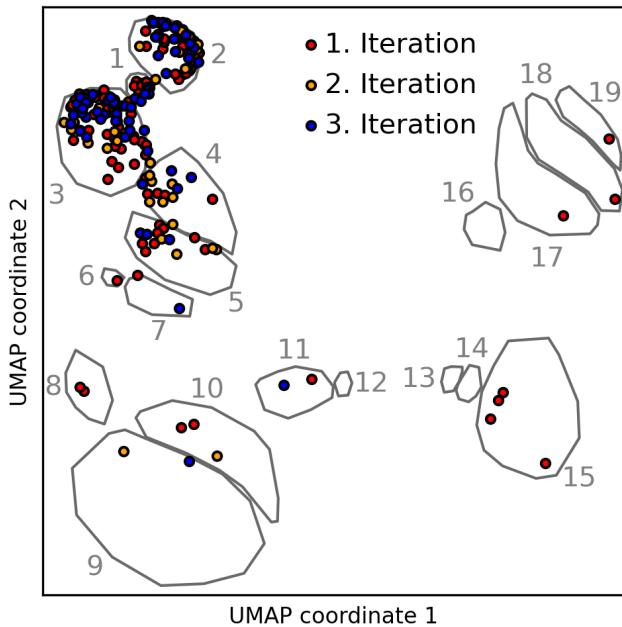


**Figure 7: Cluster classification on a rough copper surface.** Examples of clusters 1-4 from Figure 6. On the upper left we show copper atoms colored according to cluster assignment on our analyzed rough copper structure. On the right we display oxygen-copper radial distribution functions with corresponding coordination numbers, determined by integrating the first peak up to the first minimum (gray dotted line). A coordination number of 1 indicates a chemisorbed water molecule on-top of a copper atom, while 0 indicates no chemisorption. In the lower panel we show local atomic environments within 5 Å of representative atoms for each cluster. Semi-transparent oxygen and hydrogen atoms lie beyond the spatial cutoff but are included to demonstrate the absence of isolated hydrogen and oxygen atoms.

presenting cluster 1-4 as representative cases. The proximity of these clusters in the UMAP embedding (Figure 6) reflects similar local copper environments. We will first focus on clusters 3 and 4. Both clusters represent edges of stepped (111) facets and closely resemble environments found on stepped surfaces like (211), (322), and (433), as reflected by the appearance of similar clusters in their UMAP embeddings. However, water adsorbs directly on top of copper atoms in cluster 3, while no analogous adsorption is observed in cluster 4. This distinction was quantified through partial radial distribution functions (Figure 7 right panel), which yields average oxygen coordination numbers and confirms water chemisorption occurs on all atoms in cluster 3 but is absent in cluster 4.

Clusters 1 and 2 represent corner atoms on (111) facet with only two or three in-plane neighbors, Figure 7. These distinctive configurations are exclusively found on rough surfaces, as low-index crystalline planes lack such sites. Remarkably, persistent water chemisorption occurs on all atoms within these clusters, with no equivalent copper coordination environments lacking water chemisorption. This highlights a unique structure-adsorption correlation with potential catalytic relevance and underscores the necessity of modeling non-ideal surfaces in theoretical investigations. The radial distribution functions in Figure 7 further reveal a sharper first peak for cluster 1 relative to cluster 2, indicating stronger water binding to these atoms. This enhanced bonding aligns with the undercoordination of Cluster 1 sites and likely facilitates the approach of secondary water molecules within 3 Å of the adsorption site, as suggested by distinct features beyond the primary hydration shell.

Finally, the classification of local atomic environments provides insights into the active learning process by enabling systematic categorization of structures added during iterative sampling. Interfacial atoms consistently exhibited the highest local uncertainties, ensuring that the fixed region of interest (see workflow Fig. 3) exclusively sampled these. Overall, during our active learning workflow, we sampled 238 configurations. For each configuration, we show the UMAP embedding of the central copper atom, or the closest copper atom if the central atom is not copper, in Figure 8. Analysis reveals that the majority of these



**Figure 8: Classification of environments extracted during active learning into the UMAP landscape.** The UMAP embedding shows copper atoms with the highest uncertainty from the structures extracted during active learning from the large-scale simulations. In cases where a non-copper atom had the highest uncertainty, we selected the closest copper atom to the atom with the highest uncertainty. The gray clusters are identical to those used for data classification in Figure 6.

atoms' local environments belong to clusters 1 to 3. Clusters 1 and 2, identified above as corner sites with low in-plane copper coordination, were absent from the initial training database, confirming that spatially resolved uncertainties successfully pinpoint underrepresented structural motifs. Interestingly, the significant representation in cluster 3 suggests that additional training data for chemisorbed water at edge sites was required for accurate system description. This is unexpected, as these sites should be covered by prior training data from low-index copper-water interfaces, see Ref. 14. We also observe that clusters 11, 12, 13, 14, 18, and 19 are rarely sampled despite being unique to the rough interface. This rarity may arise because these environments may already be similar to bulk copper (due to minimal surface exposure) and lack water adsorption, reducing chemical complexity and subsequently the uncertainty. Finally, we can assess how the distribution of environments changes over the active learning cycles. The initial cycle features a broad sampling across clusters, most of which represent surfaces not previously observed by the model. However, in subsequent cycles, the sampling shows a preference for the corner-like sites within clusters 1, 2, and 3. In summary, our workflow's ability to automatically target complex geometries underscores a key advantage of local uncertainty quantification over structure-wide measures, which would fail to enable such precise sampling.

## Conclusion

We present an explicit atomistic simulation of a rough metal-water interface. To develop an accurate machine-learning interatomic potential (MLIP) for this system, we implemented an active learning workflow that leverages spatially resolved uncertainties to identify substructures within the interface. This approach enabled the extraction of DFT-feasible configurations and provides a generalizable workflow for generating MLIP training data for large-scale systems.

Unsupervised classification of local atomic environments reveals a diverse ensemble of

structural motifs at the Cu-H<sub>2</sub>O interface. While individual model interfaces capture subsets of these environments, they fail to capture the full complexity present on rough surfaces. Water chemisorption occurs predominantly at undercoordinated Cu atoms, with step-edge intersections showing the strongest binding, suggesting these sites as primary candidates for catalytic reaction centers. In contrast, chemisorption on terrace atoms accounts for only a minor fraction of the observed sites. Our unsupervised approach identifies chemically distinct local structures without requiring a priori knowledge of relevant features.

This work establishes a foundation for simulating interfaces with realistic surface morphologies, while also highlighting directions for future investigation. Although the present focus is on qualitative interfacial structure assessment, extensions could include carbon-based reactants, e.g. by monitoring carbon environments to identify species such as CO<sub>2</sub> and CO, as well as variable copper oxidation states, which play an important role in CO<sub>2</sub> reduction.<sup>22,57</sup> Finally, emerging methodologies<sup>58</sup> may soon enable studies of electrified interfaces and their influence on structural properties.

## Data availability

The MLIP models we trained, the input files for training, the databases, and the rough copper surfaces will be made available on Zenodo upon publication under <https://doi.org/10.5281/zenodo.17119744>.

## Acknowledgement

This research was funded in part by the Austrian Science Fund (FWF) 10.55776/COE5 and 10.55776/PAT3175423. The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC). For open access purposes, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. The work leading to this publication was supported by the PRIME programme

of the German Academic Exchange Service (DAAD) with funds from the German Federal Ministry of Research, Technology and Space (BMFTR).

## References

- (1) Calvin, K. et al. *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]*. IPCC, Geneva, Switzerland.; 2023.
- (2) Hori, Y. In *Modern Aspects of Electrochemistry*; Vayenas, C. G., White, R. E., Gamboa-Aldeco, M. E., Eds.; Springer: New York, NY, 2008; pp 89–189.
- (3) Nitopi, S.; Bertheussen, E.; Scott, S. B.; Liu, X.; Engstfeld, A. K.; Horch, S.; Seger, B.; Stephens, I. E. L.; Chan, K.; Hahn, C.; Nørskov, J. K.; Jaramillo, T. F.; Chorkendorff, I. Progress and Perspectives of Electrochemical CO<sub>2</sub> Reduction on Copper in Aqueous Electrolyte. *Chemical Reviews* **2019**, *119*, 7610–7672.
- (4) Yamamoto, S.; Bluhm, H.; Andersson, K.; Ketteler, G.; Ogasawara, H.; Salmeron, M.; Nilsson, A. In Situ X-Ray Photoelectron Spectroscopy Studies of Water on Metals and Oxides at Ambient Conditions. *Journal of Physics: Condensed Matter* **2008**, *20*, 184025.
- (5) Keller, H.; Saracino, M.; Nguyen, H. M.; Huynh, T. M. T.; Broekmann, P. Competitive Anion/Water and Cation/Water Interactions at Electrified Copper/Electrolyte Interfaces Probed by in Situ X-ray Diffraction. *The Journal of Physical Chemistry C* **2012**, *116*, 11068–11076.
- (6) Mistry, K.; Snowden, H.; Darling, G. R.; Hodgson, A. Hydroxyl on Stepped Copper and Its Interaction with Water. *The Journal of Physical Chemistry C* **2024**, *128*, 13025–13033.
- (7) Lee, J.; Sorescu, D. C.; Jordan, K. D.; Yates, J. T. J. Hydroxyl Chain Formation on the Cu(110) Surface: Watching Water Dissociation. *The Journal of Physical Chemistry C* **2008**, *112*, 17672–17677.

