POLITECNICO DI TORINO
Repository ISTITUZIONALE

Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments

Original
Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments / Cioni, Matteo; Polino, Daniela; Rapetti, Daniele; Pesce, Luca; DELLE PIANE, Massimo; Pavan, Giovanni M.. - In: THE JOURNAL OF CHEMICAL PHYSICS. - ISSN 0021-9606. - 158:12(2023), pp. 1-16. [10.1063/5.0139010]

Availability:
This version is available at: 11583/2977388 since: 2023-05-29T07:55:40Z

Publisher:
AIP

Published
DOI:10.1063/5.0139010

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

10 April 2024

The Journal of Chemical Physics

Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments

Cite as: J. Chem. Phys. **158**, 124701 (2023); https://doi.org/10.1063/5.0139010
Submitted: 16 December 2022 • Accepted: 06 March 2023 • Accepted Manuscript Online: 06 March 2023 • Published Online: 22 March 2023

Matteo Cioni, Daniela Polino, Daniele Rapetti, et al.

COLLECTIONS

Paper published as part of the special topic on Machine Learning Hits Molecular Simulations

openaccess

View Online

Export Citation

CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Phase diagrams—Why they matter and how to predict them
The Journal of Chemical Physics **158**, 030902 (2023); https://doi.org/10.1063/5.0131028

Computing chemical potentials of solutions from structure factors
The Journal of Chemical Physics **157**, 121101 (2022); https://doi.org/10.1063/5.0107059

Learning pair potentials using differentiable simulations
The Journal of Chemical Physics **158**, 044113 (2023); https://doi.org/10.1063/5.0126475

Time to get excited.
Lock-in Amplifiers – from DC to 8.5 GHz
Find out more
Zurich Instruments

AIP Publishing

© 2023 Author(s).

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments

Cite as: J. Chem. Phys. 158, 124701 (2023); doi: 10.1063/5.0139010
Submitted: 16 December 2022 • Accepted: 6 March 2023 •
Published Online: 22 March 2023

Matteo Cioni,[1] **Daniela Polino,**[2] **Daniele Rapetti,**[1] **Luca Pesce,**[2] **Massimo Delle Piane,**[1] **and Giovanni M. Pavan**[1,2,a)]

AFFILIATIONS

[1] Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[2] Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Polo Universitario Lugano, Campus Est, Via la Santa 1, 6962 Lugano-Viganello, Switzerland

Note: This paper is part of the JCP Special Topic on Machine Learning Hits Molecular Simulations.

[a)] **Author to whom correspondence should be addressed:** giovanni.pavan@polito.it

ABSTRACT

Metals are traditionally considered hard matter. However, it is well known that their atomic lattices may become dynamic and undergo reconfigurations even well below the melting temperature. The innate atomic dynamics of metals is directly related to their bulk and surface properties. Understanding their complex structural dynamics is, thus, important for many applications but is not easy. Here, we report deep-potential molecular dynamics simulations allowing to resolve at an atomic resolution the complex dynamics of various types of copper (Cu) surfaces, used as an example, near the Hüttig (~1/3 of melting) temperature. The development of deep neural network potential trained on density functional theory calculations provides a dynamically accurate force field that we use to simulate large atomistic models of different Cu surface types. A combination of high-dimensional structural descriptors and unsupervized machine learning allows identifying and tracking all the atomic environments (AEs) emerging in the surfaces at finite temperatures. We can directly observe how AEs that are non-native in a specific (ideal) surface, but that are, instead, typical of other surface types, continuously emerge/disappear in that surface in relevant regimes in dynamic equilibrium with the native ones. Our analyses allow estimating the lifetime of all the AEs populating these Cu surfaces and to reconstruct their dynamic interconversions networks. This reveals the elusive identity of these metal surfaces, which preserve their identity only in part and in part transform into something else under relevant conditions. This also proposes a concept of "statistical identity" for metal surfaces, which is key to understanding their behaviors and properties.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0139010

I. INTRODUCTION

Metals are traditionally considered a reference material class in hard matter. Their structure, characterized by crystalline solid lattices with well-defined order and symmetries,[1] imparts them a variety of technologically relevant properties. Nonetheless, the dynamics of metals is perhaps as much interesting but also far more elusive.

It is known that metals may assume a non-trivial dynamic behavior well below their melting temperature. Above the Hüttig temperature (~1/3 of the melting temperature), their atomic surface enters a dynamic equilibrium in which the atoms may move in the lattice.[2–4]

In particular, it has been experimentally observed that above this temperature, metal surfaces may undergo different structural transformations, such as "deconstruction," "preroughening," faceting, roughening, and surface premelting. However, although a large number of experimental and theoretical studies focused on these phenomena in the past,[5–8] a general understanding is still far from complete. Nonetheless, the appearance or disappearance of specific atomic environments and their dynamic evolution on a metal surface may be crucial in determining properties that are directly related to the structural features of their surface, such as e.g., their reactivity in heterogeneous catalysis.[9,10]

Obtaining an atomic-level insight into the dynamics of metals is key to understanding their properties but is non-trivial. Widely used for studying metals,[11–18] computational modeling holds considerable potential in this sense. However, in several computational approaches, such as in heterogeneous catalysis studies, metal surfaces are usually treated as static.[14–18] Accounting for the atomic dynamics of metal lattices, nevertheless, becomes crucial in those conditions where this is a determinant for the material's properties. As a notable example, Gazzarrini *et al.* demonstrated how atoms' mobility in copper (Cu) nanoparticles may produce variations in the number of vertex, edge, and face atoms, affecting the nanoparticles' efficiency in catalyzing CO_2 conversion to methane.[19] Nelli *et al.* studied the dynamic diffusion of atomic impurities in copper–cobalt icosahedral nanoparticles via metadynamics simulations.[20] These works offer important preliminary evidence of how metal lattices cannot be simply studied as static materials, but in certain regimes, these are rather complex systems where atoms are in dynamic equilibrium and continuous exchange. Intriguingly, this is close to what it has been recently seen in soft self-assembling systems.[21–24]

While disentangling such complex structural dynamics is non-trivial, machine learning (ML) is useful to this end. On one hand, ML potentials trained on quantum mechanical (QM) data offer the opportunity to obtain accurate force fields allowing to reliably simulate metals on relevant spatiotemporal scales. Since the pioneering work of Behler and Parrinello[25] that introduced the concept of high-dimensional neural network potentials, different approaches to build ML potentials have been developed in the last decade,[26–33] allowing to investigate, with *ab initio* accuracy, systems of increasing complexity (we refer the reader to Ref. 34 for a description of various widely used methods). On the other hand, high-dimensional descriptors and advanced statistical analyses are extremely useful to unravel the complexity of dynamic molecular systems. For example, unsupervized clustering of smooth overlap of atomic position (SOAP) data[35–37] extracted from molecular dynamics (MD) simulations recently allowed to reconstruct the structural/dynamical complexity of a variety of molecular materials/systems[22,23,38] and to build robust data-driven metrics[37,39] for their classification.[23] Similar data-driven approaches have been also recently used to explore the mechanism of gold nanoparticles (NPs) melting.[40] In this work, we report a data-driven approach that allows resolving, at an atomistic resolution, the complex dynamics of metal surfaces under relevant conditions. As an example of a dynamic metal widely used for technological applications, we focus on Cu (the approach is nonetheless versatile and applicable to a variety of metals). An extensive campaign of density functional theory (DFT)[41] MD simulations generated Cu surface configurations and interaction energy data that we used to train a deep neural network (*NN*) potential.[31] This provided a dynamically accurate atomistic force field allowing to simulate various types of Cu surfaces near the Hüttig temperature on relevant spatiotemporal scales. Bottom-up and top-down SOAP-based analyses reveal the atomic environments that statistically populate Cu surfaces, including non-native atomic environments, typical of other types of surfaces, which continuously emerge/resorb in the various Cu surfaces in equilibrium conditions. This ultimately allows estimating an "equivalent identity" for such metal surfaces on a purely statistical basis from the data emerging along the deep-potential molecular dynamics (DPMD) simulations, changing the way we look at such materials.

II. RESULTS

A. Modeling the dynamics of copper surfaces

Here, we use Cu as an example of a dynamic metal widely used for various applications.[42–46] We focus on the study of Cu surface models at 500–700 K, just above the Hüttig temperature (447 K for Cu). Noteworthy, although this is not central in this work, such conditions are of interest, e.g., for catalytic applications (CO_2 hydrogenation),[47,48] where the structural/dynamical features of the Cu surface are important.

Studying the structural dynamics of metal surfaces requires sufficient accuracy in the treatment of the Cu–Cu interactions and, at the same time, one needs to simulate these atomic systems on sufficiently large spatiotemporal scales to prevent finite-size effects and guarantee that microscopic dynamic transitions (and not only vibrations) are sampled with sufficient statistics. Accurate DFT[41] calculations are limited to timescales and sizes that are too short/small to this end, whereas classical atomistic force fields[49–54] may not guarantee sufficient accuracy in the treatment of the structural dynamics of the metal surface. To obtain a dynamically accurate atomistic force field, we turned to ML. In particular, we developed an inter-atomic interaction potential for Cu surfaces via training a deep *NN* on data extracted from DFT calculations adopting the Deep Potential Molecular Dynamics (DeePMD) scheme developed by Wang *et al.*[31,55] [Fig. 1(a)].

A campaign of DFT MD simulations of small Cu FCC bulk, (100), (110), (111), (211), and (210) surface environments conducted at temperatures between 500 and 700 K [Fig. 1(b)] provided a rich dataset of atomic configurations, forces, and energies used to train a first guess *NN* potential. An active (iterative) learning strategy[56,57] was then adopted to ensure a good sampling of the reconstructed configurations in the Cu surface and the local atomic transitions leading to such reconstructions. It is important to note that, although the surface portions simulated at the DFT level used for the training [Fig. 1(b): ~100 atoms, depending on the surface type—sufficient to account up to the ~4 – 5 atom neighbors] are smaller than the surfaces that are then simulated at atomistic level [Figs. 1(c) and 1(d): containing ~2400 atoms], the iterative nature of the approach and our tests demonstrate that the obtained potential is robust and reliable (see also Sec. IV for details). In fact, even if the first trained *NN* potential initially contains DFT-level information only on the small ideal surface patches (limited sampling), the new conformations that are then discovered via DPMD simulations using such incomplete *NN* potential are then re-simulated at the DFT level and added to the DFT dataset. Such a process is conducted iteratively, and at each iteration, the DFT data on the newly discovered atomic configurations are added to the training set. The training process ends when no new configurations are discovered in the successive iterations, and the obtained *NN* force field can be thus considered complete (the reduced spatial sampling is compensated by temporal sampling guaranteed by the iterative DPMD simulation scheme).

