Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Material Science (36^{th}cycle)

Exploring the Importance of Dynamics in Materials from the Atomic to the Supramolecular Scale Using Advanced Computational Methods

By

Matteo Cioni

******

Supervisor(s):
Prof. Giovanni Maria Pavan

Doctoral Examination Committee:
Prof. Piero Ugliengo, Referee, University of Torino, Torino, IT
Prof. Maria Clelia Righi, Referee, University of Bologna, Bologna IT

Politecnico di Torino
2024

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Matteo Cioni
2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Contents

List of Figures

List of Tables

List of Acronyms

AA All-Atom

ADF-STEM Annular Dark Field Scanning Transmission Electron Microscopy

AE Atomic Environment

AIMD Ab-initio Molecular Dynamics

a-CNA adaptive Common Neighbor Analysis

CSP Centrosymmetry parameter

CV Collective Variable

DFT Density Functional Theory

DPMD Deep Potential Molecular Dynamics

EAM Embedded-Atom Method

FES Free Energy Surface

FCC Face-centered cubic

GAFF Generalized AMBER Force Field

GGA Generalized Gradient Approximation

GNN Graph Neural Network

HC Hierarchical Clustering

ID Identity

LDA Local Density Approximation

LENS Local Environments and Neighbors Shuffling

LJ Lennard-Jones

MD Molecular Dynamics

MetaD Metadynamics

ML Machine Learning

NMA N-Methylacetamide

NN Neural Network

NP Nanoparticle

NPT Isothermal-Isobaric Ensemble

NVE Micro-Canonical Ensemble

NVT Canonical Ensemble

PBC Periodic Boundary Conditions

PES Potential Energy Surface

PCA Principal Component Analysis

SASA Solvent Accessible Surface Area

SCC-DFTB Self-Consistent Charge Density Functional Tight Binding

SOAP Smooth Overlap of Atomic Positions

TS Transition State

WT-MetaD Well-Tempered Metadynamics

Chapter 1

Introduction

1.1 Metals

In the field of materials science, metals constitute a fundamental class of material that possess unique properties and that have been crucial in technological and industrial advancements. The distinctive attributes of metals, such as high electrical and thermal conductivity, prominent mechanical properties, ductility, and malleability, are highly related to their atomic structure and to the nature of the metallic bonding[1–5]. The unique characteristics of metallic bonding arise from the fundamental properties of metal atoms, which distinguish them from nonmetals. Metal atoms typically possess a high number of electrons, yet only a few reside in the valence shell. Indeed, these valence electrons are relatively distant from the nucleus and are shielded by numerous inner-shell electrons, resulting in a weak binding force. Consequently, metal atoms are unlikely to form one-to-one covalent bonds. Instead, they organize into crystalline lattice structures, where multiple neighboring atoms surround each atom. In these configurations, the valence electrons are not localized into the individual atoms but are shared collectively across the entire lattice, creating a "sea of electrons." This delocalization enables the electrons to move freely throughout the lattice, which accounts for the exceptional thermal and electrical conductivity characteristic of metals[6–9]. Furthermore, these delocalized electrons interact with positively charged metal ions through electrostatic forces, creating a cohesive lattice structure with characteristic ductility and malleability, which maintain its integrity even under significant mechanical stress. All these peculiar features make metallic bonding very different from other types of chemical bonds.

The strength of metallic bonds, rooted in the electrostatic attraction between metal ions and the delocalized electrons, ensures that metals can withstand plastic deformation before breaking. This ductility is a key feature that makes metals indispensable for applications where materials are subjected to varying levels of stress and strain[10].

The sharing of electrons among metal atoms stabilizes the lattice structure while simultaneously enhancing electrical conductivity. This shared electron cloud facilitates the movement of electrons and electric conductivity through the lattice, enabling efficient electrical conduction with minimal resistance (ohmic conductors)

Understanding the intricacies of metallic bonding not only provides insights into the fundamental properties of metals but also distinguishes them from other materials. This understanding is essential for advancing scientific knowledge and driving innovation across a wide range of technological applications.

1.1.1 Metals' role in modern technological applications

Metals play an important role across a wide range of technological applications, showcasing their versatility in modern industry and technology. The inherent properties of metals, such as strength, ductility, and conductivity, are not only fundamental to their application but also highlight the dynamic interplay between their structural characteristics and functional performance[1–3]. This short overview provides insight into the modern technological applications of metals across various sectors.

Metals are essential to technological advancements in many industries, serving as the foundation for modern engineering applications. In the aerospace sector, metals are chosen for their strength-to-weight ratio, crucial for airframes and propulsion systems[11, 12]. Alloys like titanium and aluminum are valued for their weight and high mechanical strength, leading to more fuel-efficient and safe aircraft.[11–14]

In electronics, metals such as copper and silver are indispensable for their excellent electrical properties[9, 14–16]. Their high electrical conductivity is essential for reliable electrical circuits and connectors, ensuring consistent performance in electronic devices.

The automotive industry relies on metals for their structural integrity and durability. High-strength steels are used for their impact resistance in critical components like chassis and brakes, meeting safety and durability standards[17–19].