- (8) Mehlhorn, M.; Carrasco, J.; Michaelides, A.; Morgenstern, K. Local Investigation of Femtosecond Laser Induced Dynamics of Water Nanoclusters on Cu(111). *Physical Review Letters* **2009**, *103*, 026101.
- (9) Heenen, H. H.; Gauthier, J. A.; Kristoffersen, H. H.; Ludwig, T.; Chan, K. Solvation at Metal/Water Interfaces: An Ab Initio Molecular Dynamics Benchmark of Common Computational Approaches. *The Journal of Chemical Physics* **2020**, *152*, 144703.
- (10) Izvekov, S.; Mazzolo, A.; VanOpdorp, K.; Voth, G. A. Ab Initio Molecular Dynamics Simulation of the Cu(110)–Water Interface. *The Journal of Chemical Physics* **2001**, *114*, 3248–3257.
- (11) Natarajan, S. K.; Behler, J. Neural Network Molecular Dynamics Simulations of Solid–Liquid Interfaces: Water at Low-Index Copper Surfaces. *Physical Chemistry Chemical Physics* **2016**, *18*, 28704–28725.
- (12) Natarajan, S. K.; Behler, J. Self-Diffusion of Surface Defects at Copper–Water Interfaces. *The Journal of Physical Chemistry C* **2017**, *121*, 4368–4383.
- (13) Gäding, J.; Della Balda, V.; Lan, J.; Konrad, J.; Iannuzzi, M.; Meißner, R. H.; Tocci, G. The Role of the Water Contact Layer on Hydration and Transport at Solid/Liquid Interfaces. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2407877121.
- (14) Schörghuber, J.; Bučková, N.; Heid, E.; Madsen, G. K. H. From Flat to Stepped: Active Learning Frameworks for Investigating Local Structure at Copper–Water Interfaces. *Physical Chemistry Chemical Physics* **2025**, *27*, 9169–9177.
- (15) Auer, A.; Sarabia, F. J.; Winkler, D.; Griesser, C.; Climent, V.; Feliu, J. M.; Kunze-Liebhäuser, J. Interfacial Water Structure as a Descriptor for Its Electro-Reduction on Ni(OH)<sub>2</sub>-Modified Cu(111). *ACS Catalysis* **2021**, *11*, 10324–10332.

- (16) Auer, A.; Sarabia, F. J.; Griesser, C.; Climent, V.; Feliu, J. M.; Kunze-Liebhäuser, J. Cu(111) Single Crystal Electrodes: Modifying Interfacial Properties to Tailor Electrocatalysis. *Electrochimica Acta* **2021**, *396*, 139222.
- (17) Simon, G. H.; Kley, C. S.; Roldan Cuenya, B. Potential-Dependent Morphology of Copper Catalysts During CO<sub>2</sub> Electroreduction Revealed by In Situ Atomic Force Microscopy. *Angewandte Chemie International Edition* **2021**, *60*, 2561–2568.
- (18) Cheng, D.; Nguyen, K.-L. C.; Sumaria, V.; Wei, Z.; Zhang, Z.; Gee, W.; Li, Y.; Morales-Guio, C. G.; Heyde, M.; Roldan Cuenya, B.; Alexandrova, A. N.; Sautet, P. Structure Sensitivity and Catalyst Restructuring for CO<sub>2</sub> Electro-reduction on Copper. *Nature Communications* **2025**, *16*, 4064.
- (19) Gunathunge, C. M.; Li, X.; Li, J.; Hicks, R. P.; Ovalle, V. J.; Waegele, M. M. Spectroscopic Observation of Reversible Surface Reconstruction of Copper Electrodes under CO<sub>2</sub> Reduction. *The Journal of Physical Chemistry C* **2017**, *121*, 12337–12344.
- (20) Kim, Y.-G.; Baricuatro, J. H.; Javier, A.; Gregoire, J. M.; Soriaga, M. P. The Evolution of the Polycrystalline Copper Surface, First to Cu(111) and Then to Cu(100), at a Fixed CO<sub>2</sub>RR Potential: A Study by Operando EC-STM. *Langmuir* **2014**, *30*, 15053–15056.
- (21) Huang, J.; Hörmann, N.; Oveisi, E.; Loiudice, A.; De Gregorio, G. L.; Andreussi, O.; Marzari, N.; Buonsanti, R. Potential-Induced Nanoclustering of Metallic Catalysts during Electrochemical CO<sub>2</sub> Reduction. *Nature Communications* **2018**, *9*, 3117.
- (22) Lee, S. H.; Lin, J. C.; Farmand, M.; Landers, A. T.; Feaster, J. T.; Avilés Acosta, J. E.; Beeman, J. W.; Ye, Y.; Yano, J.; Mehta, A.; Davis, R. C.; Jaramillo, T. F.; Hahn, C.; Drisdell, W. S. Oxidation State and Surface Reconstruction of Cu under CO<sub>2</sub> Reduction Conditions from In Situ X-ray Characterization. *Journal of the American Chemical Society* **2021**, *143*, 588–592.

- (23) Amirbeigiarab, R.; Tian, J.; Herzog, A.; Qiu, C.; Bergmann, A.; Roldan Cuenya, B.; Magnussen, O. M. Atomic-Scale Surface Restructuring of Copper Electrodes under CO<sub>2</sub> Electroreduction Conditions. *Nature Catalysis* **2023**, *6*, 837–846.
- (24) Nguyen, K.-L. C.; Bruce, J. P.; Yoon, A.; Navarro, J. J.; Scholten, F.; Landwehr, F.; Rettenmaier, C.; Heyde, M.; Cuenya, B. R. The Influence of Mesoscopic Surface Structure on the Electrocatalytic Selectivity of CO<sub>2</sub> Reduction with UHV-Prepared Cu(111) Single Crystals. *ACS Energy Letters* **2024**, *9*, 644–652.
- (25) Jiang, K.; Huang, Y.; Zeng, G.; Toma, F. M.; Goddard, W. A. I.; Bell, A. T. Effects of Surface Roughness on the Electrochemical Reduction of CO<sub>2</sub> over Cu. *ACS Energy Letters* **2020**, *5*, 1206–1214, Publisher: American Chemical Society.
- (26) Ebaid, M.; Jiang, K.; Zhang, Z.; Drisdell, W. S.; Bell, A. T.; Cooper, J. K. Production of C<sub>2</sub>/C<sub>3</sub> Oxygenates from Planar Copper Nitride-Derived Mesoporous Copper via Electrochemical Reduction of CO<sub>2</sub>. *Chemistry of Materials* **2020**, *32*, 3304–3311, Publisher: American Chemical Society.
- (27) Drautz, R. Atomic Cluster Expansion for Accurate and Transferable Interatomic Potentials. *Physical Review B* **2019**, *99*, 014104.
- (28) Thompson, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral Neighbor Analysis Method for Automated Generation of Quantum-Accurate Interatomic Potentials. *Journal of Computational Physics* **2015**, *285*, 316–330.
- (29) Nguyen-Cong, K.; Willman, J. T.; Moore, S. G.; Belonoshko, A. B.; Gayatri, R.; Weinberg, E.; Wood, M. A.; Thompson, A. P.; Oleynik, I. I. Billion Atom Molecular Dynamics Simulations of Carbon at Extreme Conditions and Experimental Time and Length Scales. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. St. Louis Missouri, 2021; pp 1–12.

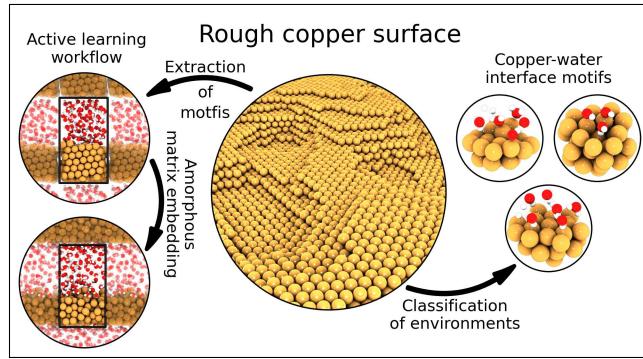
- (30) Erhard, L. C.; Otzen, C.; Rohrer, J.; Prescher, C.; Albe, K. Understanding Phase Transitions of  $\alpha$ -Quartz under Dynamic Compression Conditions by Machine-Learning Driven Atomistic Simulations. *npj Computational Materials* **2025**, *11*, 58.
- (31) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nature Communications* **2022**, *13*, 2453.
- (32) Bochkarev, A.; Lysogorskiy, Y.; Drautz, R. Graph Atomic Cluster Expansion for Semilocal Interactions beyond Equivariant Message Passing. *Physical Review X* **2024**, *14*, 021036.
- (33) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Advances in Neural Information Processing Systems* **2022**, *35*, 11423–11436.
- (34) Leimeroth, N.; Erhard, L. C.; Albe, K.; Rohrer, J. Machine-Learning Interatomic Potentials from a Users Perspective: A Comparison of Accuracy, Speed and Data Efficiency. *Modelling and Simulation in Materials Science and Engineering* **2025**, *33*, 065012.
- (35) Heid, E.; Schörghuber, J.; Wanzenböck, R.; Madsen, G. K. H. Spatially Resolved Uncertainties for Machine Learning Potentials. *Journal of Chemical Information and Modeling* **2024**, *64*, 6377–6387.
- (36) Erhard, L. C.; Rohrer, J.; Albe, K.; Deringer, V. L. Modelling Atomic and Nanoscale Structure in the Silicon–Oxygen System through Active Machine Learning. *Nature Communications* **2024**, *15*, 1927.
- (37) Lysogorskiy, Y.; Bochkarev, A.; Mrovec, M.; Drautz, R. Active Learning Strategies for Atomic Cluster Expansion Models. *Physical Review Materials* **2023**, *7*, 043801.