© Author(s) 2023

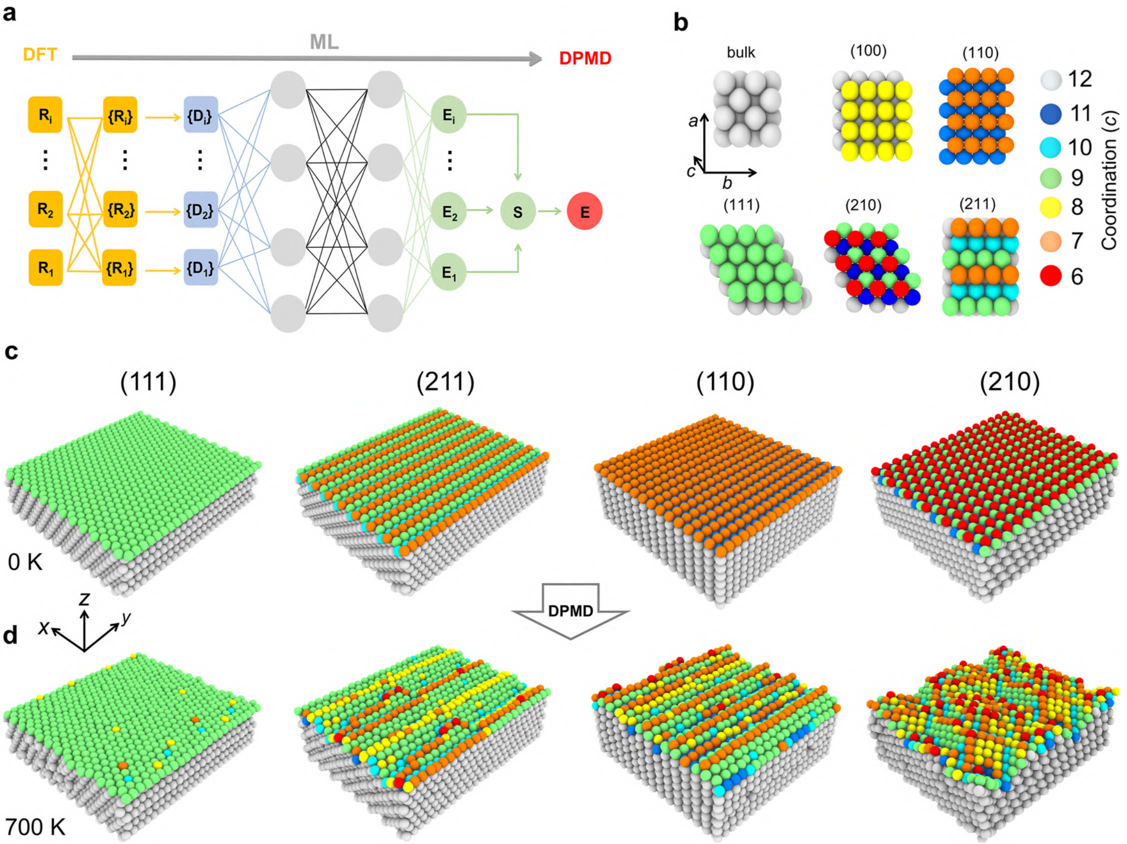

FIG. 1. Atomistic DPMD simulations of Cu surfaces near the Hüttig temperature. (a) Scheme of the deep MD[31] model developed in this work: atomic configurations and energies obtained from DFT calculations have been used to train an inter-atomic *NN* potential for the Cu surfaces. (b) Atomistic models for bulk and surface Cu environments (top view) used for the training (left: atoms are colored based on their coordination). DFT configurations and energies extracted in the temperature range of 500–700 K have been used to train the Cu *NN* potential. (c) Starting ideal configurations of the Cu surface models investigated in this work: we compare the behaviors of (111), (211), (110), and (210) Cu surfaces (atoms colored based on their coordination). (d) Cu surfaces after 150 ns of DPMD simulations at 700 K show structural rearrangements, atomic mobility, and coordination changes (left-to-right, the surfaces are ordered according to their mobility).

To ensure that no residual spurious finite-size effects could affect the trained *NN* force field, we also conducted additional tests using larger surface patches than those shown in Fig. 1(b) (~600 atoms, six times larger). Nonetheless, these tests demonstrated that the maximum deviation in the forces provided by the obtained *NN* potential in the two cases is negligible (estimated forces within the training error in 99.95% of the cases), confirming the robustness and consistency of the obtained Cu *NN* potential. We also note that the discovery-and-sampling approach adopted herein to create the training dataset is based on the collection and sampling of configurations along the DPMD simulations with a time frequency suitable to effectively follow with fine-temporal resolution the motions of the atoms on the surfaces. In particular, the DFT training set contains information not only on the local minima configurations but also on the intermediate configurations and on the transition barrier states that are visited. In this way, the trained *NN* potential has DFT accuracy in reproducing both the energy differences and the transition barriers between the various atomic configurations visited along the DPMD simulations. This provides a structurally and dynamically accurate force field having DFT precision in the treatment of the atomic configurations (energies, forces, etc.) and of their dynamic interconversion within the Cu surfaces (see Sec. IV for complete details).

The *NN* potential was finally validated against some copper bulk and surface properties. In particular, we calculated the lattice constant, the vacancy, and interstitial formation energies of the surface energies at 0 K of the unreconstructed (100), (110), (111), (211), and (210). The values computed, given in Table S1, are in good agreement with the values computed at the DFT level, with the embedded atom model (EAM) of Mendelev *et al.*,[54] and with available experimental values.[58–60] To ascertain the reliability of the *NN* potential, we calculated the surface energy of the (110) (1× 2) missing-row reconstruction at 0 K and found it to be larger than that of the unreconstructed (110) surface. This is in agreement with the experimental evidence that the Cu(110) surface at low temperatures does not undergo the (1 × 2) missing-row reconstruction typical of other noble metals, e.g., Au and Ag. Finally, we computed the diffusion barriers of single adatoms on the (100), (110), and (111). The

© Author(s) 2023

results, given in Table S2, are in good agreement with available theoretical and experimental data.[61–70] This last test validates the ability of the potential to correctly reproduce the dynamic properties of the surface atoms.

We used such *NN* potential to simulate large FCC (111), (211), (110), and (210) surface models composed of 2400 Cu atoms [2304 for (210)] via DPMD simulations. All Cu surface models have depth >15 Å and replicate on the *xy* plane through periodic boundary conditions, effectively modeling a portion of the bulk of infinite surfaces [Fig. 1(c)]. During the DPMD simulations, the two bottom layers in the surface models are kept fixed, while all other atoms are free to move. Such a setup and system sizes prevent finite-size effects and guarantee reliable modeling of the structural dynamics of these Cu surfaces (see Sec. IV for details).

All systems underwent 150 ns of DPMD simulations at 500, 600, and 700 K, the last 75 ns of the simulations being representative of the equilibrium of the surfaces (see the supplementary material, Figs. S2–S4, and Sec. IV). Changes in the atoms' colors in Fig. 1(d) indicate changes in atomic coordination, atomic movements, and reconfigurations along the DPMD simulations. The DPMD simulations show how the various surfaces have different dynamic behaviors: (111) is substantially persistent/static, while (211) and (110) surfaces are much more dynamic. The (210) surface is found the most dynamic of the studied surfaces. These results are in agreement with the order of stability determined by the corresponding surfaces energies, and more importantly, with experimental evidence. For example, experimental studies proved that the (111) surface does not undergo any structural modification or premelting below the melting temperature, but the surface enters a disordered state only at the melting temperature.[71,72]

Regarding the (110) surface, a large body of experimental evidence shows that this orientation is characterized by an increasing disorder induced by thermal energy already at T > 550 K. This degree of disorder could not be simply assigned to thermal anharmonicity because it was much larger and for this reason was named "enhanced anharmonicity."[73] Trayanov *et al.* attributed this phenomenon to a missing-row type reconstruction, given the small difference between the surface energies of the reconstructed and unreconstructed surfaces.[5,74] Our simulations confirm this interpretation, as shown in Fig. S5. Here, we compare the snapshot of our Cu(110) surface after 150 ns at 700 K with the corresponding ideal (110) (1 × 2) and (1 × 3) reconstructions coloring the atoms according to their coordination. The formation of a (1 × 3) missing-row type reconstruction in the section highlighted by the dashed box can be readily seen.

More recently, the Cu(211) surface has been the object of several studies, given its reactivity toward the methanol synthesis.[47,75] Witte *et al.*[76] found that upon oxygen adsorption already below room temperature, the (211) surface also undergoes reconstruction forming a (211) (2 × 1) surface. The same type of reconstruction is observed in our simulations at 600 and 700 K. In Fig. S6, we compare the ideal (211) (2 × 1) reconstructed surface with a snapshot of

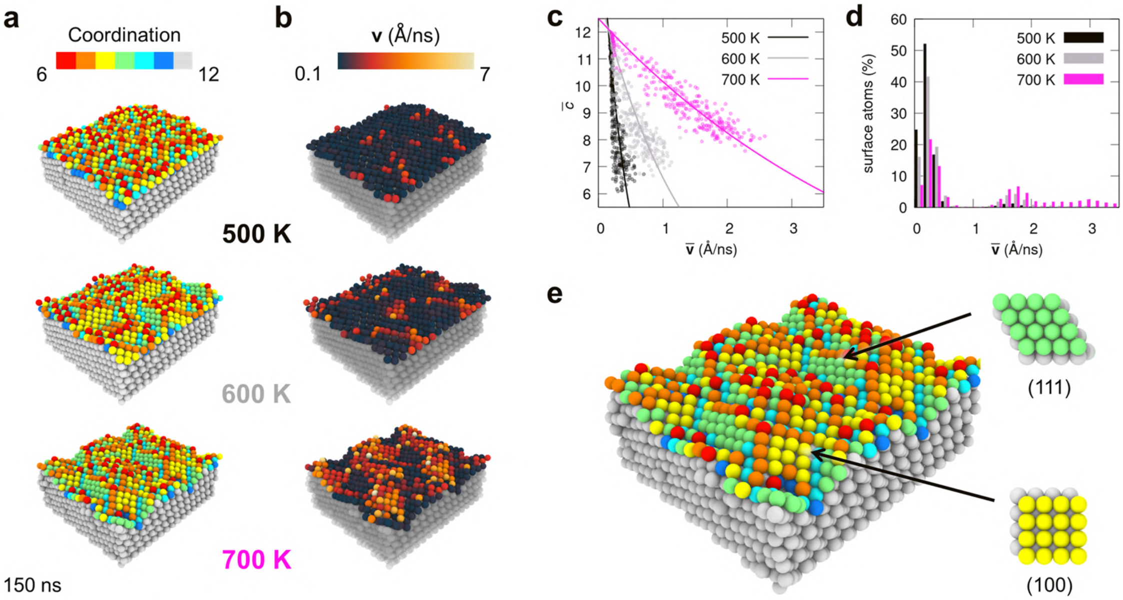

FIG. 2. Dynamic diversity in a metal surface. (a) The Cu(210) surface after 150 ns of DPMD simulations at 500, 600, and 700 K (top-to-down). Atoms are colored based on their coordination. (b) Relative diffusion velocity in the (210) surface at different temperatures: the Cu atoms are colored with the relative velocity, estimated for each atom in the system as the atomic displacement in the time interval of $\Delta t = 1500$ ps ($\Delta r/\Delta t$, expressed in Å ns^{-1}). Dark vs light colors in the snapshots identify static vs dynamic regions on the Cu surfaces (bulk and sub-surface atoms are shown transparent for clarity). (c) Correlation between the average coordination number and velocity for the atoms in the (210) surface at the three temperatures. Solid lines are exponential fits to guide the eye. (d) Histograms of the atomic velocity distributions in the (210) surface at 500, 600, and 700 K. (e) Zoomed detail of the coordination analysis for the (210) surface at 700 K, highlighting the emergence of (100) and (111) domains along the DPMD simulation (yellow and green details).

© Author(s) 2023

the (211) surface after 150 ns at 700 K. As for the (110), in this case, a reconstruction of the surface toward the (211) (2× 1) surface is easily detectable.

From a dynamic perspective, the (210) surface is the most interesting as it is found dynamic at all temperatures, with its dynamics increasing with temperature [Fig. 2(a)]. Deeper microscopic analyses reveal a considerable dynamic diversity in the behaviors of the atoms in the (210) surface. While the temperature is set globally in these DPMD simulation (see Sec. IV for details), the data show that there are atoms that move faster and atoms that are more static on the surface. This can be inferred via estimating along the DPMD the displacement of the individual atoms in the surface in the time interval $\Delta t = 1500$ ps ($\Delta r/\Delta t$).

For example, bright atoms shown in Fig. 2(b) move by tens of nanometers during the DPMD simulations, while black atoms just vibrate around their lattice position.