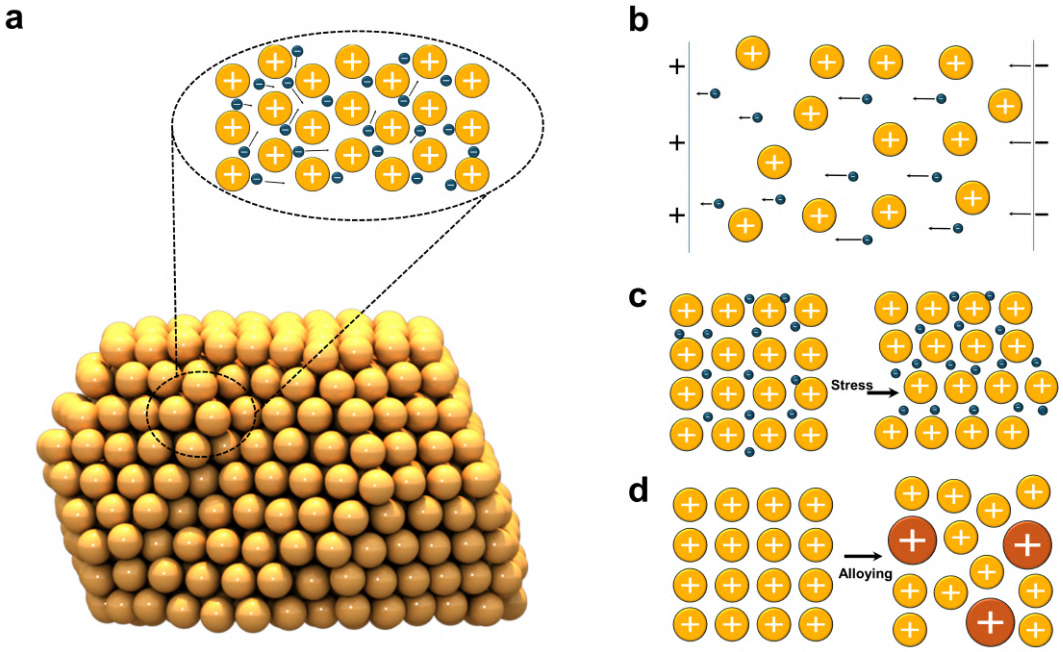


Fig. 1.1 **Illustration of metallic bonding and related phenomena (a)** presents an enlarged view of the 'electron sea' model within a metallic lattice, highlighting the fluid mobility of electrons amongst the positively charged metal ions, a key factor in the conductivity and malleability of metals. **(b)** depicts the mechanism of electrical conductivity, characterized by the free flow of electrons across the lattice, facilitating electrical current. **(c)** captures the deformation of the metallic lattice under applied mechanical stress, displaying both distortion and subsequent realignment of ions that contribute to the metal's ductility. Finally, **(d)** shows the process of alloying, where the introduction and substitution of different metallic elements alter the lattice structure and improve the material's overall properties.

In renewable energy, metals are crucial for constructing wind turbines and photovoltaic panels. Alloys of steel, iron, and copper provide the necessary durability and efficient energy transmission, ensuring long-lasting and effective renewable energy installations.[20–22] Metals combine strength and ductility, allowing them to undergo plastic deformation while maintaining structural integrity[23–28]. Alloying, the process of combining metals with other elements, enhances properties like stiffness, strength, and corrosion resistance, resulting in materials with unique characteristics[29–32].

Metal drawing is a key manufacturing process for producing wires, rods, and tubes. This plastic deformation process involves compressing the metal's cross-sectional area and extending its length, reshaping the metal efficiently[33, 34].

In this context, the concept of tribology—the science of wear, friction, and lubrication—becomes crucial. Tribology in metals is vital in numerous industrial applications where metals undergo continuous mechanical interaction, either with other metals or different materials. The wear and tear metals endure, the friction they encounter in different environments, and the need for effective lubrication to minimize damage are all factors that significantly influence their performance and lifespan [35, 36].

Understanding the tribological properties of metals is crucial for designing more durable and efficient mechanical systems[37, 38]. In the automotive industry, the wear resistance of engine components, the efficiency of transmission systems, and the reliability of braking systems all rely on the tribological properties of the metals used. Likewise, in aerospace, the performance and safety of aircraft heavily depend on the tribological characteristics of their metal components, which must endure extreme pressure, temperature, and mechanical stress conditions. Furthermore, advancements in nanotechnology have facilitated the creation of nanostructured surfaces with tailored tribological properties. Nanoscale surface features such as nanocrystalline grains, nanotubes, and nanoporous structures have shown promise in reducing friction and enhancing wear resistance by modifying the contact mechanics at the surface interface.

In addition to their crucial role in tribology, metals are also fundamental in the field of heterogeneous catalysis. Heterogeneous catalysis is a process in which the catalyst and the reactants exist in different phases, most commonly involving solid catalysts and gaseous or liquid reactants. This type of catalysis is highly valued for its efficiency and selectivity in facilitating chemical transformations, which are essential in numerous industrial processes. The significance of metals in heterogeneous catalysis arises from their electronic structures and surface geometries, which enable them to efficiently adsorb and activate reactant molecules. Transition metals are particularly effective catalysts due to their partially filled d-orbitals. These d-orbitals can engage in a variety of bonding interactions with the orbitals of reactant molecules, which is critical for several reasons.

Firstly, the partially filled d-orbitals of transition metals can accommodate electrons from adsorbed molecules, weakening the bonds within these molecules and making them more reactive. This is often described by the d-band model, which explains that the energy levels of the d-electrons play a critical role in determining the strength of adsorption and the activation energy of the reaction. For example,

metals like platinum (Pt) and nickel (Ni) have d-band centers that are well-suited for interacting with a variety of reactant molecules, thus lowering the activation energy and increasing the reaction rate.