- (38) Kong, L.; Li, J.; Sun, L.; Yang, H.; Hao, H.; Chen, C.; Artrith, N.; Torres, J. A. G.; Lu, Z.; Zhou, Y. Overcoming the Size Limit of First Principles Molecular Dynamics Simulations with an In-Distribution Substructure Embedding Active Learner. 2023.
- (39) Hodapp, M.; Shapeev, A. In Operando Active Learning of Interatomic Interaction during Large-Scale Simulations. *Machine Learning: Science and Technology* **2020**, *1*, 045005.
- (40) Kývala, L.; Montero de Hijes, P.; Dellago, C. Unsupervised Identification of Crystal Defects from Atomistic Potential Descriptors. *npj Computational Materials* **2025**, *11*, 50.
- (41) Cioni, M.; Polino, D.; Rapetti, D.; Pesce, L.; Delle Piane, M.; Pavan, G. M. Innate Dynamics and Identity Crisis of a Metal Surface Unveiled by Machine Learning of Atomic Environments. *The Journal of Chemical Physics* **2023**, *158*, 124701.
- (42) Hjorth Larsen, A. et al. The Atomic Simulation Environment—a Python Library for Working with Atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.
- (43) Stukowski, A. Visualization and Analysis of Atomistic Simulation Data with OVITO—the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering* **2010**, *18*, 015012.
- (44) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Computer Physics Communications* **2022**, *271*, 108171.
- (45) Leimeroth, N.; Rohrer, J.; Albe, K. General Purpose Potential for Glassy and Crystalline Phases of Cu-Zr Alloys Based on the ACE Formalism. *Physical Review Materials* **2024**, *8*, 043602.

- (46) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Physical Review B* **1993**, *47*, 558–561.
- (47) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Computational Materials Science* **1996**, *6*, 15–50.
- (48) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Physical Review B* **1996**, *54*, 11169–11186.
- (49) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved Adsorption Energetics within Density-Functional Theory Using Revised Perdew-Burke-Ernzerhof Functionals. *Physical Review B* **1999**, *59*, 7413–7421.
- (50) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.
- (51) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Physical Review B* **1999**, *59*, 1758–1775.
- (52) Lysogorskiy, Y.; van der Oord, C.; Bochkarev, A.; Menon, S.; Rinaldi, M.; Hammer-schmidt, T.; Mrovec, M.; Thompson, A.; Csányi, G.; Ortner, C.; Drautz, R. Performant Implementation of the Atomic Cluster Expansion (PACE) and Application to Copper and Silicon. *npj Computational Materials* **2021**, *7*, 97.
- (53) Bochkarev, A.; Lysogorskiy, Y.; Menon, S.; Qamar, M.; Mrovec, M.; Drautz, R. Efficient Parametrization of the Atomic Cluster Expansion. *Physical Review Materials* **2022**, *6*, 013804.
- (54) Kocer, E.; Mason, J. K.; Erturk, H. Continuous and Optimally Complete Description

- of Chemical Environments Using Spherical Bessel Descriptors. *AIP Advances* **2020**, *10*, 015021.
- (55) Montes-Campos, H.; Carrete, J.; Bichelmaier, S.; Varela, L. M.; Madsen, G. K. H. A Differentiable Neural-Network Force Field for Ionic Liquids. *Journal of Chemical Information and Modeling* **2022**, *62*, 88–101.
- (56) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **2018**, *3*, 861.
- (57) Favaro, M.; Xiao, H.; Cheng, T.; Goddard, W. A.; Yano, J.; Crumlin, E. J. Subsurface oxide plays a critical role in CO<sub>2</sub> activation by Cu(111) surfaces to form chemisorbed CO<sub>2</sub>, the first step in reduction of CO<sub>2</sub>. *Proceedings of the National Academy of Sciences* **2017**, *114*, 6706–6711.
- (58) Bergmann, N.; Bonnet, N.; Marzari, N.; Reuter, K.; Hörmann, N. G. Machine Learning the Energetics of Electrified Solid/Liquid Interfaces. 2025; <https://arxiv.org/abs/2505.19745>.

# TOC Graphic



**Supporting Information:**

**How Realistic are Idealized Copper Surfaces?**

**A Machine Learning Study of Rough**

**Copper-Water Interfaces**

Linus C. Erhard,<sup>\*,†,‡</sup> Johannes Schörghuber,<sup>†,‡</sup> Aleix Comas-Vives,<sup>†</sup> and Georg K.

H. Madsen<sup>\*,†</sup>

<sup>†</sup>Institute of Materials Chemistry, TU Wien, A-1060 Vienna, Austria

<sup>‡</sup>These authors contributed equally

E-mail: linus.erhard@tuwien.ac.at; georg.madsen@tuwien.ac.at

**Table S1: Detailed simulation protocol to produce rough copper surfaces by nanoparticles placed on top.** We used two different temperature protocols for producing rough surfaces. In version 1, we used lower temperatures of maximum 1100 K, but therefore longer simulation times. In version 2, we used higher temperatures of a maximum of 1300 K, but for much shorter times. The temperature of the bottom part of the slab was always fixed at 500 K. In an intermediated region of 10 Å we used starting from the second step a NVE ensemble to allow for a temperature gradient.

Protocol	Simulation Time [ps]	Temperature [K] Top Layer	Temperature [K] Bottom Layer	Box Change Allowed in Surface Plane
<b>Version 1</b>				
Step 1.	20	500	500	Yes
Step 2.	200	500 → 1100	500	No
Step 3.	400	1100	500	No
Step 4.	4000	1100 → 500	500	No
<b>Version 2</b>				
Step 1.	20	500	500	Yes
Step 2.	20	500 → 1300	500	No
Step 3.	10	1300	500	No
Step 4.	20	1300 → 900	500	No
Step 5.	1000	900 → 500	500	No

**Table S2: Detailed simulation protocol to produce rough copper surfaces by inserting indenters.** The whole simulation is performed at constant volume conditions. We start by equilibrating the system and then increasing the temperature of the top layer. Between the top layer and the bottom layer is a 10 Å thick region, where we do not apply a thermostat. The roughness of the surface is then induced by indenters inserted from above the slab into the top layer. The indenters are implemented by dummy particles, which interact by a purely repulsive Lennard-Jones potential with the copper. The radius of the indenters, as well as the intrusion depth, is varied randomly for different structures. We apply a reflecting wall at the bottom and well above the top of the slab to prevent atoms from escaping. After inserting, we keep the indenters at a constant position in the simulation, while we cool down the top layer.

Protocol	Simulation Time [ps]	Temperature	Temperature	Indenters	
		[K] Top Layer	[K] Bottom Layer	Active	Moving Down
Step 1.	10	500	500	No	No
Step 2.	40	500 → 1500	500	No	No
Step 3.	200	1500	500	Yes	Yes
Step 4.	40	1500	500	Yes	No
Step 5.	2000	1500 → 500	500	Yes	No