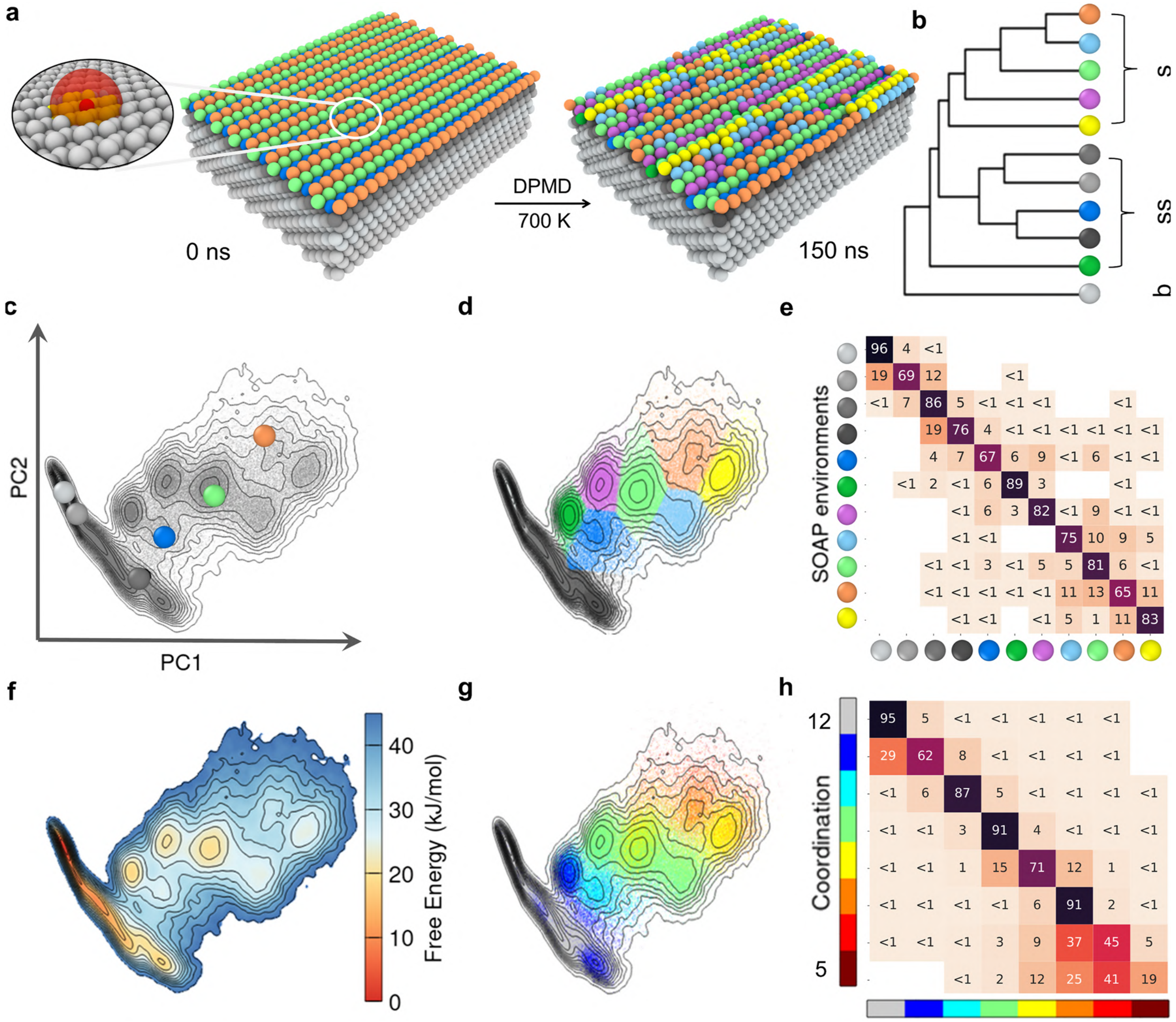

FIG. 3. ML of atomic environments in the Cu(211) surface and of their dynamics. (a) Cu atoms on the (211) surface colored based on the SOAP environments emerging along the equilibrium (last 75 ns) DPMD simulation at 700 K. (b) Hierarchical dendrogram connecting the 11 detected SOAP clusters, corresponding to six internal atomic environments—five sub-surface (ss) and one bulk (b)—plus five surface environments (s). (c) Projection on the first two principal components of the SOAP data PCA with density isolines. The colored dots indicate the position on the PCA of the native SOAP states present in the ideal (211) surface at the DPMD simulation start [(a), left: at 0 K]. (d) Unsupervized clustering (HDBSCAN*) of SOAP data identifies 11 main SOAP environments (microstates) that emerge at the equilibrium in the (211) surface at 700 K. (e) Transition matrix reporting the normalized probabilities (in %) for atoms to undergo transition between the SOAP clusters in a sampling time interval of *dt* = 300 ps. (f) Free energy surface (FES) computed from the PCA of the SOAP density data (c). (g) PCA projection of the SOAP dataset colored based on atomic coordination. (f) Transition matrix reporting the normalized probabilities (in %) for atomic transitions between the various coordination states along the DPMD (*dt* = 300 ps).

© Author(s) 2023

While in these simulations the atomistic surface models are thermostated and, on average, their temperature plateau to 500, 600, and 700 K, these analyses demonstrate that on a nanosecond timescale window, the structure and dynamics on the surfaces is not uniform. In particular, the plots in Fig. 2(b) show that these surfaces are dynamically diverse, being populated by domains that are more static and others that are more dynamic. Still, it is noteworthy to add that all these domains are continuously destroyed and formed with different timescales and are also in continuous dynamic exchange with each other, as shown in detail in Figs. 3, 5, and 6. The plots in Fig. 2(c) relate the average velocities and coordination numbers of the atoms. The minimum atomic coordination increases from ~6 (black data) to ~8 (in pink) while increasing the temperature in the (210) surface. The histograms of the atomic velocity distributions [Fig. 2(d)] indicate the variability with which the (210) surface atoms move, a few having high relative mobility, while the majority of them are more static. Figure 2(a) shows that at 500 K (top), the ideal structure of the Cu(210) surface is better preserved during the DPMD than at 700 K (bottom), where the increased kinetic/thermal energy triggers more considerable disordering and faceting, which in turn produces a surface configuration populated by more stable atomic arrangements (facets) with increased coordination that corresponds to (111) facets (in green) and (100) facets (in yellow). A zoom onto the (210) surface at 700 K [Fig. 2(e)] shows how such green domains correspond to (111) islands, non-native of this surface. Yellow-colored (100) square domains, non-native of (210), are also visible. It is interesting to notice that similar results were observed in the experimental work of Kirby *et al.*[77] In particular, they detected a faceting phenomenon of the Cu(210) surface induced by activated nitrogen already at room temperature. Another interesting aspect revealed by our analysis comes from the comparison of the colored domains in Fig. 2(e) vs Fig. 2(b), which provides an insight into the dynamic diversity of the surface, revealing how non-native (111) domains [Fig. 2(e), green] are more static, while (100) ones (in yellow) correspond to more dynamic regions in Fig. 2(b). Noteworthy, along the DPMD simulations, the atoms change color dynamically in these Cu surfaces, demonstrating the dynamic equilibrium present within them (see the supplementary material, Movies S1 and S2).

B. Unsupervized machine learning of the structure and dynamics of a copper surface

To obtain a more robust and general quantitative analysis, we turned to an advanced data-driven approach recently proven useful for reconstructing the structural/dynamical complexity of various types of molecular systems.[22,23,37–39] We use SOAP vectors as high-dimensional descriptors of the local environments surrounding each atom in these surfaces. Calculation of the SOAP spectra of all atoms along the DPMD simulations allows (i) to classify the local atomic environments that populate/emerge within the Cu surface in equilibrium conditions based on their levels of order/disorder and similarity, and (ii) to reconstruct the entire Cu surface dynamics (see Sec. IV for details).[22,23,38,39] Key advantages of such an analysis are that the SOAP descriptor is agnostic[35] and the analysis is unsupervized and data-driven, i.e., it does not require prior knowledge of the systems, while the SOAP-detected environments emerge directly from the DPMD trajectories (bottom-up analysis).[22,23,38,39]

We perform our SOAP analysis on the DPMD trajectories of the various simulated Cu surfaces (see the supplementary material, Fig. S12). After 75 ns of DPMD, all surfaces reach a microscopic equilibrium where the atomic environments populating the systems do not change anymore (see the supplementary material, Figs. S2–S4). The last 75 ns of DPMD, representative of the equilibrium of the modeled surfaces, are thus retained for the analyses. From these, 250 snapshots—one every $\Delta t = 300$ ps—are extracted and analyzed. In particular, we calculate the SOAP spectra of each of the top-most 1300 atoms in the Cu surface, obtaining at each DPMD sampled snapshot 1300 SOAP spectra representative of the pristine arrangements/dynamics of the atomic surfaces in the studied regimes, and a dataset of 325 000 SOAP spectra in total for each simulated case (see Sec. IV). Unsupervized clustering of the SOAP data via the Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN*) algorithm[78,79] identifies the main SOAP clusters (states/environments) populating the equilibrium DPMD trajectories [Figs. 3(a) and 3(b)]. As an example, HDBSCAN* identifies 11 main SOAP clusters in the Cu(211) surface at 700 K [Figs. 3(a) and 3(b)]. These correspond to the density peaks seen in the Principal Component Analysis (PCA) of the SOAP data in Figs. 3(c) and 3(d), namely, to the most "visited" atomic environments during the DPMD simulation. The dendrogram of Fig. 3(b) shows the adjacency between the detected SOAP micro-clusters (based on their similarity), revealing three main macro-groups: surface (s), sub-surface, and bulk (b) atomic environments. Shades of gray in Figs. 3(a) and 3(b) correspond to deeper sub-surface layers connected to the bulk (light gray). Dark blue and green identify states at the interface between sub-surface and surface. Brighter colors identify the different surface states (s). The colored dots on the PCA projection of Fig. 3(c) indicate which SOAP states are present in the ideal (211) surface at 0 K (at DPMD start). Comparing Figs. 3(c) and 3(d), it is clear how, in terms of external surface states, only the orange, light-green, and blue SOAP environments are native of the ideal (211) surface (at 0 K). All other surface environments that emerge along the DPMD at 700 K (yellow, cyan, purple, etc.) are non-native states, which emerge with temperature.

Since the detected SOAP environments are well-sampled along the DPMD, we know the clusters' density at the equilibrium, and as we have information on the SOAP environment each atom belongs to at every sampled DPMD snapshot, we can reconstruct the dynamics and thermodynamics of the Cu surface. The transition matrix of Fig. 3(e) reports the normalized probabilities for an atom belonging to a given SOAP environment at a time t to remain in that environment (diagonal entries) or to undergo transition to a different SOAP environment (off-diagonal entries) at $t + \Delta t$ (i.e., after $\Delta t = 300$ ps in this analysis) in the Cu(211) surface at 700 K. Such transition matrices are non-symmetric as they are normalized to have the rows summing to 100, while the raw non-normalized matrices are conversely symmetric, as the Cu surface is at the equilibrium (see the supplementary material, Fig. S10). In general, the higher the numbers on the diagonal of the matrix, the more persistent is the surface. Vice versa, the higher the off-diagonal probabilities, the more probable are the atomic transition between the SOAP states in Δt and the more dynamic is the surface. The fact that

© Author(s) 2023

the transition matrix of Fig. 3(e) is rich of off-diagonal entries demonstrates the rich dynamics present in this surface. From such transition probabilities, one can estimate the characteristic transition rates/frequencies (i.e., by dividing the off-diagonal entries by $100 \times \Delta t$, being $\Delta t = 300$ ps in the analyses reported herein). For example, in the (211) surface at 700 K, the atoms belonging to the purple SOAP environments undergo transition to blue with a probability of ~6% every 300 ps, which corresponds to a transition rate of ~0.2 ns^{-1} and transitions occurring in the timescale of tens of nanoseconds (assuming a single-step transition). A consistent dynamics is obtained even changing the temporal resolution of the analysis (Δt), which proves the robustness of the obtained results (see the supplementary material, Fig. S11).[80–82] All entries <1% should be taken as qualitative, as they refer to events observed only sparsely along the DPMD.

This analysis shows there is no direct communication/exchange between bulk and surface states, while these may communicate only through intermediate sub-surface states. Even at 700 K, where the (211) surface is considerably dynamic, the transitions occur on a nanosecond timescale. This demonstrates how the (211) surface has a "discrete" dynamics.