Secondly, the d-orbitals of transition metals can overlap with the orbitals of reactant molecules, forming transient bonds. This interaction brings the reactants into close proximity and correctly orients them for the reaction. For instance, in heterogeneous catalysis, the reactant molecules are adsorbed onto the surface of the metal catalyst. The d-orbitals of the transition metal interact with the electron clouds of the reactants, stabilizing the transition state and lowering the activation energy required for the reaction to proceed.[39, 40]

Additionally, the variable oxidation states of transition metals allow them to participate in redox reactions, alternating between different oxidation states. This ability is essential in processes such as the Haber-Bosch process for ammonia synthesis, where Fe acts as a catalyst by facilitating the conversion of nitrogen (N_2) and hydrogen (H_2) into ammonia (NH_3) through repeated oxidation and reduction cycles. Also copper surfaces demonstrate exceptional catalytic efficiency in the synthesis of methanol from carbon dioxide and hydrogen. The unique electronic structure of copper enhances the adsorption and activation of the reactants, facilitating the reaction under milder conditions than traditional stoichiometric processes [41–46]. This shows the intricate relationship between surface geometry and catalytic activity, where the unique electronic structure of copper enhances reactant adsorption and activation, allowing for reactions to proceed under significantly milder conditions than traditional stoichiometric processes [47, 48]. Transition metals can also form complex compounds with ligands, stabilizing reactant intermediates through coordination bonds, thus increasing the reaction rate. For example, in homogeneous catalysis, transition metal complexes, such as those involving Pd in the Heck reaction, can activate alkene substrates through coordination with the metal center, making the reactants more reactive towards nucleophilic attack [49].

Furthermore, the surface geometry of metals significantly impacts their catalytic properties.[39] Metal surfaces are composed of various atomic arrangements, including terraces, steps, and kink sites, each offering different catalytic properties. The availability of these different sites means that metal surfaces can provide multiple pathways for reactions, enhancing overall catalytic activity. For instance, the step and kink sites on a metal surface are often more reactive than the flat terrace sites

because they have under-coordinated atoms, which are more likely to form bonds with reactant molecules[50].

On the other hand, nanoparticles, namely particles with sizes between 1 and 100 nanometers, have significantly advanced the field of catalysis due to their unique properties. These systems are more effective than larger bulk materials in catalytic processes mainly because of their greater surface area relative to their volume and more sites available for chemical reactions. This increased surface area allows for more reactant molecules to interact with the catalyst, improving the efficiency and speed of the reactions.[41, 42, 51–54] For instance, gold nanoparticles can catalyze the oxidation of carbon monoxide at room temperature, demonstrating abilities that go beyond the expected chemical activity of bulk gold.[55–57] The strategic design of these nanoparticle catalysts, grounded in the Sabatier principle [58] and enhanced through site coordination strategies, is instrumental in crafting nanocatalysts with tailored reactivity and selectivity.

By adjusting the electronic structure and surface configuration of nanoparticles, it is possible to control catalytic behavior, optimizing efficiency and selectivity for specific reactions.[54] Furthermore, the concept of site coordination plays a vital role in fine-tuning catalytic activity. By strategically arranging and electronically tuning the catalyst's active sites, it is possible to directly influence the reaction's activation energy and the orientation of reactants, steering the reaction kinetics and pathways. This level of control at the nanoscale allows for the facilitation of complex reactions that were once considered unfeasible, broadening the scope of catalytic chemistry.[59–61] For example, in the production of pharmaceuticals, nanoparticles engineered with specific site coordination can catalyze the synthesis of complex molecules with high purity and fewer by-products.[62–64] In environmental applications, catalysts designed at the nanoscale are used to efficiently convert harmful pollutants like nitrogen oxides in automotive exhaust into harmless nitrogen and water, showcasing a significant reduction in environmental impact.[65] Additionally, in the energy sector, nanoparticles with tailored electronic properties enhance the conversion efficiency of solar cells by optimizing light absorption and charge separation processes[66–68], demonstrating the broad applicability of these advanced catalytic systems [69, 70]

In situ characterization techniques are indispensable in this context, to provide deep insights into the active sites and underlying mechanisms of catalytic processes. These techniques enable real-time investigation of catalysts under experimentally-

relevant conditions, revealing the dynamic interaction between the catalyst's structure and its performance[71–73].

This quick overview of the technological applications of metals highlights the extensive influence metals have on modern technology and industry. By highlighting specific applications and the role of metals within them, we not only appreciate their versatility but also understand the critical need for continuous research and development in material science, with a constant pursuit to optimize metal properties for current and future applications, ensuring their relevance and sustainability in addressing global challenges.

1.1.2 Shifting perspectives: the dynamic nature of metals under experimentally-relevant conditions

The field of materials science has increasingly recognized the importance of understanding the behavior of metals under operational conditions. Before the melting temperature, metals are perceived as a perfect and static lattice structure, and their theoretical models are built on this assumption. However, real-world applications expose metals to many environmental stresses that lead to significant deviations from these idealized structures [74–77]. Advances in *operando* characterization techniques have allowed the observation of these deviations in real-time, shedding light on the structural transformations that occur and how they impact the material's properties [78–80].

For instance, Shi *et al.*(2021) emphasized the importance of *operando* conditions in understanding the real-time dynamics of heterogeneous catalytic processes. They highlighted how these techniques provide insights into reaction mechanisms and active sites that would be impossible to observe under *ex situ* conditions. Their work shows that advanced characterization techniques reveal dynamic structural changes that directly correlate with catalytic performance[79]. Understanding metals in *operando* conditions is thus essential to unlocking their full potential and understanding how changes in structure affect their performance. This need stems from the differences often seen between how metals behave in real-world situations compared to how they are expected to behave according to theoretical models [74, 81, 82]. In experimental setups, metals face various environmental factors such as temperature, pressure, and exposure to reactants that can drastically change their structure [83–87, 81]. This is different from theoretical models and simulations, where met-

als are often assumed to have perfect and static crystal structures. Even at lower temperatures, metals can undergo significant structural changes under *operando* conditions, which can influence their mechanical and catalytic properties; Yang *et al.* (2021) demonstrated that advanced *operando* methods enable direct visualization of changes in metal catalysts under realistic conditions. Their work showed that surface restructuring, defect formation, and lattice relaxation significantly impact catalytic efficiency and selectivity [87].