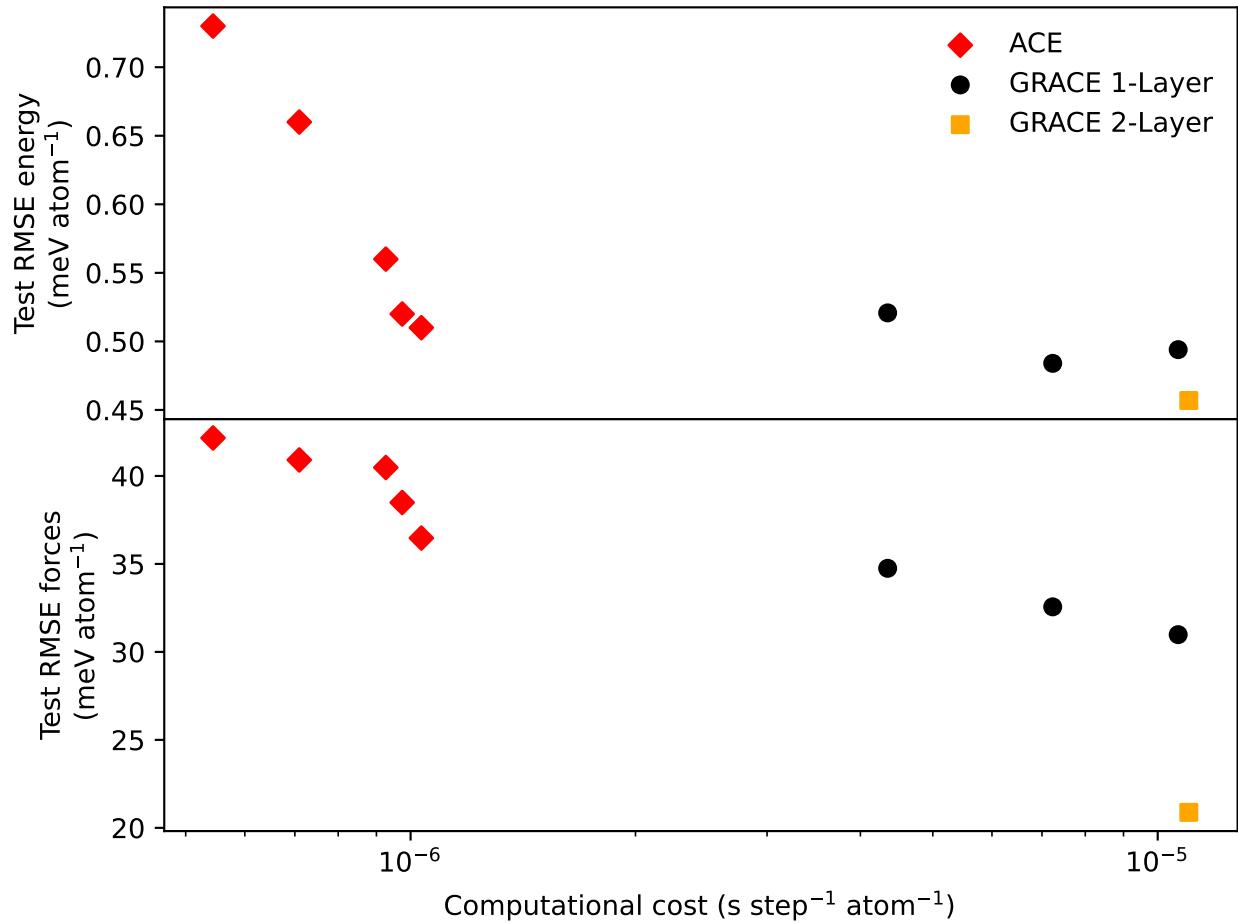
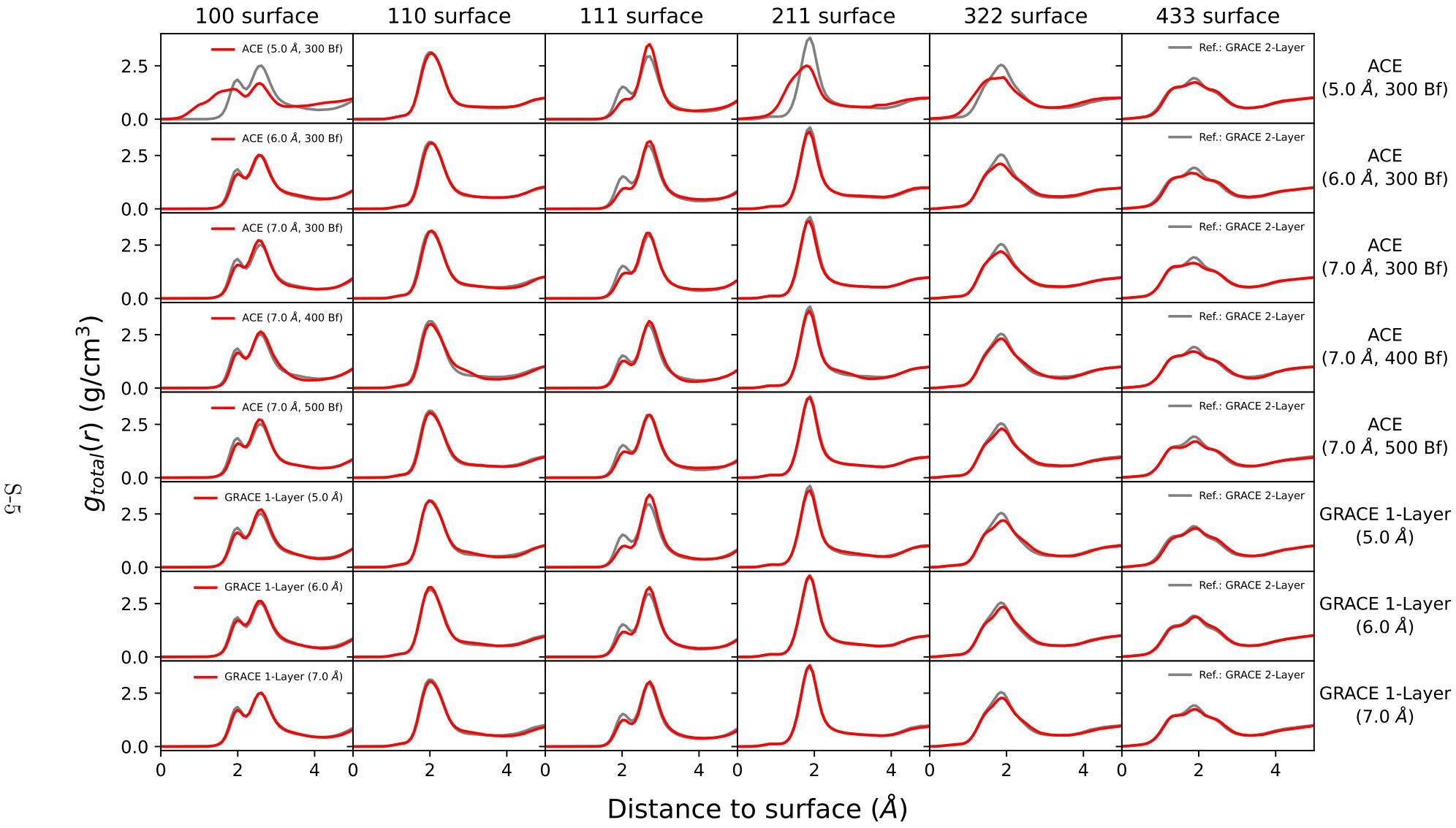
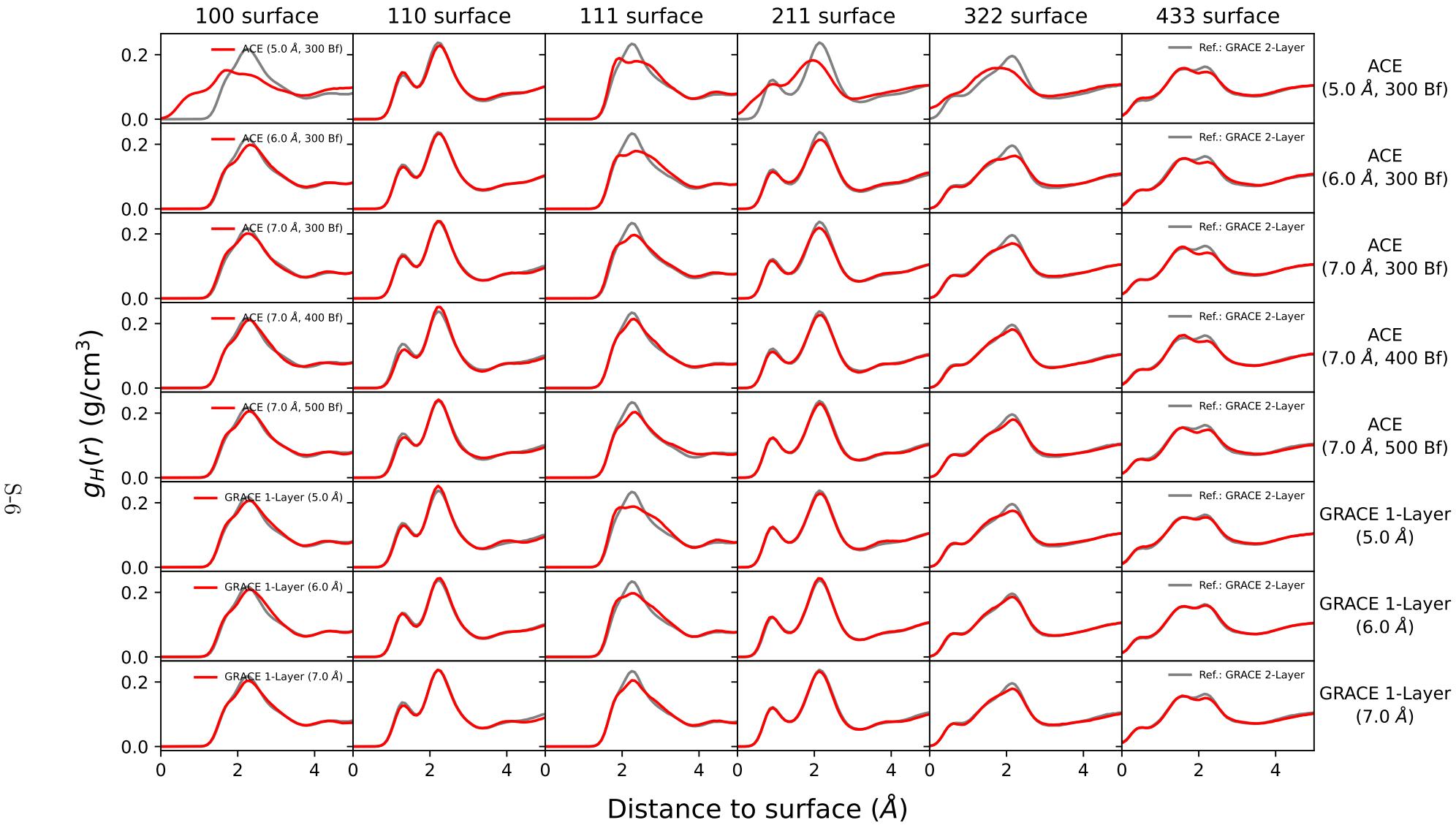