A similar "discrete" dynamics is observed for the (110) surface [supplementary material, Figs. S12(a)–S12(e)]. In addition, here, the atoms of the first layers move via jumps across different crystallographic positions. This type of motion is correlated with the fact that when computing the $g(r)$ of the surface the peak positions remain unchanged; however, the surface atoms' motion leads to a reduction in the peak intensities and a more diffuse background, as shown in Fig. S9. This type of atomic motion could explain the anomalous reduction in the intensity peaks of Cu(110) at T > 550 K obtained by diffraction experiments.[5,6,8] This phenomenon was referred to as "enhanced anharmonicity" since the only corresponding property was an enhanced mean square displacement of the surface atoms. However, here, we reinterpret this "enhanced anharmonicity," suggesting that its origin comes from the frequent jump-motion of the surface atoms.

The same analysis for the other surfaces studied herein shows that (210) is the most dynamic surface, with more fluid-like dynamics [supplementary material, Figs. S12(f)–S12(l)]. This behavior is also mirrored by a quasi-liquid-like $g(r)$ at 700 K (see Fig. S9). On the other hand, the (111) surface is more static: only sparse transitions can be observed even at 700 K.

From the inverse exponential of the PCA density, it is also possible to estimate a pseudo free energy surface (FES) for the Cu(211) surface model at 700 K [Fig. 3(f)]. Such FES shows how, at 700 K, all surface SOAP states are separated from each other by relatively low free energy barriers within ~10 – 15 kJ mol^{-1} (10 kJ mol^{-1} corresponds to ~2 kT at 700 K). Their transitions can be thus efficiently sampled during an equilibrium DPMD (see the supplementary material, Movie S3). The SOAP analysis of Figs. 3(d)–3(f) reveals how—since the SOAP states have characteristic lifetimes and transition rates, and are in continuous interchange with each other—at 700 K, the (211) surface has just an average configuration that is purely statistical. The great flexibility and the agnostic nature of such data-driven analysis come with the disadvantage of a non-straightforward interpretation. There is no-straightforward correspondence between the detected SOAP states and their physical differences. Figure 3(g) shows the same SOAP PCA colored according to the coordination number of each atom. This shows how the SOAP analysis captures very well differences in the coordination number between the atomic environments in the Cu surface, while at the same time, a simple coordination analysis is less sensitive—e.g., green domains in the PCA having the same coordination number (~9) but corresponding to different SOAP density peaks [Fig. 3(g)].

From a broader perspective, this data-driven analysis shows that such Cu surfaces possess a non-trivial structural/dynamical complexity well below melting. At the same time, these results underline the importance of relying on a structurally and dynamically accurate force field (as the *NN* potential used herein) to obtain meaningful insights on such a complexity. The fact that new states, non-native/present in a given (ideal) surface, may appear at finite temperature along the DPMD simulations poses fundamental questions on the elusive identity of these surfaces. For example, are the new environments that emerge with a temperature closer to native environments present in the ideal surface (at 0 K), or, e.g., to other ones, native of different types of surfaces? To answer such questions, we designed another complementary (top-down) analysis.

C. A dictionary of SOAP atomic environments

We created a dictionary of SOAP atomic environments by computing the SOAP spectra for all atoms in the optimized Cu(111), (110), (211), and (210) surfaces (at 0 K). The analysis identifies three SOAP atomic environments in the ideal (111) surface—Fig. 4(a) (yellow): one bulk ($b_{(111)}$), one sub-surface ($ss_{(111)}$), and one surface environment ($s_{(111)}$). The other surfaces are more structurally diverse: the ideal (110) is characterized by five main SOAP atomic environments (in red), while the ideal (210) and (211) are characterized by seven SOAP environments [Fig. 4(a): blue and green, respectively]. In total, we obtain 22 distinct SOAP spectra characteristic of the bulk, surface, and sub-surface atomic environments proper of these ideal surfaces. We created a unique SOAP dataset containing all these SOAP environments and computed from their characteristic SOAP power spectra their mutual distances (d_{soap}) in the global high-dimensional SOAP feature space (see Sec. IV for details).[23,37,39] Such d_{soap} metrics allows quantifying the similarity between the various characteristic SOAP spectra, providing a rich data-driven classification of all detected atomic environments present in these ideal surfaces. The result is the distance matrix of Fig. 4(b). The colors of the matrix cells represent the SOAP distance (d_{soap}) between the various SOAP atomic environments: dark colors indicate very similar environments ($d_{soap} \sim 0$), while light colors identify structurally different atomic environments. The dendrogram adjacent to the matrix [Fig. 4(b): left] shows the hierarchical clustering of the various SOAP environments based on their similarity. The matrix reveals dark macro-areas indicating SOAP environments that are nearly identical in the various surfaces, i.e., bulk (b) environments and some high-coordination sub-surface ones (ss). In general, deep atomic environments are found quite similar in the various ideal surfaces. The matrix also reveals non-obvious similarities between the high, medium, and low-coordination surface environments, e.g., $s_{(111)}$ vs $s^{m}_{(211)}$ and $s^{m}_{(210)}$, or $s^{l}_{(110)}$ vs $s^{l}_{(210)}$.

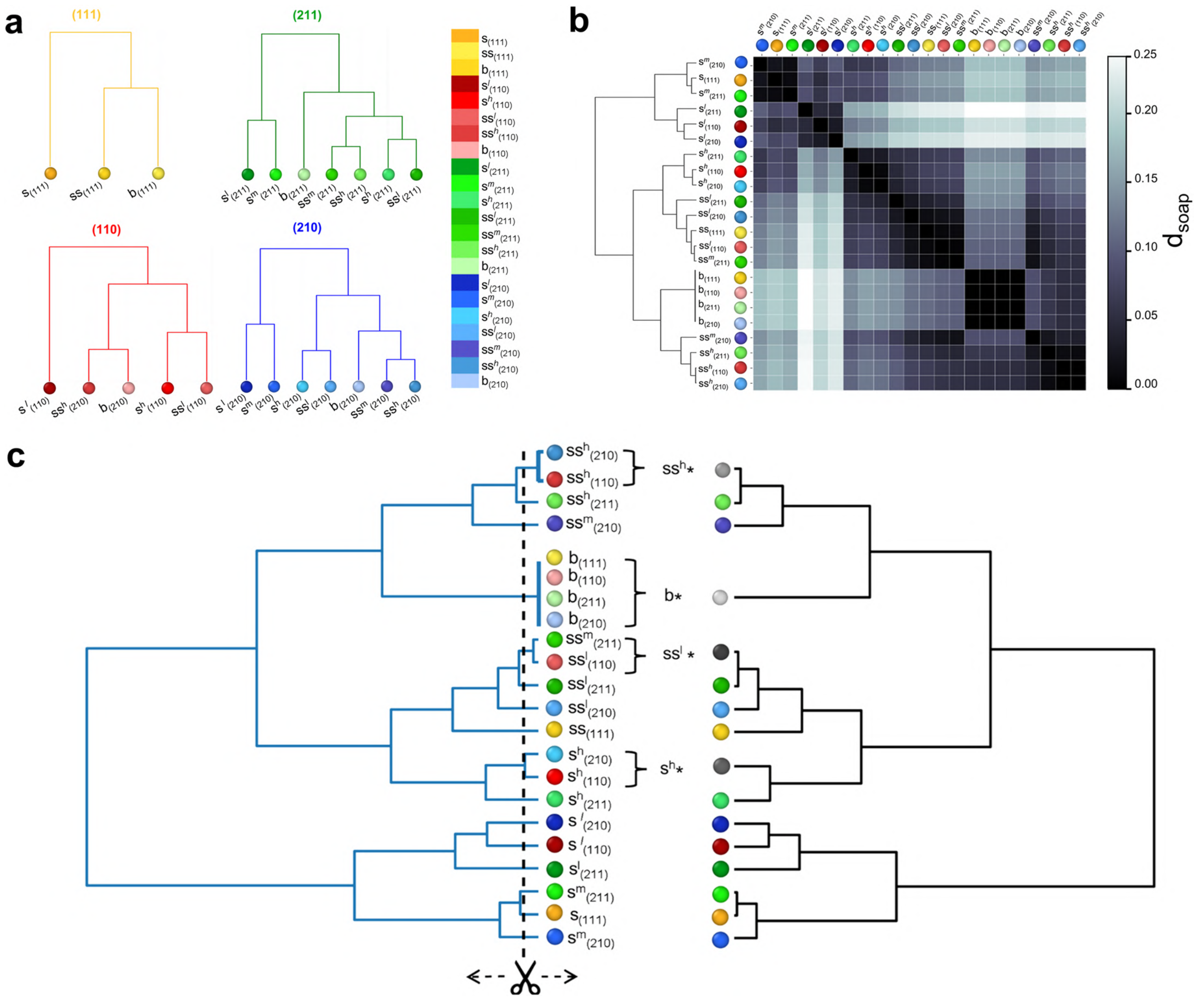

FIG. 4. A SOAP dictionary for classifying atomic environments in Cu surfaces. (a) Hierarchical dendrograms for the SOAP environments detected in the different ideal Cu surfaces (at 0 K): Cu(111) environments are shown in yellow, (211) in green, (110) in red, and (210) in blue. (b) Distance matrix: the color scale indicates the distance in the high-dimensional SOAP feature space (d_{soap}) between all SOAP environments in the Cu surfaces. (c) Left: hierarchical dendrogram (in blue) showing the similarity between all SOAP environments [also reported on the left of the d_{soap} matrix, (b)]. Right: same dendrogram cut at $d_{soap} \geq 0.01$ (in black). This results in grouping very similar SOAP environments into common macro-clusters: e.g., bulk (b*), sub-surface (ss*), and surface (s*).

Figure 4(c) shows how the complete dendrogram (left, in blue) can be cut in order to consider only detected SOAP distances greater than a minimum value, offering the opportunity of a variable resolution in the analysis. For example, Fig. 4(c) (right) shows what the dendrogram becomes when considering only $d_{soap} \geq 0.01$ (in black). At this level of resolution, the bulk environments of all surfaces are grouped in a single bulk state (b*). The same happens for other very similar sub-surface (ss*) and surface (s*) environments. While complete information is encoded in the pristine dendrogram, this offers the opportunity to modulate the noise/relevance trade-off of the analysis, focusing on differences that are really meaningful (e.g., distinguishing between the bulk states of these surfaces is useless, as these are identical SOAP environments). As it will be demonstrated in Sec. II D, this is important, for example, when using such SOAP data and d_{soap} metric to track the similarity between the atomic environments emerging in the metal surface in equilibrium conditions and those included in the SOAP dictionary (top-down classification).

D. Dynamic reconstructions and statistical equivalent identities of copper surfaces

Starting from the Cu(211) surface at 700 K, at each time step along the DPMD simulation, we measure the d_{soap} distance between the SOAP spectrum of each atom and all SOAP spectrum characteristic environments present in the dictionary of Fig. 4. At each DPMD

© Author(s) 2023

time step, each atom is then attributed to the closest SOAP environment/class (smallest d_{soap}) in the SOAP dictionary. This allows us to track the transformations in the surface along the DPMD and to estimate the reconstruction of non-native domains, their lifetime, and dynamics in terms of atomic transitions between them.