Colic and Bandarenka (2016) also pointed out that the surface structures of metal alloys change significantly in real-world conditions, which are different from idealized models. They showed that varying temperatures and pressures create stress and rearrange the atoms, leading to unexpected behaviors [81]. As this thesis progresses, we will introduce the mechanisms of structural changes, such as defect formation, lattice relaxation, and surface reconstruction, and explore their effects on structural-related properties. This examination aims to highlight the intricacies of metal behavior under *operando* conditions, emphasizing the critical role of dynamic structural transformations in advancing material science and engineering for diverse applications. This process is crucial for understanding how surface properties differ from bulk properties and how these differences impact the overall behavior of the material. By exploring surface relaxation in detail, we can gain insights into the initial steps of surface reconstruction, setting the foundation for a deeper understanding of how metals adapt to their environments in real-world conditions.

1.1.3 Surface relaxation

Every crystalline solid consists of crystallographic lattice planes arranged periodically in space. Within the bulk of the material, these planes exhibit a consistent structure when separated by a multiple of the crystal's spatial period. However, near the surface, the arrangement and crystallographic properties of these planes, as well as the behavior of constituent atoms, can differ significantly.

Surface relaxation refers to the rigid shift of one or more entire lattice planes near the surface relative to their equilibrium positions in the bulk. This displacement, which can be either parallel or perpendicular to the surface, occurs as the system responds to the imbalance of forces created when a surface is cleaved from the crystal [88]. This process is illustrated in Figure 1.2, which shows the atomic positions in a five-layer metal slab model before and after relaxation. The left side of the figure

depicts the original slab with atoms at bulk-like positions, while the right side shows the model after the top three layers have relaxed. Surface relaxation generally leads to a lower energy state compared to the original unreconstructed surface.

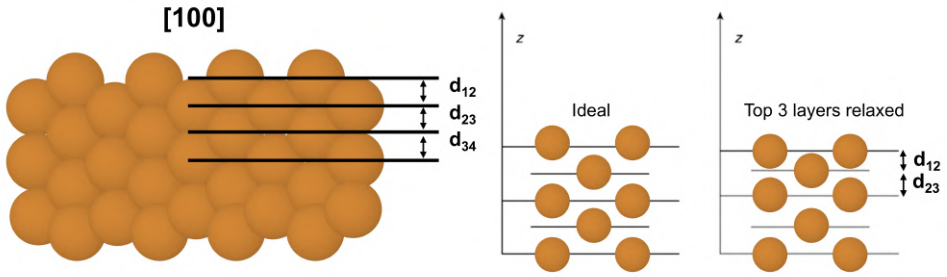


Fig. 1.2 **Schematic illustration of relaxation of surface atoms in a slab model**. The top three layers were allowed to relax, while the bottom two layers remained fixed at ideal, bulk positions.

The abrupt termination of a surface significantly modifies the atomic coordination, often leading to a reduction in the distance between the first and second atomic layers, denoted as d_{12}. The variation in this distance, Δd_{12}, is expressed as a percentage relative to the interlayer spacing in the bulk material. Table 1.1 presents the interlayer spacing relaxations for Cu(100) and Cu(111) surfaces, computed using DFT for various slab thicknesses.

	Cu(100)			Cu(111)		
Layers	$\delta\ d_{12}$ (%)	$\delta\ d_{23}$ (%)	$\delta\ d_{34}$ (%)	$\delta\ d_{12}$ (%)	$\delta\ d_{23}$ (%)	$\delta\ d_{34}$ (%)
5	3.84	0.50	0.53	0.61	0.08	+0.08
6	1.93	+0.83	+0.37	0.64	0.11	+0.27
7	2.30	+0.55	-0.25	0.56	0.04	+0.32
8	2.14	+0.85	+0.00	0.59	0.32	+0.51
Expt.	2.0±0.5[89]	+1.0±0.7[89]		0.7±0.5[90]		

Table 1.1 Interlayer relaxations for Cu(100) and Cu(111) surfaces computed using Density Functional Theory with the PBE exchange-correlation functional.

The Cu(100) surface, with its square planar arrangement, tends to exhibit more significant changes in interlayer spacing due to its more open structure, as illustrated in Figure 1.2. In contrast, the Cu(111) surface, which features a hexagonal close-packed arrangement, exhibits smaller relaxations due to its more tightly packed atomic arrangement. This difference in atomic arrangement and surface density

directly impacts their surface energy and stability, making Cu(100) surfaces more likely to relaxation phenomena as they seek to minimize surface energy, resulting in greater changes in interlayer spacing.