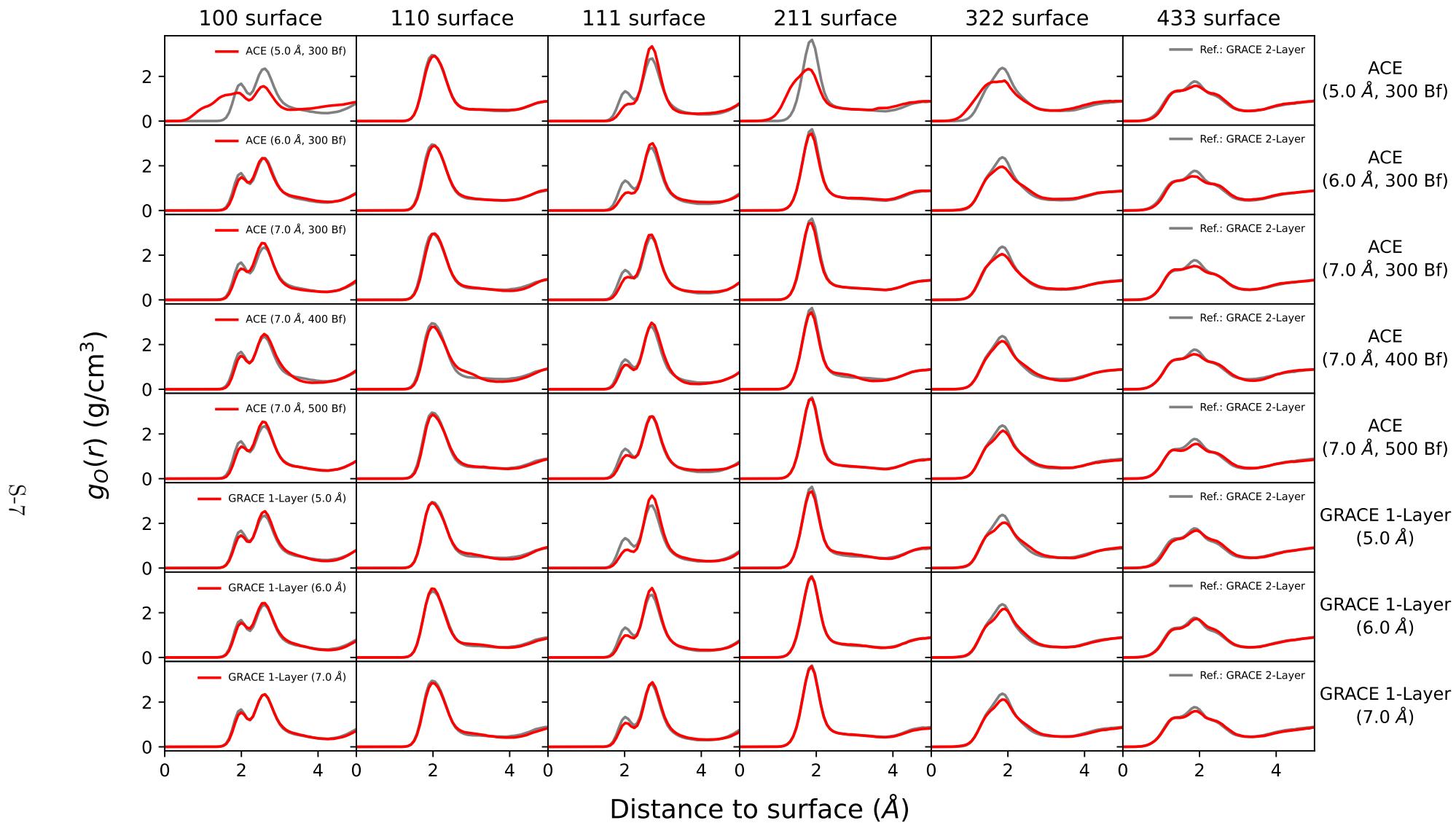
Figure S1: **Accuracy and computational costs of ACE and different types of GRACE.** ACE and GRACE 1-Layer are compared with regard to their accuracy and their computational efficiency to GRACE 2-Layer. In the case of ACE, we varied the cutoff radius and the number of basis functions, in the case of GRACE 1-Layer, we varied only the cutoff radius, leading to several points in the plot.



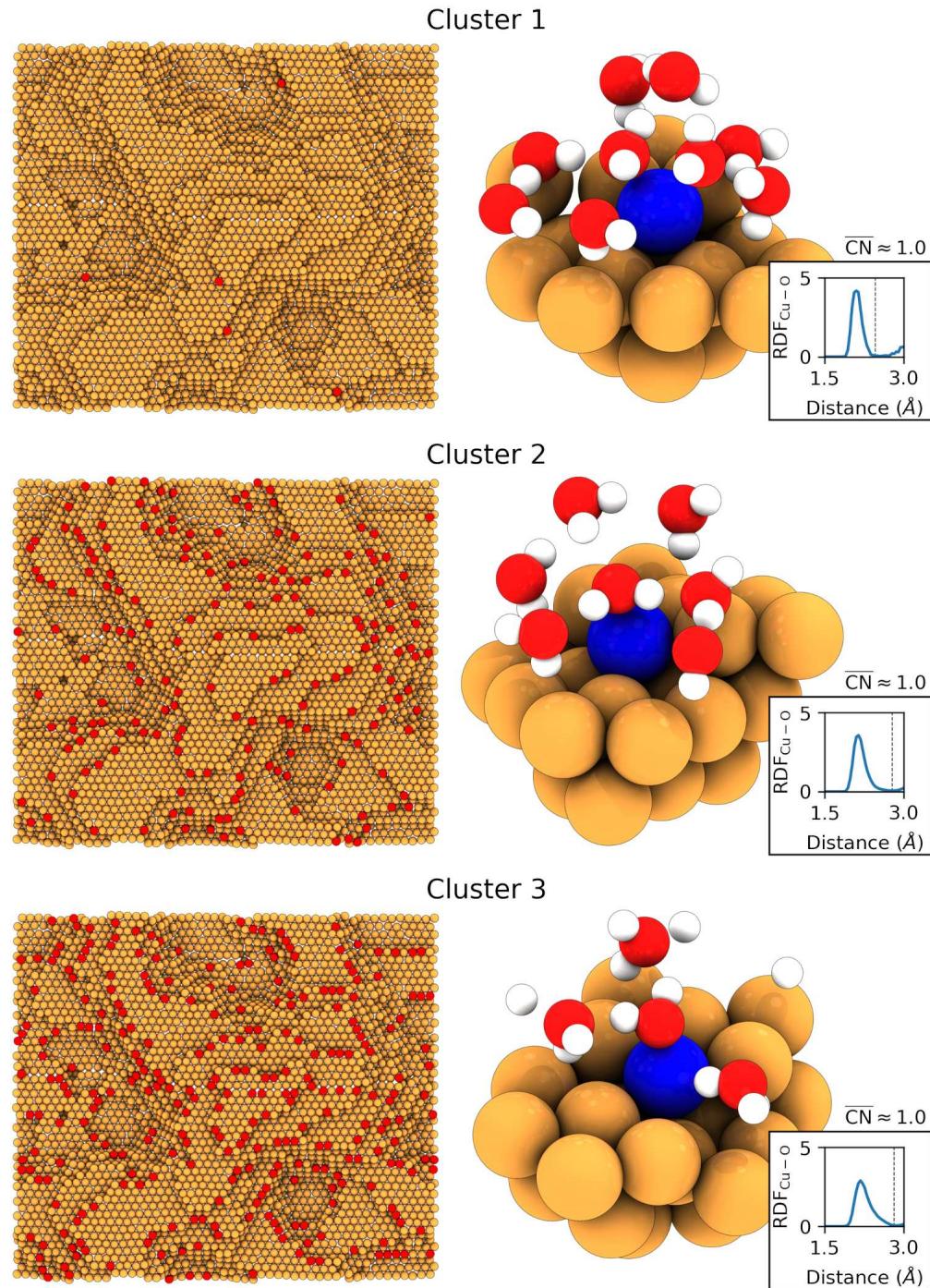
**Figure S2: Density of water as a function of the distance to various copper surfaces.** These plots show how water is distributed as a function of the distances on the copper 100, 110, 111, 211, 322, and 433 surfaces. We evaluated the curves with various machine-learning interatomic potentials, e.g., ACE, GRACE 1-Layer, and GRACE 2-Layer. All potentials have been fitted to the same databases. As GRACE 2-Layer provides the highest accuracy (see Figure S1), we used it in all plots as a reference line (gray).