Since we are interested in surface reconstructions and dynamics, we next focus only on the five top-most layers of the simulated surfaces (see Sec. IV). Figure 5(a) shows how the (211) surface, ideally composed only of green native SOAP sites at 0 K (left), convert into local non-native domains, proper of (111), (110), and

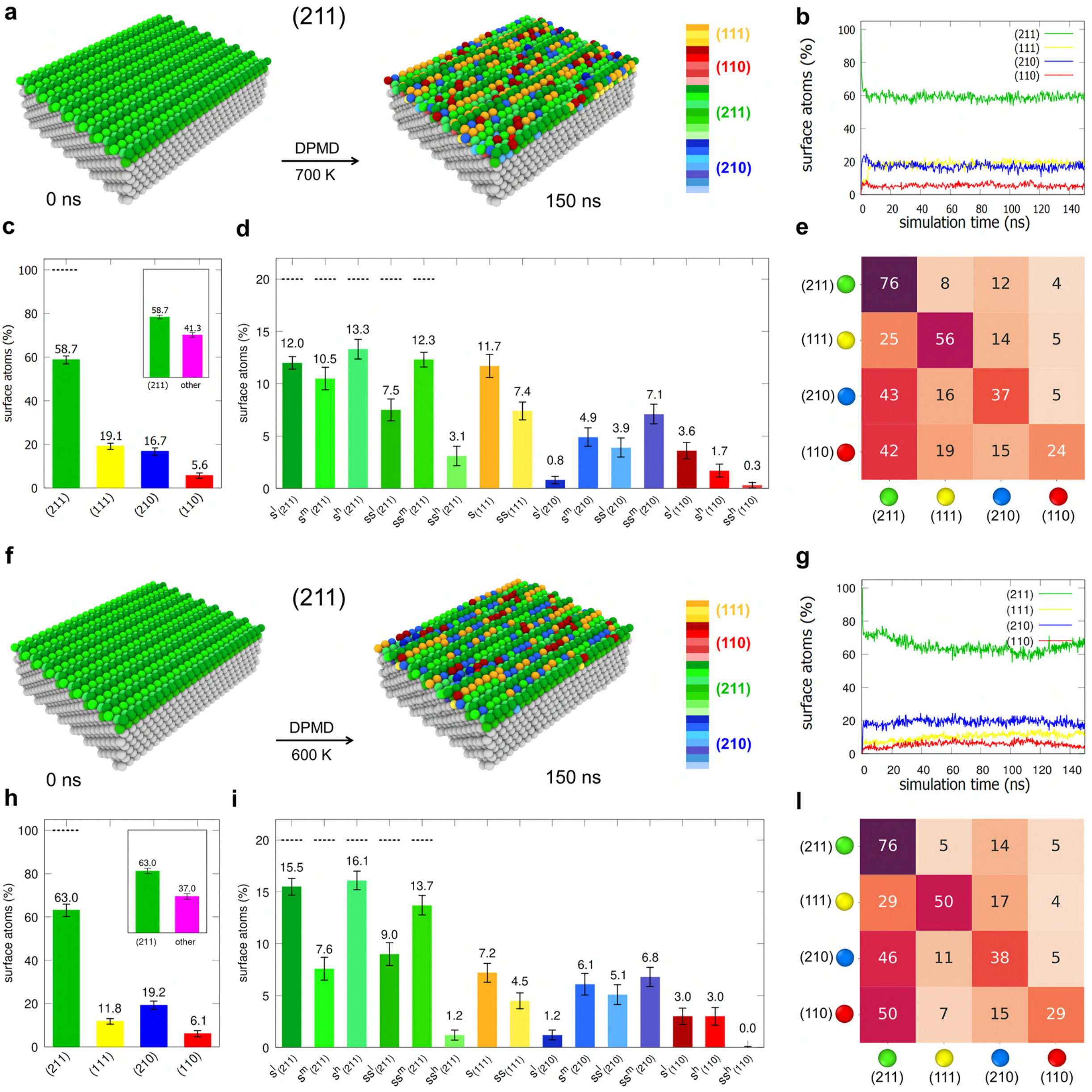

FIG. 5. Dynamic reconstructions and equivalent identity of a Cu(211) surface. (a) Cu(211) at 0 ns (left) and after 150 ns of DPMD at 700 K (right): SOAP environments native to the ideal (211) surface are colored in green. Red, blue, and yellow colors identify non-native atomic domains, proper of (110), (210), and (111) surfaces, respectively. (b) Populations of the native and non-native environments (in %) in the (211) surface at 700 K as a function of DPMD time. (c) Equilibrium composition of (211) at 700 K (% and standard deviations) in terms of native (green) and non-native domains (yellow, blue, and red, combined in pink in the inset). (d) Breakdown of (211) composition at 700 K. Dashed lines indicate the composition at the DPMD start. (e) Transition matrix showing the probabilities for atomic transitions in (211) between native and non-native environments at 700 K (within Δt = 300 ps). (f)–(l) Same analyses for the (211) surface at 600 K.

© Author(s) 2023

(210) ideal surfaces (right: yellow, red, and blue, respectively), along 150 ns of DPMD at 700 K. The environment populations of Fig. 5(b) show that the surface reaches a microscopic equilibrium along the DPMD, being populated of native and non-native domains. Along the DPMD, ~40% of native domains disappear, converting into non-native domains [Figs. 5(b) and 5(c)]. On a statistical level, the (211) surface thus preserves its own identity only by ~60% at 700 K [Fig. 5(c): inset]. In particular, ~19% of the emerging domains correspond to (111) environments, ~17% to (210), and ~6% to (110) ones [Fig. 5(c)]. Such surface reconfiguration is rather fast in this system [Fig. 5(b)]. Figure 5(d) shows a breakdown of the detailed native and non-native environments that populate the surface in equilibrium conditions. Among all the emerging non-native environments, $s_{(111)}$ is the predominant one (~12%), followed by $ss_{(111)}$ and $ss^{m}_{(210)}$, both constituting ~7% of the (211) surface at 700 K. While such an analysis provides detailed quantitative information on the composition, structural diversity, and reconfiguration in the metal surface, it is interesting to investigate the dynamical features of such phenomena.

The transition matrix of Fig. 5(e) reports the normalized probabilities for the atom transitions between native and non-native environments in the (211) surface at 700 K (in $\Delta t = 300$ ps). The diagonal entries show that only the native (green) and yellow (111) domains have a residence probability >50%. This means that at the temporal resolution of this analysis, these are somewhat persistent domains. On the contrary, (210) and (110) atomic environments have persistence probabilities well <50%. Such domains are considerably more dynamic, and the atoms composing them have a higher probability to re-convert into native domains (red-to-green and blue-to-green transition probabilities ~42% – 43%) than to remain there in $\Delta t = 300$ ps. This provides a picture reminiscent to that shown in Fig. 2 for the (210) surface, where, above the Hüttig temperature, such dynamically diverse surface appears to be composed of stable domains coexisting in equilibrium with more dynamic ones (supplementary material, Movie S4).

Decreasing the temperature to 600 K, the behavior of (211) does not change substantially [Figs. 5(f)–5(l)]. In general, the transformation of the (211) surface is rather similar, while it occurs slightly slower at 600 K than at 700 K [Fig. 5(g)]. The appearance of (111) domains is just slightly reduced than at 700 K [Figs. 5(h) and 5(i): ~12% at 600 K vs ~19% at 700 K]. The transition matrix shown in Fig. 5(l) is also very similar to that shown in Fig. 5(e). This demonstrates how the phenomena occurring in these regimes are thermodynamically driven, being, e.g., the surface energy of (111) lower than that of (211) (see Table S1). Conversely, the same analysis demonstrated how the (211) surface appears as substantially static at 500 K (Hüttig temperature of Cu: 447 K) in the same DPMD timescales. In such a regime, the thermal bath is evidently insufficient to trigger the reconfiguration and the surface remains trapped in the ideal (211) configuration (supplementary material, Fig. S13). In Sec. II A, the similarity between the final state of the simulated Cu(211) surface and the corresponding missing-row (211) (2 × 1) reconstruction is highlighted. We deemed it interesting to analyze with our SOAP-based approach the ideal (211) (2 × 1) reconstructed surface, and noticeably, this surface shows a number of yellow (111) atomic domains, which is in line with what has been observed for the Cu(211) surface simulated at 700 K (see Fig. S8). This confirms the ability of the present approach to detect (111) facets in reconstructed surfaces. The differences between the simulated and reconstructed (211) surfaces originate from the actual motion of the atoms that undergo frequent changes in their surrounding environment.

Comparing the behavior of the other Cu surfaces, the same analysis reveals that the (110) surface is highly dynamic and substantially unstable at 700 K [Figs. 6(a)–6(e)]. The snapshots of Fig. 6(a) show how during 150 ns of DPMD, the surface becomes largely populated with non-red colors and mainly green (211) domains (supplementary material, Movie S5). Figure 6(b) indicates a substantial reconfiguration of the (110) surface. The instability of (110) at 700 K is manifest in the fact that the surface is reconstructed by ~85%. The native red environments drop to <20% of the surface, reconstructing in large part (211) green domains [Fig. 6(c): rising to ~50%]. Starting from an ideal (110) configuration, such a surface evolves toward reconstructing a different, more stable surface. Similar to what was observed for (210) at 700 K, the transition matrix of Fig. 6(e) reveals how the residual red native domains are also highly dynamic (survival probability <40%), which fits well with their relatively high surface energy,[83,84] while the persistent domains in this surface are non-native environments. As seen experimentally for other metals,[85–87] also the Cu(110) surface appears unstable at 700 K and reconstructs large surface domains structurally/dynamically similar to the (211) surface ones (Fig. 5). In addition, at 600 K, the (110) surface has a dynamics similar to that of (211) at the same temperature, while the reconstruction of (211) domains is much slower than at 700 K. Similar to (211), at 500 K, the (110) surface is also substantially immobile and preserves its identity: the thermal bath is insufficient to trigger the reconstruction in the timescales accessible via these DPMD simulations (see the supplementary material, Fig. S14).

The emergence of a large number of green (211) atomic domains can be quite puzzling. However, it can be explained through a SOAP analysis of the missing-row (110) (1 × 3) reconstructed surface. When coloring this surface with our SOAP-based dictionary, we observe that the atoms of the surface are composed mainly of green (211) and yellow (111) atomic domains. This evidence shows that the atomic environments that are native to the (110) (1× 3) reconstructed surface are actually very close in terms of soap distance to those of the (211) surface (see Fig. S7).

The Cu(210) surface is more dynamic than both (211) and (110) surfaces. Among all studied surfaces, (210) has the least coordinated atoms: the $s^{l}_{(210)}$ are the only atoms with coordination 6 in the whole dataset [see Fig. 1(c)]. Figures 6(f)–6(l) show the analysis for (210) at 500 K. Even so close to the Hüttig temperature, this surface undergoes considerable reconstruction (see the supplementary material, Movie S6). The (210) reconstructs non-native domains by >60%, preserving its identity only by <40% [Figs. 6(g) and 6(h)]. Increasing the temperature to 600 or 700 K has the unique effect of accelerating such reconfiguration, while the equilibrium populations remain substantially preserved (supplementary material, Fig. S15). This fits well with the higher energy of this Cu surface.[83,84] One difference is in the dynamics of such atomic environments. The transition matrix of Fig. 6(l) shows diagonal entries very close to or higher than 50%. At the resolution of our analysis, the dynamics that emerges in this surface at 500 K is discrete (solid dynamics). Conversely, increasing the temperature to 700 K creates dynamically