1.1.4 Defects in metals

The study of point defects in metals is critical for understanding their mechanical and physical properties, as Lomer (1959) emphasized in his work on defects in pure metals [91]. Point defects, which result from deviations in the perfect lattice arrangement, significantly impact key material properties such as diffusion rates, mechanical resilience, and electrical conductivity. Vacancies and interstitials are significant among these defects. The former is missing atoms in the lattice structure, while the latter occurs when extra atoms occupy spaces between regular lattice positions. The presence of these defects can fundamentally change how a material behaves, influencing its ability to conduct electricity, withstand mechanical stress, and enable the movement of atoms through diffusion. Li and Lu (2017) highlight that managing and manipulating defects in metals, such as vacancies and dislocations, can significantly enhance their mechanical properties[92]. By strategically controlling these defect structures, it is possible to modify metals to have specific desirable properties. The generation of defects can be intentional, through processes like alloying, heat treatment, or mechanical deformation, or unintentional, due to environmental factors like radiation or temperature changes. Understanding how to control these factors and their impact on defect formation is crucial for optimizing the performance of metals in various applications.[92] Factors such as temperature, pressure, and environmental interactions are critical in driving structural transformations in metals. Elevated temperatures increase atomic mobility, leading to diffusion processes that promote defect formation or the migration of existing defects, resulting in surface reconstruction or phase changes that alter the crystal lattice. Environmental interactions, including chemical reactions, radiation, and corrosion, can introduce impurities or alter the atomic structure of metals, leading to surface modifications, defect generation, and changes in lattice structure[93]. Incorporating these factors into the study of metals helps understand their behavior under various *operando* conditions, influencing the development of materials with tailored properties; these transformations are known to significantly improve properties like electrical conductivity and mechanical strength by modifying dislocation dynamics.[94–96]. As noted by Gourdin *et. al*

(2021), employing in situ and *operando* characterization techniques is crucial for detailing these changes, thereby advancing our understanding and application of electrochemical devices.[97]

Furthermore, the dynamic equilibrium maintained by operational environments suggests that metal structures continually adapt, exhibiting transient properties that are difficult to capture with traditional analysis methods. For instance, Li *et al*. (2015) demonstrated that correlated imaging and spectroscopy probes under *operando* conditions could uncover the complex structural dynamics of nanocatalysts, providing invaluable insights into the behavior of metal species across various states.[72]

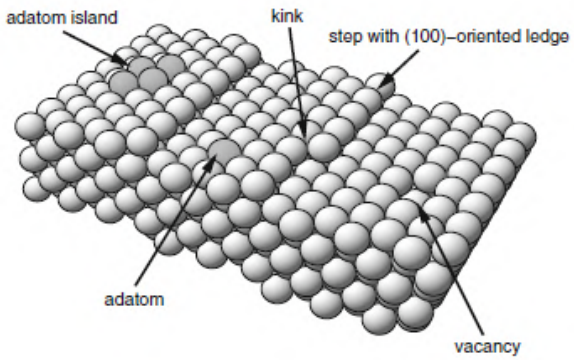


Fig. 1.3 **Illustration of various surface defects on a metal crystal lattice.** Isolated atoms on the metal surfac can migrate and cluster together to form structures known as surface islands. Kinks are abrupt changes in the direction at the edge of atomic planes, while steps with (100)-oriented ledges are linear defects where additional atomic layers begin. The presence of vacancies—sites where an atom is missing—also affects the surface properties. These features play a significant role in the properties and behavior of metal surfaces *Adapted with permission from Springer Nature*[88].

Incorporating these observations strengthens the understanding of how defects are integral to the dynamic and responsive nature of metal structures to environmental conditions, leading to profound changes in material behavior and performance. This understanding aligns with findings from other significant works [98–100].

The study of these defects has advanced through both theoretical and experimental lenses. Theoretically, the exploration of the energy landscapes that govern defect formation, migration, and interaction provides critical insights into their behavior. This approach allows for the prediction of equilibrium concentrations of defects and their impact on material properties at various temperatures. Experimentally, sophisti-

cated techniques such as electron microscopy, positron annihilation spectroscopy, and electrical resistivity measurements have enabled the direct observation and quantitative analysis of these defects, shedding light on their roles in the underlying mechanisms of material behavior[101–105].

Figure 1.3 serves as a visual compendium of surface defects that can exist on metal crystals. Vacancies, kinks, and steps are illustrated, each playing a distinct role in the surface dynamics and overall behavior of the metal.

The understanding of point defects is further complicated when considering the effects of operational conditions on defect formation and evolution. High temperatures, for instance, can increase the mobility of defects, accelerating diffusion processes [106], while applied stresses can drive the formation of new defects or the migration of existing ones, impacting mechanical properties such as creep and fatigue [107].

Advancements in computational modeling and *in situ* characterization techniques are enhancing our ability to simulate and observe these defects in real-time, providing unprecedented insights into their formation, migration, and interactions at the atomic scale [108]. This progress paves the way for developing new metal alloys and materials engineered to optimize performance and longevity by utilizing the intrinsic properties of point defects [109].

1.1.5 Surface reconstruction

Before exploring surface reconstruction, it is important to distinguish this phenomenon from surface relaxation, as both alter surface properties through different mechanisms. Surface relaxation involves minor adjustments in the positions of atoms near the surface, responding to unbalanced forces to lower surface energy while maintaining the basic lattice structure. In contrast, surface reconstruction entails significant atomic rearrangement, creating new surface structures and periodicity to minimize energy under varying conditions. This often leads to changes in surface geometry and properties, affecting the material's overall performance[110–112]. Indeed, the electronic structure of closed-packed metal surfaces is different from that in the bulk, with an abrupt decrease in the density of conduction electrons. This can result in a change in bond lengths between surface atoms and occasionally a reconstruction involving a change in the number of atoms in the first surface layer. These transformative adaptations significantly deviate from the internal crystalline arrange-

ment, influencing the surface's chemical, physical, and electronic properties, and thereby affecting a material's overall performance in various applications[111, 113].