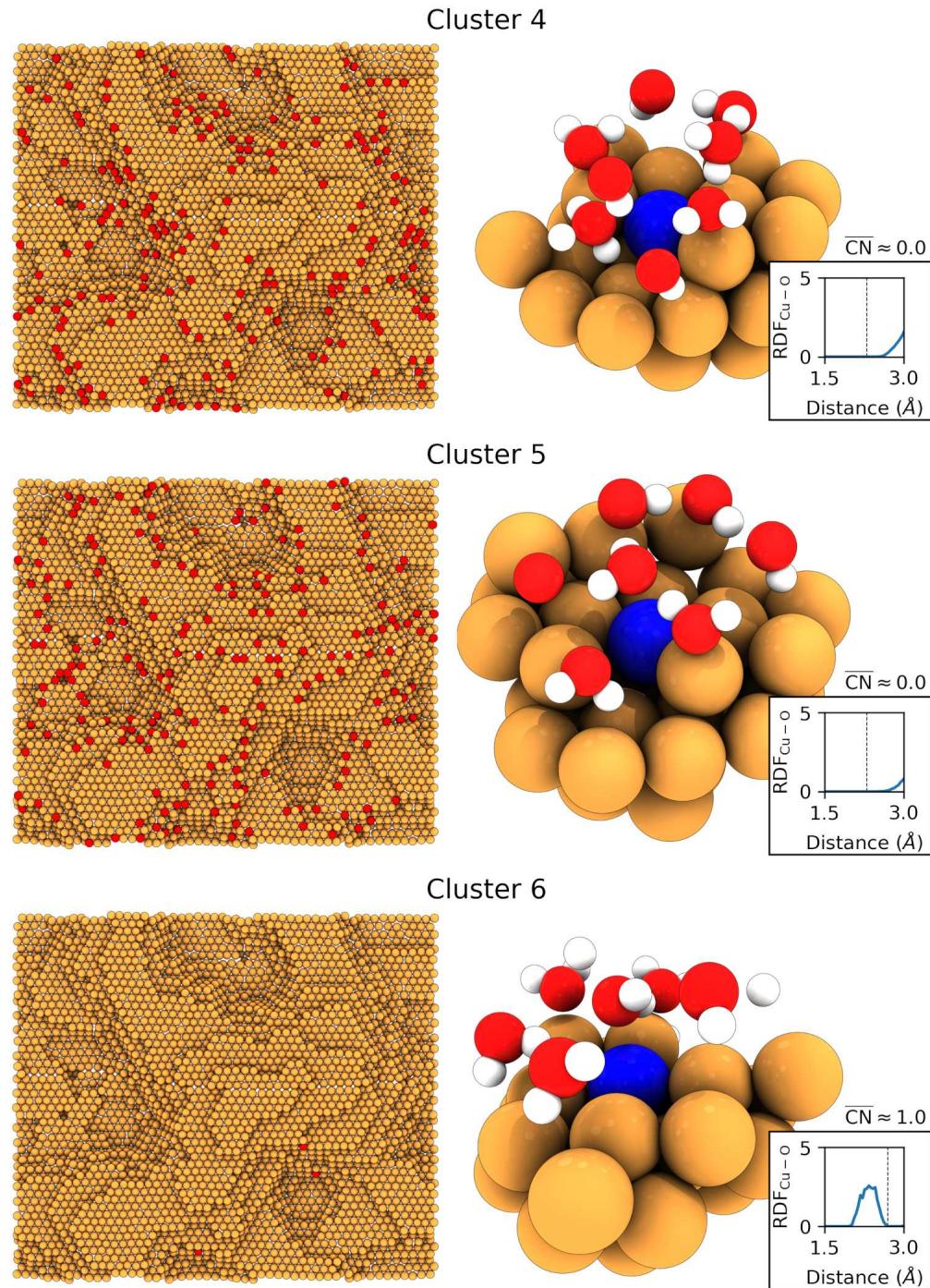
**Figure S3: Density of hydrogen in water as a function of the distance to various copper surfaces.** These plots show how hydrogen, as part of the water molecules, is distributed as a function of the distances on the copper 100, 110, 111, 211, 322, and 433 surfaces. We evaluated the curves with various machine-learning interatomic potentials, e.g., ACE, GRACE 1-Layer, and GRACE 2-Layer. All potentials have been fitted to the same databases. As GRACE 2-Layer provides the highest accuracy (see Figure S1), we used it in all plots as a reference line (gray).



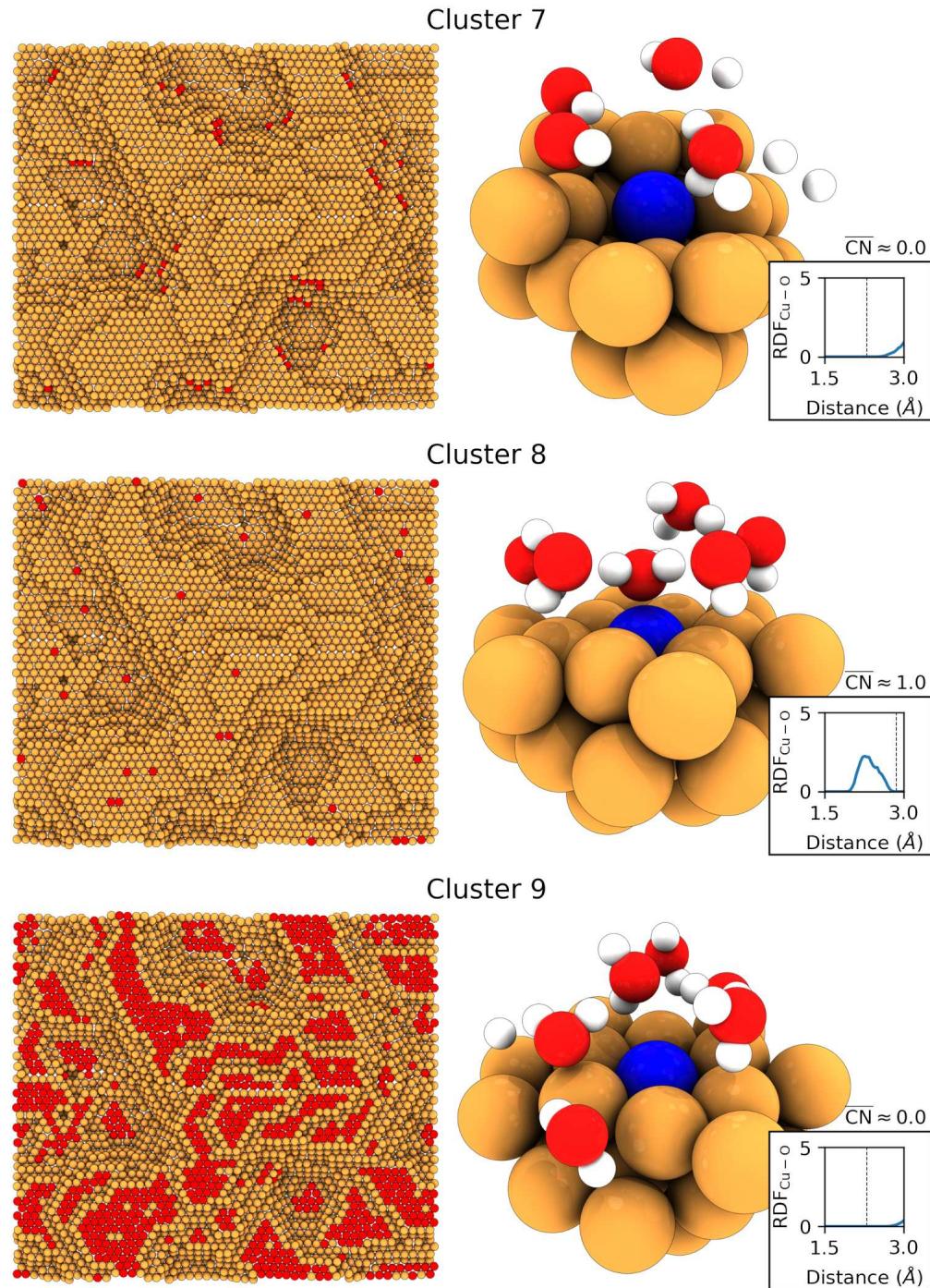
**Figure S4: Density of oxygen in water as a function of the distance to various copper surfaces.** These plots show how oxygen, as part of water molecules, is distributed as a function of the distances on the copper 100, 110, 111, 211, 322, and 433 surfaces. We evaluated the curves with various machine-learning interatomic potentials, e.g., ACE, GRACE 1-Layer, and GRACE 2-Layer. All potentials have been fitted to the same databases. As GRACE 2-Layer provides the highest accuracy (see Figure S1), we used it in all plots as a reference line (gray).