© Author(s) 2023

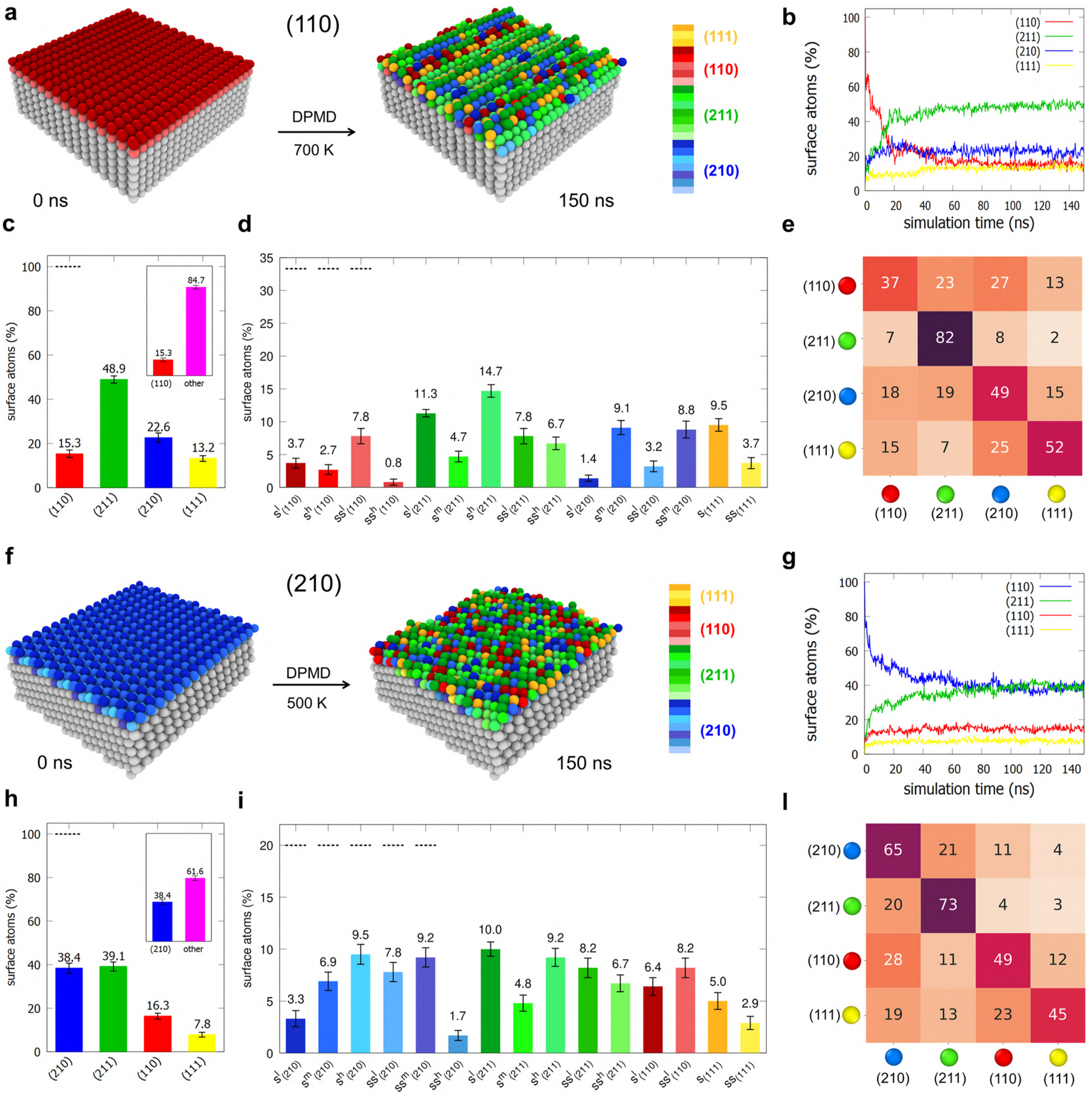

FIG. 6. Dynamic reconstructions and equivalent identity of other Cu surfaces. (a) Cu(110) at 0 ns (left) and after 150 ns of DPMD at 700 K (right): SOAP environments native to ideal (110) are colored in red. Green, blue, and yellow colors identify non-native environments proper of (211), (210), and (111) surfaces, respectively. (b) Environment populations in the (110) surface at 700 K as a function of DPMD time. (c) Equilibrium composition of (110) at 700 K (% and standard deviations). Inset: native domains in red and non-native in pink. (d) Breakdown of (110) composition at 700 K. The dashed lines indicate the composition at DPMD start. (e) Transition matrix showing the probabilities for atomic transitions in (110) at 700 K (within Δt = 300 ps). (f)–(l) Same analyses for the (210) surface at 500 K. Same analyses for surfaces at other temperatures are reported in the supplementary material.

persistent solid-like domains—native blue (210) and non-native green (211) domains—coexisting with dynamic domains (supplementary material, Fig. S15). This also confirms that, in this case, increasing the temperature does not generate a uniform increase in the dynamics of atoms but the emergence of local dynamic domains and a non-uniform dynamically diverse surface.

The last case that we compare is the close-packed Cu(111) surface. Even the simple coordination analysis of Fig. 1 clearly shows that this surface is very stable[83,84] and does not undergo any considerable reconstructions in such regimes. Even at 700 K, surface atoms with coordination ≠9 emerge only sparsely and as statistical local fluctuations, indicating vibrations rather than reconstructions.

© Author(s) 2023

III. DISCUSSION

We report a data-driven approach that allows resolving at an atomistic resolution the complex structural dynamics of metal surfaces above the Hüttig temperature. As a case study, we use Cu surfaces. However, the approach is versatile and can be applied to other metal systems. The approach provides a detailed microscopic characterization of the atomic environments composing such dynamically diverse surfaces, the rates with which these emerge/disappear, their residence time, and persistence (see, e.g., Fig. 3). The development of a dictionary of SOAP atomic Cu surface environments (Fig. 4) allows for a data-driven analysis of the similarity/differences between the local motifs that appear in dynamic equilibrium conditions on the different Cu surfaces [Figs. 5, 6(c), 6(d), 6(h), and 6(i)]. This provides an exquisitely statistical picture of these metal surfaces, and a data-driven estimation of their "equivalent identity" in dynamic regimes (Figs. 5 and 6). Knowing what local environments emerge, how often, and for how long is a prime requisite to understand what a metal surface looks like and its properties in determined thermodynamic conditions.

The developed *NN* potential allows dynamically reliable DPMD simulations of relatively large Cu surface models composed of 2,400 atoms (replicating on *xy* through periodic boundary conditions). Noteworthy, the transition matrices of Figs. 5 and 6 show transition probabilities ranging ~5% – 40% in $\Delta t = 300$ ps, revealing a rich microscopic dynamics in such metal surfaces with characteristic times for the transitions between the various environments in the order of nanoseconds. On a technical standpoint, this shows how such simulations provide access to information extremely difficult to attain with other approaches.

From a scientific point of view, metal surfaces in most cases are still studied treating the surface as a rigid object; however, the rich structural dynamics seen in these metallic materials at temperature regimes of 500–700 K in our simulations indicate that the actual scenario is much more complex and that the intrinsic dynamics of the metal surface must be explicitly accounted to understand surface properties. In particular, it is intriguing to note that the results of Figs. 3, 5, and 6 provide a picture of such metal surfaces that is quite far from that of hard materials, revealing internal dynamic equilibria and a structural/dynamical diversity that, in a sense, is reminiscent of that of soft dynamic materials.[22–24]

In perspective, our data-driven approach offers remarkable opportunities to relate the innate structural dynamics of metals to their properties. While we are interested in resolving the complex structural dynamics of metal surfaces *per se*, we envisage that this will have a considerable impact on various fields, from the study of their mechanical properties to, e.g., heterogeneous catalysis, where the dynamical emergence of local atomic environments with utterly different reactivities and survival lifetimes may have a strong impact on the catalytic activity.[19,47,48,88–91] In general, this study reveals and resolves the complex dynamic character of metals in dynamic regimes, demonstrating how these cannot be simply studied as static structures even far from the melting temperature. This changes the way we look at such materials, opening new exciting directions toward data-driven statistical reinterpretations of their properties.

IV. METHODS

A. Neural network representation of the inter-atomic interaction potential

The DFT database needed to train the Cu *NN* potential was generated extracting configurations along *ab initio* molecular dynamics (AIMD) trajectories of small Cu systems. Simulations were performed using the PWscf code of Quantum ESPRESSO.[92] In these calculations, Perdew–Burke–Ernzerhof (PBE) exchange–correlation density functional[93] forces were used to propagate the nuclear dynamics. Preliminary tests demonstrated that this functional provides a reliable representation of the metal surface, free of empirical corrections, and constitutes an ideal compromise between accuracy and computational cost. Moreover, the same approach was also demonstrated as reliable for other metals, e.g., for gold and silver surfaces.[94] Ultrasoft Rappe-Rabe-Kaxiras-Joannopoulos (RRKJ) pseudopotentials[95] replaced explicit core-valence electron interactions, while electron density and wavefunctions were expanded in plane-waves with energy cutoffs of 220 and 50 Ry, respectively. Occupation was treated by the cold smearing technique of Marzari *et al.*[96] with a Gaussian spreading of 0.01 Ry. The Brillouin zone was sampled using a $2 \times 2 \times 2$ Monkhost–Pack k-point grid[97] for the bulk structure while a $2 \times 2 \times 1$ k-point grid was used for the slab models.

Convergence against cutoff energy, Monkhost–Pack sampling, and occupation was tested, and the setup described was chosen as the best compromise between feasibility and accuracy. The AIMD simulations were carried out with a time step of 1.0 fs in a constant volume and temperature (NVT) ensemble using the stochastic velocity rescaling thermostat of Bussi *et al.*[98] In order to span a larger portion of the configurational space, we simulated the systems at temperatures ranging between 500 and 700 K.

The AIMD-simulated systems include both bulk and surface structures. The bulk was modeled by a periodically repeated supercell of size $7.1638 \times 7.1638 \times 7.1638$ Å^3 containing 32 atoms. For surface calculations, four slabs models were used to construct the (100), (110), (111), (211), and (210) copper surfaces. Slabs with a different number of atomic layers were built and a vacuum layer was set in the *z* direction. The first two bottom Cu layers were kept fixed during optimization and AIMD simulations.

We trained the Cu *NN* potential using the DeePMD-kit package.[31,55] The smooth version of the deep potential model was adopted, with a cutoff radius of 6.0 Å. To remove the discontinuity introduced by the cutoff, the 1/r term in the network construction is smoothly switched-off by a cosine shape function from 1.0 to 6.0 Å. The filter (embedding) network has three layers with (25, 50, and 100) nodes/layer and the fitting net is composed of three layers, with 240 nodes each. The network is trained with the ADAM optimizer,[99] with an exponentially decaying learning rate from 1.0×10^{-3} to 5.0×10^{-8}. The batch size was chosen as 4. The prefactors of the energy and the force terms in the loss function change during the optimization process from 1 to 10 and from 1,000 to 1, respectively. The final model used for the production run was trained for 10.0×10^6 steps (see the supplementary material, Fig. S1).

The choice of the training dataset is a crucial step in the training of an *NN* potential. For this reason, we used the configurations collected along the DFT MD simulations to train a first "guess" Cu

© Author(s) 2023

NN potential and we then used it to run several DPMD simulations for the different surfaces at different temperatures. This allowed us to explore a larger portion of the configurational space accessible in the temperature range of interest (500–700 K) and to efficiently extract new configurations to add to the training dataset, further enriching the *NN* potential. This active learning protocol follows the same procedure introduced by Deringer and Csányi[56] and later implemented in the DeepMD software package by Zhang *et al.*[57] Here, the criterion used to select new configurations is based on the agreement on the forces' predictions made by an ensemble of four *NN* potentials, which have been trained on the same reference dataset but with different initial weights. We measured the model deviation as the maximum (over the force components) of the standard deviation on the forces predicted by such an ensemble of *NN* potentials. Whenever the model deviation for one configuration was in the range of $[27–240 \times 10^{-3}]$ eV/Å, the corresponding structure was included in the new training dataset. The iterative process continues until when no new configurations are visited so that the *NN* potential can be considered "complete" for the sampled conditions. A total of 10 000 configurations were used for the training, and the final root mean square errors of the *NN* potential on the testing set are equal to 1.0 meV/atom for the energy and 40 meV/Å for the force.

To verify that the use of ~100 atom surface environment patches (accounting up to the ~4 – 5th neighbor) for the training allows providing a *NN* potential that is robust and complete (e.g., free from spurious finite-size effects), we also performed additional tests using larger surface patches (~600 atoms) for the training. Our tests demonstrated that the deviations in the forces and energies thus estimated are systematically found within the training-testing errors, confirming, thus, the robustness of the approach adopted for the study of such systems and of the obtained *NN* potential.

Moreover, the fine sampling of atomic configurations and associated energies guaranteed by the DFT MD calculations allowed to include in the *NN* potential training dataset configurations corresponding both to the local minima and transition states. In this way, the obtained *NN* potential is trained to represent with DFT accuracy the differences in energy between different sampled configurations, the transition barriers between them, and thus the transition kinetics.

This provided us with an atomistic *NN* potential with DFT precision in the treatment of the structure, energy, and dynamics of Cu surfaces. The *NN* potential is also rich in that, thanks to the inclusion of various types of surface environments in the training set [Fig. 1(b)]. This is trained to cover a variety of atomic configurations—from those, more stable, present in the FCC bulk and (111) surface at 500 K, to those emerging, e.g., in the Cu(210) surface, which is highly dynamic (and nearly pre-melted) at 700 K. The use of a structurally and dynamically accurate Cu force field is key in our case, as it allows us to reconstruct the internal atomic dynamics of Cu surfaces in an accurate and reliable way. Nonetheless, on a qualitative point of view, rather similar internal dynamics of the Cu surfaces has been obtained also in MD simulations using a different general-purpose and widely used force field for copper available in the literature,[49,51,100] confirming how the rich internal dynamics seen in our DMPD simulations is not exclusive of our *NN* potential and is somehow innate in these surfaces at this regimes.