While some surfaces may retain the structural integrity of the bulk crystal, dubbed as 'ideal unreconstructed surfaces,' these instances represent exceptions. The majority of surfaces encountered in practical situations exhibit some degree of reconstruction, where the arrangement of atoms in the topmost layers diverges from the bulk structure. Such reconstructed surfaces manifest a two-dimensional periodicity that may diverge from that of the bulk lattice planes, introducing novel properties and behaviors not present in the idealized models. The importance of surface reconstruction is underscored by its significant impact on material performance, particularly in environments subjected to high temperatures, corrosive media, or mechanical stresses. These reconstructed surfaces may exhibit improved resistance, enhanced catalytic activity, or superior electronic properties, diverging markedly from expectations based on the bulk properties alone[110, 111, 114]. The modification in the electronic structure brought about by surface reconstruction alters surface energy states, affecting electron mobility and thereby the electrical conductivity and catalytic efficiency of the material. Furthermore, the operational durability of metallic systems, especially in catalytic and sensor applications, is closely linked to their surface configuration. Reconstructed surfaces can present an energetically more favorable arrangement under certain environmental conditions, thus enhancing the material's stability and extending its useful applications under those specific conditions[110, 115].

1.1.6 Advancements in *in situ* experimental characterization techniques

Recent developments in *in situ* characterization techniques have significantly enhanced the understanding of the complex interplay between the structure and reactivity of metallic systems at the atomic and molecular scales [116–119]. *In situ* and *operando* spectroscopy, along with state-of-the-art electron microscopy, stand at the forefront of these advancements, offering unprecedented insights into the dynamic nature of metallic systems and the active sites critical for their properties. Technologies such as synchrotron-based spectroscopy bring to light the detailed electronic structures and bonding states within metals, thanks to their high-intensity X-rays. X-ray Absorption Spectroscopy (XAS) explores the local atomic structure

and electronic states, making it invaluable for assessing changes in oxidation states and coordination environments [120–123]. Ambient Pressure X-ray Photoelectron Spectroscopy (AP-XPS)[124] pushes the boundaries of surface chemistry analysis, allowing for investigations under more realistic pressure and temperature conditions. Meanwhile, Environmental Transmission Electron Microscopy (ETEM)[125] and Scanning Transmission Electron Microscopy (STEM)[126, 127] provide a lens into the nanoscale world, capturing real-time structural changes, defect formations, and the dynamics of catalytic processes. Collectively, these methods have revolutionized our ability to monitor the electronic and geometric alterations in metallic systems under actual reaction conditions.

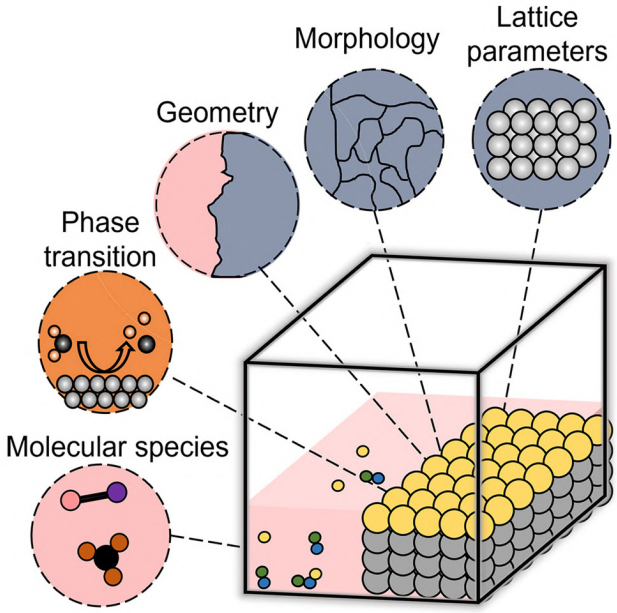


Fig. 1.4 **Illustration of *in situ* characterization techniques used to analyze a metallic slab.** These techniques, including synchrotron-based spectroscopy, X-ray Absorption Spectroscopy (XAS), Ambient Pressure X-ray Photoelectron Spectroscopy (AP-XPS), and advanced electron microscopy methods like Environmental Transmission Electron Microscopy (ETEM) and Scanning Transmission Electron Microscopy (STEM), provide detailed insights into the electronic structures, bonding states, and dynamic nature of metallic systems under realistic conditions. This comprehensive approach enables the observation of electronic and geometric changes in real-time, facilitating a deeper understanding of the interplay between structure and reactivity in catalytic processes. *Reproduced with permission from Elsevier*[119].

Advancements in technology have demonstrated how temperature influences structure and active sites, yet these experimental methods come with inherent lim-

itations. Their inability to track individual atoms or capture the rapid processes fundamental to observation highlights the critical role of computational methods. By simulating and tracking the behavior of individual atoms across various timescales, which are difficult to observe directly, these methods fill the gaps left by experimental techniques.

1.2 The role of simulations in the study of metals

Computational methods have significantly advanced the study of metallic systems by providing profound insights into their complex behaviors at the atomic and electronic levels. These techniques facilitate a comprehensive understanding of the intricate interactions within these systems under diverse conditions, often challenging to replicate precisely in traditional experimental setups. Unlike experimental methods, which can struggle with replicating exact operational environments or directly observing atomic-level phenomena, computational models offer a detailed perspective of the microscopic interactions governing metal properties. This paradigm shift has enhanced the understanding of metallic systems, illustrating how minor atomic-scale adjustments can profoundly influence macroscopic properties such as electrical conductivity, mechanical strength, and chemical reactivity[128–130].

Computational models allow us to simulate a wide range of conditions, adjusting parameters to explore the effects of temperature, pressure, and chemical environments on metal performance. This flexibility enables a thorough investigation of metal behavior across various scenarios, significantly advancing our understanding beyond what is typically achievable in experimental settings. Moreover, the efficiency and cost-effectiveness of computational simulations stand in contrast to the typically time-consuming and expensive nature of experimental research.