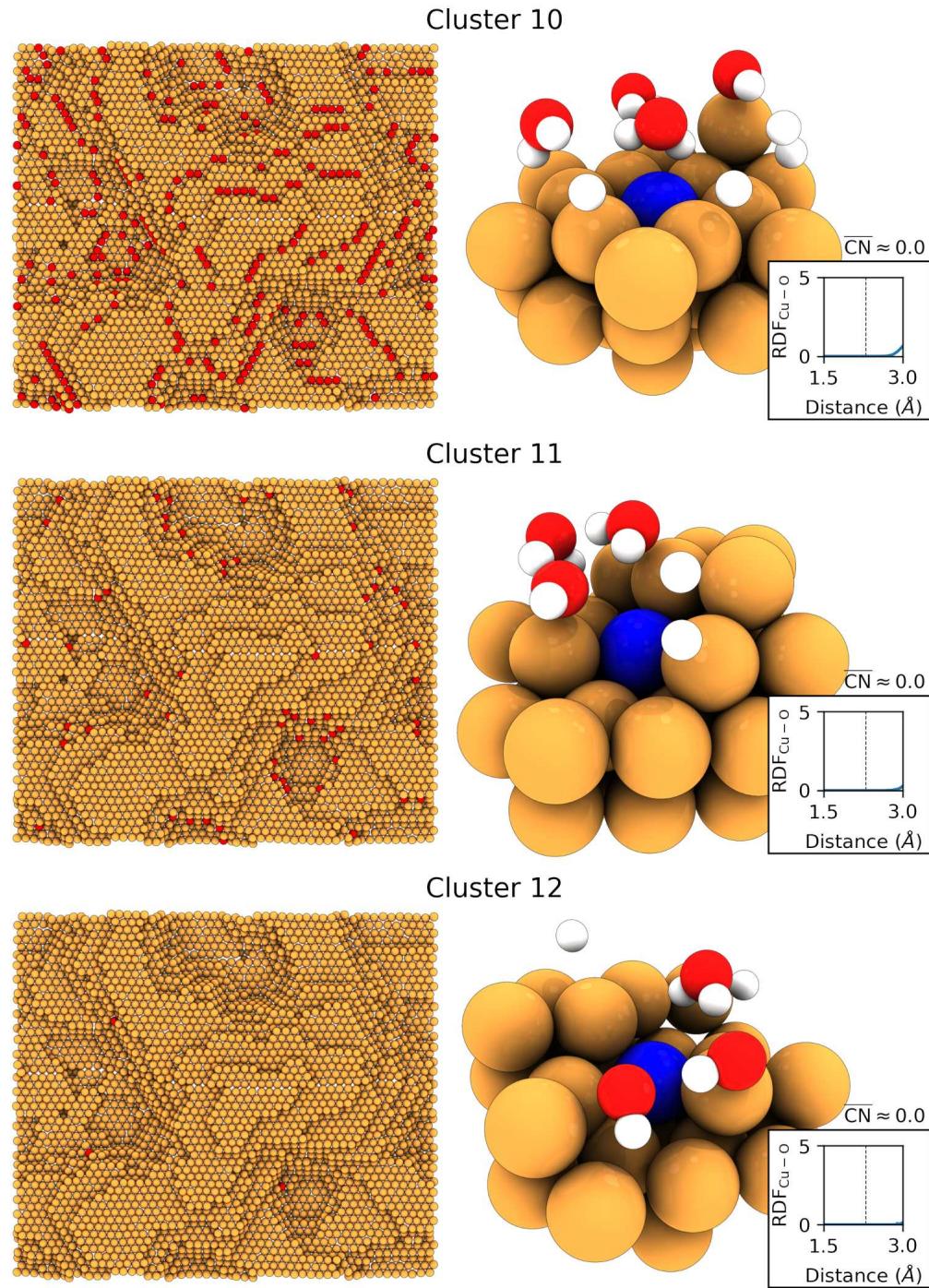
**Figure S5: Rough copper surface with atoms marked according to their cluster classification.** The left side of the figure depicts the rough copper surface, with atoms colored according to their cluster assignment as specified in Figure 6. On the right, an exemplary environment within a 5 Å radius of a cluster atom (marked in blue) is shown. In this representation, copper atoms are colored brown, oxygen atoms are colored red, and hydrogen atoms are colored white. Additionally, the average radial distribution function between copper atoms in this cluster and oxygen atoms is plotted. The average coordination number of copper with respect to oxygen is determined by integrating the first peak of the radial distribution function up to the point indicated by the black line.



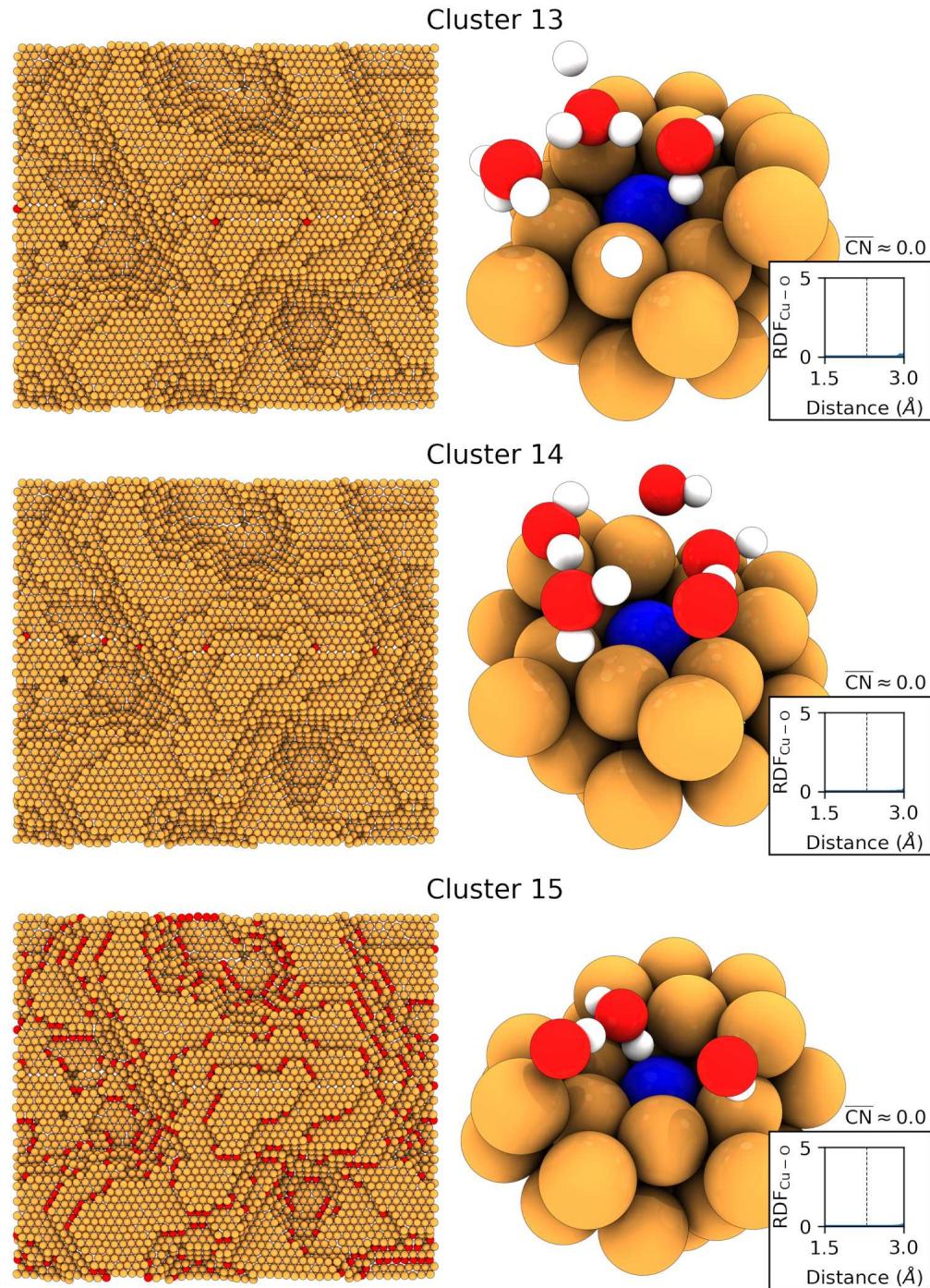
**Figure S6: Rough copper surface with atoms marked according to their cluster classification.** The left side of the figure depicts the rough copper surface, with atoms colored according to their cluster assignment as specified in Figure 6. On the right, an exemplary environment within a 5 Å radius of a cluster atom (marked in blue) is shown. In this representation, copper atoms are colored brown, oxygen atoms are colored red, and hydrogen atoms are colored white. Additionally, the average radial distribution function between copper atoms in this cluster and oxygen atoms is plotted. The average coordination number of copper with respect to oxygen is determined by integrating the first peak of the radial distribution function up to the point indicated by the black line.