B. Atomistic surface models and DPMD simulations

All production simulations have been conducted using FCC Cu(111), (110), (210), and (211) surface models composed of 2,400 atoms [only (210) has 2304] replicating on the *xy* plane through periodic boundary conditions, and having thickness >15 Å. During all simulations, the two bottom-most atomic layers were kept fixed (these two layers are then not considered in all our analyses). Preliminary tests demonstrated that such a size is large enough to prevent finite-size effects on the structural dynamics of the surfaces with the adopted setup.

All DPMD simulations were performed using a large-scale atomic/molecular massively parallel simulator (LAMMPS).[101] The temperature was controlled using the stochastic velocity rescaling thermostat of Bussi *et al.*[98] The integration time step used is 1 fs and the relaxation time for the thermostat was set to 0.1 ps. Coordination and velocity analyses of surface atoms were performed with OVITO software.[102] Coordination was calculated as the number of neighbors within a certain cutoff ($R_{cutoff} = 3.1$ Å) that corresponds to the first minimum of the Cu–Cu radial distribution function. Surface atoms' velocities in the analysis of Fig. 2(b) were calculated as $\Delta r/\Delta t$, where Δr was computed as the atomic position displacement within a certain time window $\Delta t = 1500$ ps. Both atomic coordination numbers and velocities reported in Figs. 2(c) and 2(d) were averaged over the second half of the trajectory.

C. Smooth overlap of atomic position

The Smooth Overlap of Atomic Position (SOAP) is a many-body atomic descriptor, conceived to accurately reproduce particle densities of a given system of interest.[35] In our analyses, we center one SOAP vector in each atom of the surface models. For each atom at each time step, this provides a SOAP "power spectrum," which is a high-dimensional descriptor of the level of order/disorder in the surrounding of each atom in the system. The advantage of the SOAP descriptor is that it is general and abstract, and thus, it can also be used in a flexible way, in principle, for other metals and other types of molecular systems.[22,23,37–39]

Given a system conformation Γ in the 3D space, the local atomic density, $\rho_i(\Gamma,\vec{r})$, is defined in the neighborhood of every SOAP center within a spatial cutoff r_{cut}; the SOAP power spectrum can be evaluated projecting $\rho_i(\Gamma,\vec{r})$ onto a basis of orthogonal radial functions $g_n(r)$ and spherical harmonics $Y_{lm}(\theta,\phi)$, which for the *i*th can be expressed as

$$\rho_i(\Gamma,\vec{r}) = \sum_{j \in r_{cut}} \sum_{nml} c^j_{nlm}(\Gamma) g_n(r) Y_{lm}(\theta,\phi), \tag{1}$$

where c^j_{nlm} are the spherical harmonics and radial functions' expansions coefficient, with that *j* runs over all the sites within r_{cut}.

The SOAP power spectrum vector can be analytically derived[35] from Eq. (1) and can be written as

$$p(\Gamma)_{nn'l} = \pi \sqrt{\frac{8}{2l+1}} \sum_{m=-l}^{l} c^*_{nlm}(\Gamma) cn'lm(\Gamma). \tag{2}$$

Once SOAP feature vectors have been calculated, the similarity between the various atomic environments can be inferred via a

© Author(s) 2023

distance metric, as long as the dimensionality of the compared SOAP vectors is the same, as in our case, where we used for $n_{\max}$, $l_{\max} = 8$ (found effective and sensitive enough in our case).

More precisely, we can define a measure of the similarity between two environments centered in two sites, building a linear polynomial kernel of their density representations; this can be simply reduced to the dot product of power spectra, defined in Eq. (2). The SOAP distance between two SOAP spectra $\vec{a}$ and $\vec{b}$ can be calculated by

$$d_{SOAP}\left(\vec{a},\vec{b}\right) = \sqrt{2 - 2\mathscr{K}\left(\vec{a},\vec{b}\right)}, \tag{3}$$

where, with the SOAP power spectrum representation that we are using, $\mathscr{K}\left(\vec{a},\vec{b}\right) = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\|\vec{b}\|}$.

For all systems, the SOAP descriptors were calculated under periodic boundary conditions along the xy dimensions using $n_{\max} = 8$ radial basis function and $l_{\max} = 8$ maximum number of spherical harmonics. The choice of the cutoff radius (r_{cut}) determines the size and shape of the neighborhood considered in characterizing the atomic environment for each SOAP center; in this work, we opted for $r_{\text{cut}} = 6.0$ Å, the same cutoff adopted for the training of the Cu *NN* potential. The remaining parameters were set to default values of the DScribe library[103] employed in this analysis.

D. Bottom-up SOAP analysis

In the bottom-up production analyses, we did not include the two bottom-most atomic layers (kept fixed) and the three closest neighbor layers of the bulk, thus considering the first 1,300 atoms of the Cu(211). This was done in order to guarantee the correct analysis of the atomic dynamics in the simulated metal surface. We extracted the characteristic SOAP spectra for each of the 1,300 Cu atoms in the surfaces from 250 snapshots taken from the last 75 ns of DPMD simulations (one every $\Delta t = 300$ ps of simulation), corresponding in all simulated cases to the equilibrated phase of the DPMD simulations. In total, for the bottom-up analysis of Fig. 3, we obtain a dataset of 325 000 SOAP power spectra in total.

The SOAP data output obtained from the DPMD trajectories consists of a set of feature vectors of high dimensionality that, while rich in information, are not convenient for visualization and classification. To overcome these limitations, we opted to employ a widely used dimensionality reduction method, performing the Principal Component Analysis (PCA) algorithm in the implementation from scikit–learn,[104] and to project the normalized SOAP vector on the principal components (PCs). The data projected on the PCs are easier to visualize [we used the first two PCs in Figs. 3(c), 3(d), 3(f), and 3(g)] and are convenient for finding patterns in the data. More precisely, in the cases reported herein, the first three PCs account for >98 % of the total variance in the simulated systems.

The SOAP data extracted from the last 75 ns of the DPMD simulations (equilibrated phase trajectories) have been then analyzed to identify the most visited atomic environments in these systems. In particular, unsupervised clustering of the SOAP data has been performed using HDBSCAN*.[78] Density-based clustering algorithms identify clusters based on a search of high-density peaks surrounded by regions where there is a lower density of points. The minimum number of points in a neighborhood for a point to be recognized as a core point (min_samples) did not impact the results. In particular, the results of Fig. 3 are obtained using the default option, with automatic set to the same value as min_cluster_size and the default excess of mass (eom) as the cluster_selection_method. As the clustering analysis is completely unsupervised, the detected clusters emerge bottom-up directly from the second half (equilibrated phase) of the DPMD trajectories.

E. Dictionary of SOAP atomic environments and top-down analysis

Figures 4–6 show a top-down SOAP-based classification method that allows tracking and classifying all the atomic environments that appear in the systems during the equilibrium DPMD trajectories based on their similarity with reference environments inserted in a SOAP dataset. As a first step, we minimized the reference structures of Cu(111), (110), (210), and (211) surface slabs. For all these minimized reference surfaces, we calculate the characteristic SOAP spectra of the atomic environments that compose them. In this case, we obtained a total of 22 characteristic SOAP spectra—Fig. 4: three for (111), five for (110), seven for (210), and seven for (211). All these SOAP spectra have been then inserted into a unique SOAP dictionary of atomic environments and used for our classification analysis. Noteworthy, this approach is flexible and general, and the dictionary can be, in principle, expanded by adding other SOAP spectra characteristic of other atomic environments if one wants to also compare other surface systems. We then hierarchically classified all of the dictionary entries using the SOAP distance[36] and by using the *complete* method of the linkage algorithm, which conserves the greater distances between the newly formed cluster and the other elements. In this analysis, our main aim is to compare the local atomic environments that emerge during the DPMD simulations and to classify these environments using the dictionary described in the dendrogram of Fig. 4(c) as a reference (on the right, in black: for this analysis, we use a criterion of $d_{SOAP} \geq 0.01$). For each atom in each simulation, we evaluated the SOAP distance (d_{SOAP}) from each dictionary entry. We then assigned each atom to the dictionary element with the lowest distance from it (i.e., smallest d_{SOAP}). We applied this process to each atom [2,304 for the Cu(210) and 2,400 for all other surfaces] along the entire 150 ns of DPMD simulations (501 frames, one every $\Delta t = 300$ ps), thus obtaining 1 202 800 SOAP spectra for Cu(111), (211), and (110) surfaces, and 1 154 304 SOAP spectra for Cu(210). This allowed us to monitor the microscopic equilibration in all surface models [Figs. 5 and 6(b)]. For all cases, we took the second half of the simulations (the last 75 ns and 250 frames of DPMD) as representative of the equilibrium configuration of the surfaces. The transition matrices of Figs. 5 and 6(b) are calculated by accumulating for each environment the number of changes to other clusters (or to self), and then by normalizing each row in the matrix to have it summing to 100, obtaining conditional probabilities for atoms belonging to a given class of environments to undergo transition into a new environments class (o to remain self) in $\Delta t = 300$ ps. Since in this analysis we are primarily interested in the mutability and transformations of the surface, only the SOAP data for the first top-most surface (s) layers of atoms are retained, while all atoms belonging to bulk or to the ss^h environments at DPMD simulation start (in the ideal surfaces) are not considered in the analysis.

© Author(s) 2023

SUPPLEMENTARY MATERIAL

The supplementary material contains details about the validation of the *NN* potential; time-series of the populations of the clusters determined by the bottom-up SOAP analysis for all the surfaces investigated; a coordination and top-down SOAP-based analyses of the reconstructed (110) (1× 2), (110) (1× 3), and (211) (2× 1) surfaces; the radial distribution function of the (111), (110), (211), and (210) surfaces at T = 0, 500, and 700 K; and additional details on the bottom-up and top-down SOAP-based analyses.

ACKNOWLEDGMENTS

G.M.P. acknowledges the funding received by the ERC under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 818776-DYNAPOL) and by the Swiss National Science Foundation (SNSF Grant No. IZLIZ2_183336). The authors also acknowledge the computational resources provided by the Swiss National Supercomputing Center (CSCS), CINECA, and HPC@POLITO (http://www.hpc.polito.it).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

M.C., D.P., and M.D.P. developed the molecular models. M.C. performed the simulations. M.C., D.P., D.R., L.P., M.D.P., and G.M.P. analyzed the results. G.M.P. conceived this research and supervised the work. M.C., D.P., D.R., M.D.P., and G.M.P. wrote the manuscript.

Matteo Cioni: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Writing – original draft (equal). **Daniela Polino**: Conceptualization (supporting); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Supervision (supporting); Writing – original draft (equal). **Daniele Rapetti**: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (lead); Supervision (supporting). **Luca Pesce**: Data curation (supporting); Investigation (supporting); Resources (supporting); Supervision (supporting); Validation (supporting); Writing – review & editing (supporting). **Massimo Delle Piane**: Data curation (supporting); Investigation (supporting); Resources (supporting); Supervision (supporting); Validation (supporting); Writing – original draft (equal); Writing – review & editing (supporting). **Giovanni M. Pavan**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Details on the procedures for the training of the neural network potential and the simulations' setup, along with additional simulation data, are provided in Sec. IV and in the supplementary material file. Complete data and materials pertaining to the molecular simulations and data analyses conducted herein (input files, model files, raw data, analysis tools, etc.) are available at https://zenodo.org/record/7738141#.ZBH5fHbMJPY. The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] M.-C. Desjonqueres and D. Spanjaard, *Concepts in Surface Physics*, 2nd ed. (Springer Science & Business Media, 1996), Vol. 30.

[2] M. S. Spencer, Nature **323**, 685 (1986).

[3] C. S. Jayanthi, E. Tosatti, and L. Pietronero, Phys. Rev. B **31**, 3456 (1985).

[4] M. Bernasconi and E. Tosatti, Surf. Sci. Rep. **17**, 363 (1993).

[5] A. Trayanov, A. C. Levi, and E. Tosatti, Surf. Sci. **233**, 184 (1990).