Historically, Density Functional Theory (DF) has played a fundamental role in studying various properties of metals. Introduced by Hohenberg and Kohn [131] and further developed by Kohn and Sham [132], DFT has become an indispensable tool in computational materials science. It allows precise calculations of electronic structures, cohesive energies, and other properties critical to understanding metallic systems, as shown in the dedicated section in the next chapter. Significant works using DFT have provided deep insights into the behavior of metals; for instance, Jones and Gunnarsson [133] discussed the successes and challenges of DFT in de-

scribing electronic properties and bonding in metals. Hafner [134] used DFT to explore the phase stability and structural properties of transition metals, contributing substantially to the understanding of their phase diagrams and transformation mechanisms. More recently, studies by Curtarolo *et al*. [128] and Rogal *et al*. [135] have employed DFT to investigate the thermodynamic properties and phase behavior of complex metallic alloys, illustrating the method's versatility and depth.

However, DFT primarily focuses on the static properties of materials, providing a snapshot of the electronic structure under specific conditions. While it excels in offering detailed information on ground-state properties, DFT's application to dynamic processes is limited due to its inherently static nature. Despite these limitations, DFT has been invaluable to metal studies. It has laid the groundwork for understanding the fundamental properties of metals and continues to be fundamental in the computational study of metallic systems.

To address the dynamic behavior of metallic systems, Car and Parrinello [136] introduced an innovative approach that combines DFT with molecular dynamics, known as Car-Parrinello Molecular Dynamics (CPMD). This method allows for the simultaneous optimization of electronic and ionic configurations, enabling the study of both electronic properties and atomic movements in a unified framework. CPMD has significantly enhanced the ability to investigate the time-dependent behavior of metals under various conditions. Furthermore, CPMD allows the application of DFT to much larger systems than was previously feasible, significantly enhancing the ability to study the time-dependent behavior of metals and other materials under various conditions.

Several studies have successfully used CPMD to explore metal systems. For example, Car and Parrinello's initial work demonstrated the application of their method to simple metals like lithium, revealing insights into atomic diffusion and phase transitions[136]. Subsequent studies have expanded the use of CPMD to more complex systems. For instance, Billeter *et al*.[137] used CPMD to study the melting behavior of aluminum, providing detailed information on the atomic-level mechanisms driving the melting process. Classical Molecular Dynamics has also played a crucial role in the study of metallic systems. Significant early works in classical MD focused on understanding fundamental properties such as diffusion, phase transitions, and mechanical properties of metals. For example, Parrinello and Rahman [138] developed a method to simulate phase transitions in crystalline solids, which has been extensively applied to study structural changes in metals under

different conditions. Another notable example is the work by Foiles *et al*. [139], who used MD simulations to investigate the diffusion of atoms in metals, providing insights into defect dynamics and materials' behavior at high temperatures.

In parallel, the accurate description of force fields has always been a central focus due to the complexity of interactions within metallic systems. Force fields are mathematical models used to predict the energy and forces in a system of atoms or molecules. The development and refinement of force fields have been crucial for studying metals, where the interactions between atoms are highly complex and require precise modeling to accurately simulate material behavior.

One of the most influential force fields for metals is the Embedded Atom Method (EAM), introduced by Daw and Baskes [140]. EAM has been widely adopted for its ability to accurately simulate metallic bonding and defect structures. It accounts for the many-body interactions that are characteristic of metallic systems, making it suitable for modeling properties such as cohesion, surface energies, and vacancy formation. Another important force field is the Modified Embedded Atom Method (MEAM), which extends EAM to handle more complex metallic alloys and compounds, as shown in the work of Baskes *et al.*[141].

Other significant force fields from the past include the Finnis-Sinclair potentials [142] and the Second Moment Approximation of tight-binding models [143], both extensively used to study a variety of metallic properties and behaviors. These force fields enable simulations that capture the essential physics of metallic bonding and have been fundamental tools for understanding the mechanical and thermal properties of metals. This historical overview highlights the crucial role of computational methods in advancing the understanding of metallic systems. Recent advancements in computational power and analytical tools have further enhanced this field. Improved computational resources now allow for more detailed simulations over longer timescales, while better analytical techniques offer improved validation and insights into models. These developments provide a deeper understanding of metal behavior under realistic conditions, driving innovation and increasing the predictive power of computational studies. This sets the stage for our next section, where we explore recent progress in computational modeling and molecular simulations.

1.2.1 Unveiling metal dynamics: computational modeling and molecular simulations

The integration of advanced computational power and sophisticated analytical tools has significantly advanced the study of metallic systems. This progression has enabled the examination of increasingly complex systems under a variety of conditions, revealing diverse properties and behaviors of metals.