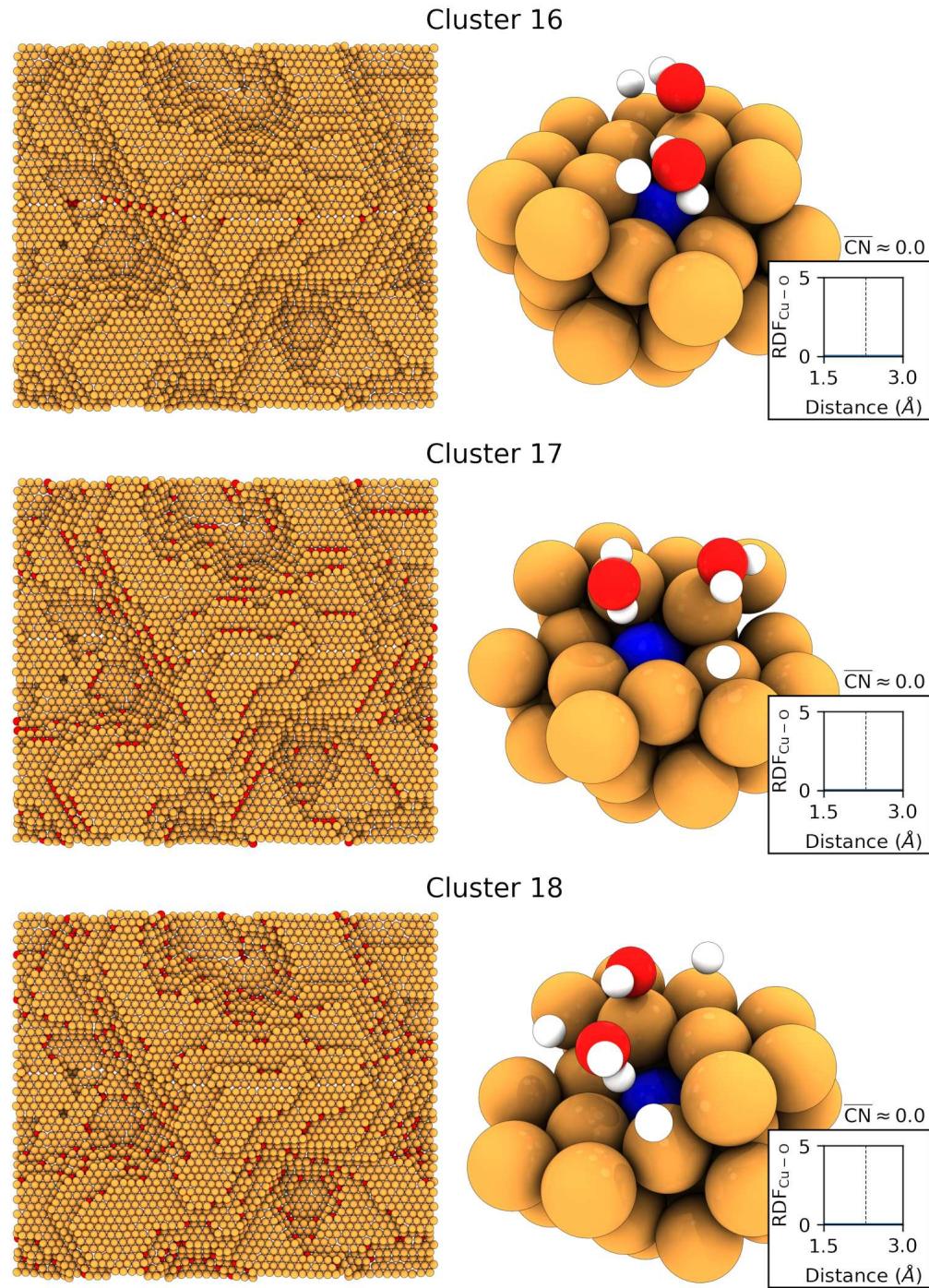
**Figure S7: Rough copper surface with atoms marked according to their cluster classification.** The left side of the figure depicts the rough copper surface, with atoms colored according to their cluster assignment as specified in Figure 6. On the right, an exemplary environment within a 5 Å radius of a cluster atom (marked in blue) is shown. In this representation, copper atoms are colored brown, oxygen atoms are colored red, and hydrogen atoms are colored white. Additionally, the average radial distribution function between copper atoms in this cluster and oxygen atoms is plotted. The average coordination number of copper with respect to oxygen is determined by integrating the first peak of the radial distribution function up to the point indicated by the black line.



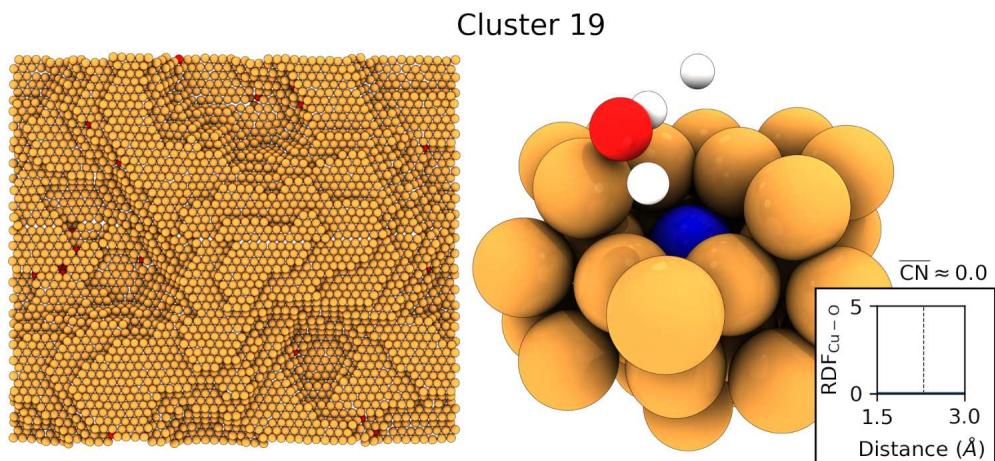
**Figure S8: Rough copper surface with atoms marked according to their cluster classification.** The left side of the figure depicts the rough copper surface, with atoms colored according to their cluster assignment as specified in Figure 6. On the right, an exemplary environment within a 5 Å radius of a cluster atom (marked in blue) is shown. In this representation, copper atoms are colored brown, oxygen atoms are colored red, and hydrogen atoms are colored white. Additionally, the average radial distribution function between copper atoms in this cluster and oxygen atoms is plotted. The average coordination number of copper with respect to oxygen is determined by integrating the first peak of the radial distribution function up to the point indicated by the black line.



**Figure S9: Rough copper surface with atoms marked according to their cluster classification.** The left side of the figure depicts the rough copper surface, with atoms colored according to their cluster assignment as specified in Figure 6. On the right, an exemplary environment within a 5 Å radius of a cluster atom (marked in blue) is shown. In this representation, copper atoms are colored brown, oxygen atoms are colored red, and hydrogen atoms are colored white. Additionally, the average radial distribution function between copper atoms in this cluster and oxygen atoms is plotted. The average coordination number of copper with respect to oxygen is determined by integrating the first peak of the radial distribution function up to the point indicated by the black line.



**Figure S10: Rough copper surface with atoms marked according to their cluster classification.** The left side of the figure depicts the rough copper surface, with atoms colored according to their cluster assignment as specified in Figure 6. On the right, an exemplary environment within a 5 Å radius of a cluster atom (marked in blue) is shown. In this representation, copper atoms are colored brown, oxygen atoms are colored red, and hydrogen atoms are colored white. Additionally, the average radial distribution function between copper atoms in this cluster and oxygen atoms is plotted. The average coordination number of copper with respect to oxygen is determined by integrating the first peak of the radial distribution function up to the point indicated by the black line.



**Figure S11: Rough copper surface with atoms marked according to their cluster classification.** The left side of the figure depicts the rough copper surface, with atoms colored according to their cluster assignment as specified in Figure 6. On the right, an exemplary environment within a 5 Å radius of a cluster atom (marked in blue) is shown. In this representation, copper atoms are colored brown, oxygen atoms are colored red, and hydrogen atoms are colored white. Additionally, the average radial distribution function between copper atoms in this cluster and oxygen atoms is plotted. The average coordination number of copper with respect to oxygen is determined by integrating the first peak of the radial distribution function up to the point indicated by the black line.

Table S3: **Cluster classification.** Classification of all clusters from Figure 6 in the corresponding parts on the rough copper surface. For each cluster, we also note whether there is water chemisorbed.

Cluster	Position	Water chemisorbed
#1	Corner of a (111) plane with two neighbors in the same plane.	Yes.
#2	Corner of a (111) plane with three neighbors in the same plane.	Yes.
#3	Edge of a (111) plane with four neighbors in the same plane. Partially atoms at the stacking fault.	Yes.
#4	Edge of a (111) plane with four neighbors in the same plane.	No.
#5	Edge of a (111) plane with five neighbors (nearly completely surrounded by copper atoms).	No.
#6	(100) surface like facets.	Yes.
#7	(100) surface like facets.	No.
#8	(111) surface.	Yes.
#9	(111) surface.	No.
#10	(111) surface not directly below an edge, but close to.	No.
#11	(100) surface like facets directly below edges.	No.
#12	Rather complex edge/corner arrangements.	No.
#13	Directly below edges at the stacking fault.	No.
#14	Directly below edges at the stacking fault.	No.
#15	Directly below edges at (111) surface facets.	No.
#16	Closely below surface at stacking fault.	No.
#17	Directly below edge of a (111) surface facet.	No.
#18	Below (111) surface corners.	No.
#19	In small holes nearly below the surface.	No.