[6] H. Dürr, R. Schneider, and T. Fauster, Phys. Rev. B **43**, 12187 (1991).

[7] J. Merikoski, H. Häkkinen, M. Manninen, J. Timonen, and K. Kaski, Int. J. Mod. Phys. B **8**, 3175 (1994).

[8] O. Zeybek, Solid State Commun. **139**, 339 (2006).

[9] S. Xie, S.-I. Choi, X. Xia, and Y. Xia, Curr. Opin. Chem. Eng. **2**, 142 (2013).

[10] F. Dattila, R. Garcıá-Muelas, and N. López, ACS Energy Lett. **5**, 3176 (2020).

[11] V. Yamakov, D. Wolf, S. R. Phillpot, A. K. Mukherjee, and H. Gleiter, Nat. Mater. **3**, 43 (2004).

[12] L. A. Zepeda-Ruiz, A. Stukowski, T. Oppelstrup, and V. V. Bulatov, Nature **550**, 492 (2017).

[13] X. Wang, S. Zheng, S. Shinzato, Z. Fang, Y. He, L. Zhong, C. Wang, S. Ogata, and S. X. Mao, Nat. Commun. **12**, 5237 (2021).

[14] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, Nat. Chem. **1**, 37 (2009).

[15] J. K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard, Proc. Natl. Acad. Sci. U. S. A. **108**, 937 (2011).

[16] F. Calle-Vallejo, J. Tymoczko, V. Colic, Q. H. Vu, M. D. Pohl, K. Morgenstern, D. Loffreda, P. Sautet, W. Schuhmann, and A. S. Bandarenka, Science **350**, 185 (2015).

[17] M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.-T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C.-S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S.-C. Lo, A. Ip, Z. Ulissi, and E. H. Sargent, Nature **581**, 178 (2020).

[18] B. W. J. Chen, L. Xu, and M. Mavrikakis, Chem. Rev. **121**, 1007 (2021).

[19] E. Gazzarrini, K. Rossi, and F. Baletto, Nanoscale **13**, 5857 (2021).

[20] D. Nelli, F. Pietrucci, and R. Ferrando, J. Chem. Phys. **155**, 144304 (2021).

[21] M. Crippa, C. Perego, A. L. de Marco, and G. M. Pavan, Nat. Commun. **13**(1), 2162 (2022).

[22] P. Gasparotto, D. Bochicchio, M. Ceriotti, and G. M. Pavan, J. Phys. Chem. B **124**, 589 (2019).

[23] A. Gardin, C. Perego, G. Doni, and G. M. Pavan, Commun. Chem. **5**(1), 82 (2022).

[24] D. Bochicchio, M. Salvalaglio, and G. M. Pavan, Nat. Commun. **8**(1), 147 (2017).

[25] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[26] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[27] S. Doerr, M. Majewski, A. Pérez, A. Krämer, C. Clementi, F. Noe, T. Giorgino, and G. De Fabritiis, J. Chem. Theory Comput. **17**, 2355 (2021).

[28] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018).

[29] O. T. Unke and M. Meuwly, J. Chem. Theory Comput. **15**, 3678 (2019).

[30] Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. **114**, 096405 (2015).

[31] H. Wang, L. Zhang, J. Han, and W. E, Comput. Phys. Commun. **228**, 178 (2018).

[32] G. P. P. Pun, R. Batra, R. Ramprasad, and Y. Mishin, Nat. Commun. **10**, 2339 (2019).

[33] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, J. Phys. Chem. A **124**, 731 (2020).

[34] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Chem. Rev. **121**, 10142 (2021).

[35] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).

[36] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. **18**, 13754 (2016).

[37] R. Capelli, F. Muniz-Miranda, and G. M. Pavan, J. Chem. Phys. **156**, 214503 (2022).

[38] T. Bian, A. Gardin, J. Gemen, L. Houben, C. Perego, B. Lee, N. Elad, Z. Chu, G. M. Pavan, and R. Klajn, Nat. Chem. **13**, 940 (2021).

[39] R. Capelli, A. Gardin, C. Empereur-Mot, G. Doni, and G. M. Pavan, J. Phys. Chem. B **125**, 7785 (2021).

[40] C. Zeni, K. Rossi, T. Pavloudis, J. Kioseoglou, S. de Gironcoli, R. E. Palmer, and F. Baletto, Nat. Commun. **12**(1), 6056 (2021).

[41] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[42] S. Nitopi, E. Bertheussen, S. B. Scott, X. Liu, A. K. Engstfeld, S. Horch, B. Seger, I. E. Stephens, K. Chan, C. Hahn *et al.*, Chem. Rev. **119**, 7610 (2019).

[43] K. Manthiram, B. J. Beberwyck, and A. P. Alivisatos, J. Am. Chem. Soc. **136**, 13319 (2014).

[44] A. J. Hickman and M. S. Sanford, Nature **484**, 177 (2012).

[45] S. Magdassi, M. Grouchko, and A. Kamyshny, Materials **3**, 4626 (2010).

[46] L. Lu, Y. Shen, X. Chen, L. Qian, and K. Lu, Science **304**, 422 (2004).

[47] M. Behrens, F. Studt, I. Kasatkin, S. Kühl, M. Hävecker, F. Abild-Pedersen, S. Zander, F. Girgsdies, P. Kurr, B.-L. Kniep *et al.*, Science **336**, 893 (2012).

[48] M. D. Higham, M. G. Quesne, and C. R. A. Catlow, Dalton Trans. **49**, 8478 (2020).

[49] R. P. Gupta, Phys. Rev. B **23**, 6265 (1981).

[50] S. M. Foiles, M. I. Baskes, and M. S. Daw, Phys. Rev. B **33**, 7983 (1986).

[51] V. Rosato, M. Guillope, and B. Legrand, Philos. Mag. A **59**, 321 (1989).

[52] M. I. Baskes, Phys. Rev. B **46**, 2727 (1992).

[53] B.-J. Lee, J.-H. Shim, and M. I. Baskes, Phys. Rev. B **68**, 144112 (2003).

[54] M. I. Mendelev, M. J. Kramer, C. A. Becker, and M. Asta, Philos. Mag. **88**, 1723 (2008).

[55] L. Zhang, J. Han, H. Wang, R. Car, and W. E, Phys. Rev. Lett. **120**, 143001 (2018).

[56] V. L. Deringer and G. Csányi, Phys. Rev. B **95**, 094203 (2017).

[57] Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang, and W. E, Comput. Phys. Commun. **253**, 107206 (2020).

[58] C. Kittel and P. McEuen, *Introduction to Solid State Physics* (John Wiley & Sons, 2018).

[59] E. A. Brandes and G. Brook, *Smithells Metals Reference Book* (Elsevier, 2013).

[60] W. R. Tyson and W. A. Miller, Surf. Sci. **62**, 267 (1977).

[61] J. Merikoski, I. Vattulainen, J. Heinonen, and T. Ala-Nissila, Surf. Sci. **387**, 167 (1997).

[62] J. Merikoski and T. Ala-Nissila, Phys. Rev. B **52**, R8715 (1995).

[63] L. S. Perkins and A. E. DePristo, Surf. Sci. **294**, 67 (1993).

[64] G. Boisvert and L. J. Lewis, Phys. Rev. B **56**, 7643 (1997).

[65] C. L. Liu, J. M. Cohen, J. B. Adams, and A. F. Voter, Surf. Sci. **253**, 334 (1991).

[66] B. D. Yu, and M. Scheffler, Phys. Rev. B **56**, R15569(R) (1997).

[67] H. Dürr, J. F. Wendelken, and J.-K. Zuo, Surf. Sci. **328**, L527 (1995).

[68] H.-J. Ernst, F. Fabre, and J. Lapujoulade, Phys. Rev. B **46**, 1929 (1992).

[69] L. Hansen, P. Stoltze, K. W. Jacobsen, and J. K. Nørskov, Phys. Rev. B **44**, 6523 (1991).

[70] F. Montalenti and R. Ferrando, Phys. Rev. B **59**, 5881 (1999).

[71] K. H. Chae, H. C. Lu, and T. Gustafsson, Phys. Rev. B **54**, 14082 (1996).

[72] A. N. Al-Rawi, A. Kara, and T. S. Rahman, Phys. Rev. B **66**, 165439 (2002).

[73] L. Yang and T. S. Rahman, Phys. Rev. Lett. **67**, 2327 (1991).

[74] A. Trayanov, A. C. Levi, and E. Tosatti, Europhys. Lett. **8**, 657 (1989).

[75] F. Studt, M. Behrens, E. L. Kunkes, N. Thomas, S. Zander, A. Tarasov, J. Schumann, E. Frei, J. B. Varley, F. Abild-Pedersen, J. K. Nørskov, and R. Schlögl, ChemCatChem **7**, 1105 (2015).

[76] G. Witte, J. Braun, D. Nowack, L. Bartels, B. Neu, and G. Meyer, Phys. Rev. B **58**, 13224 (1998).

[77] R. E. Kirby, C. S. McKee, and L. V. Renny, Surf. Sci. **97**, 457 (1980).

[78] R. J. Campello, D. Moulavi, and J. Sander, in *Advances in Knowledge Discovery and Data Mining*, edited by J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu (Springer Berlin Heidelberg, 2013) pp. 160–172.

[79] R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander, ACM Trans. Knowl. Discovery Data **10**, 1 (2015).

[80] D. M. Zuckerman, *Statistical Physics of Biomolecules: An Introduction* (CRC Press, 2010).

[81] D. E. Makarov, *Single Molecule Science: Physical Principles and Models* (CRC Press, 2015).

[82] P. B. Gupta, C. M. Fillmore, G. Jiang, S. D. Shapira, K. Tao, C. Kuperwasser, and E. S. Lander, Cell **146**, 633 (2011).

[83] Z. Jian-Min, M. Fei, and X. Ke-Wei, Chin. Phys. **13**, 1082 (2004).

[84] J. Wang and S.-Q. Wang, Surf. Sci. **630**, 216 (2014).

[85] S. Titmuss, A. Wander, and D. A. King, Chem. Rev. **96**, 1291 (1996).

[86] R. Koch, M. Borbonus, O. Haase, and K. H. Rieder, Appl. Phys. A **55**, 417 (1992).

[87] R. Koch, M. Sturmat, and J. J. Schulz, Surf. Sci. **454–456**, 543 (2000).

[88] P. Grosse, D. Gao, F. Scholten, I. Sinev, H. Mistry, and B. Roldan Cuenya, Angew. Chem., Int. Ed. **57**, 6192 (2018).

[89] Y. Li, D. Kim, S. Louisia, C. Xie, Q. Kong, S. Yu, T. Lin, S. Aloni, S. C. Fakra, and P. Yang, Proc. Natl. Acad. Sci. U. S. A. **117**, 9194 (2020).

[90] G.-C. Wang, L. Jiang, Y. Morikawa, J. Nakamura, Z.-S. Cai, Y.-M. Pan, and X.-Z. Zhao, Surf. Sci. **570**, 205 (2004).

[91] L. Delgado-Callico, K. Rossi, R. Pinto-Miles, P. Salzbrenner, and F. Baletto, Nanoscale **13**, 1172 (2021).

[92] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo *et al.*, J. Phys.: Condens. Matter **21**, 395502 (2009).

[93] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[94] C. M. Andolina, M. Bon, D. Passerone, and W. A. Saidi, J. Phys. Chem. C **125**, 17438 (2021).

[95] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, 1227 (1990).

[96] N. Marzari, D. Vanderbilt, A. D. Vita, and M. C. Payne, Phys. Rev. Lett. **82**, 3296 (1999).

[97] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[98] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).

[99] D. P. Kingma and J. Ba, arXiv:1412.6980 (2014).

[100] F. Cyrot-Lackmann and F. Ducastelle, Phys. Rev. B **4**, 2406 (1971).

[101] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, Comput. Phys. Commun. **271**, 108171 (2022).

[102] A. Stukowski, Modell. Simul. Mater. Sci. Eng. **18**, 015012 (2009).

[103] L. Himanen, M. O. J. Jäger, E. V. Morooka, F. F. Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, Comput. Phys. Commun. **247**, 106949 (2020).

[104] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).

© Author(s) 2023