Recent advancements in computational power have dramatically expanded the scale and scope of simulations that can be performed. For instance, Schlexer *et al.* (2018) highlighted the cost-effectiveness of computational methods in screening catalytic reactions to understand structural changes and predict behavior [144]. Similarly, Zhai *et al.*(2014) explored how additive manufacturing technologies, enhanced by computational modeling, have revolutionized metal design. These technologies reduce material and energy costs, facilitating more innovative and complex designs at reduced expenses. This capability enables broad exploration of metallic systems with minimal resource expenditure, accelerating research and development. It also opens the door to exploring innovative metal designs and applications that might have been considered too risky or costly to pursue through experimental methods alone.[145]

The predictive power of computational models is a crucial advantage, offering the ability to forecast the behavior of metallic systems accurately. These models are instrumental in designing and optimizing new metals and alloys, serving as essential tools in materials engineering. For instance, Gatsos *et al.* (2020) discussed how computational modeling predicts and enhances the mechanical properties and structural characteristics of metal components in additive manufacturing.[146]

In the field of tribology, computational methods provide crucial insights into wear mechanisms and lubrication processes under operational conditions. For instance, Righi and Loehlé (2018) utilized ab initio molecular dynamics simulations to investigate tribochemical reactions at sliding iron interfaces, revealing mechanisms that reduce friction and wear.[147] Further insights into the tribological behavior of lubricant additives were explored, detailing the interactions that govern the tribological properties of these interfaces. This study illustrates the broad applicability of computational models for predicting chemical interactions and optimizing tribological behavior, enabling the development of advanced materials for diverse industrial uses.[38]

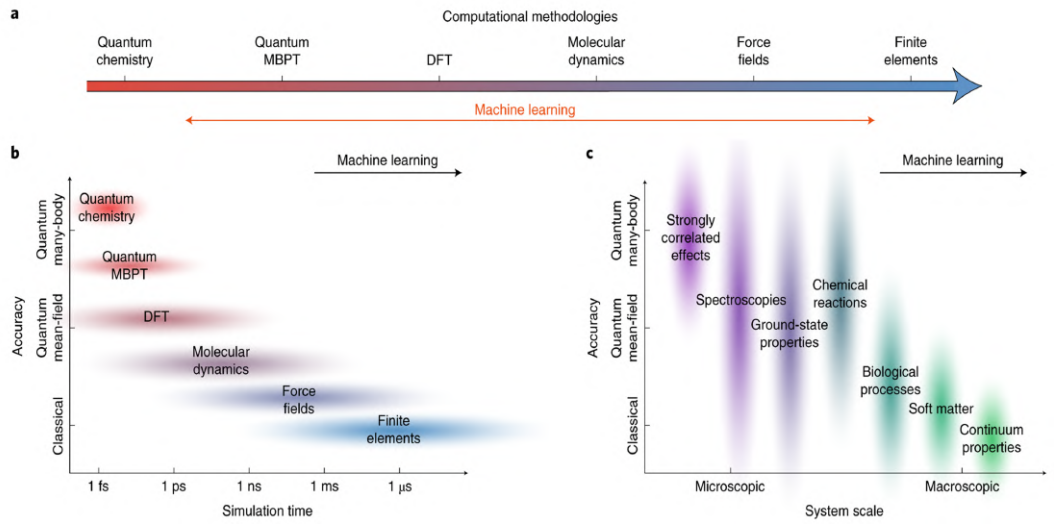


Fig. 1.5 **Overview of Computational Methodologies Across Different Length and Time Scales (a)** Schematic representation of computational methodologies spanning from quantum chemistry to finite elements. The figure illustrates the progression from quantum mechanical methods (such as quantum chemistry and many-body perturbation theory) to classical approaches (like force fields and finite element methods), with increasing system size and complexity. Machine learning is shown as a unifying tool that can enhance accuracy and efficiency across these methods. **(b)** Trade-off between accuracy and simulation time for different computational methods. Quantum chemistry offers high accuracy but is limited to short simulation times, while methods like molecular dynamics and finite elements allow for longer simulations at the expense of accuracy. Machine learning aims to bridge this gap by improving the efficiency and accuracy of simulations across different scales. **(c)** The application of various computational methods to different scientific problems. Quantum many-body methods and mean-field theories are essential for studying strongly correlated effects and spectroscopies at microscopic scales. As the system scale increases, methods such as molecular dynamics and force fields are used for chemical reactions, ground-state properties, and biological processes. Finite elements are applied to macroscopic phenomena, like the continuum properties of soft matter and biological tissues. *Reproduced with permission from Springer Nature*[148].

Moreover, computational approaches excel in elucidating aspects of metal behavior that are beyond the reach of direct experimental observation. These models provide a comprehensive understanding of the principles governing metal behavior by investigating transient and elusive phenomena, such as intermediate species, electronic density distributions, and the initiation of deformation and wear processes. For instance, Vogiatzis *et al.* (2018) explored the role of 3d transition metals in catalysis, shedding light on the intricate structure and behavior of complex molecular species, often elusive in experimental setups. This work underscores the critical

role of computational methods in revealing the dynamic processes of intermediate species formation and their influence on catalytic activity[149]. Similarly, Harvey *et al*. (2019) emphasized the scope and challenges of computational methods in studying the mechanism and reactivity in homogeneous catalysis, highlighting the formation and role of intermediates in these systems. These studies demonstrate the capability of computational tools to predict and analyze the dynamic and complex nature of catalytic processes, offering insights into the behavior of metals that are often impossible to observe in traditional experiments.[150]

Given the insights shared thus far and the technological advancements in computational modeling, it is becoming increasingly evident that incorporating the effects of dynamic changes is essential for studying metals. Indeed, as discussed in subsection 1.1.6, recent advancements in technology have enabled *in situ* and *in operando* characterization of catalytic materials, highlighting the substantial impact these conditions have on the structure and corresponding activity [151, 152]. These techniques allow for real-time observation of changes, providing crucial insights into the dynamic behavior of metals under realistic operating conditions. This progress has underscored the importance of considering the dynamic aspects of metallic systems to fully understand their behavior and properties.

Theoretical calculations have also emphasized the necessity of considering dynamics [153–162]. For instance, Gazzarrini *et al*. demonstrated that atom mobility in copper nanoparticles can cause variations in the number of vertex, edge, and face atoms, subsequently affecting the efficiency of these nanoparticles in catalyzing CO_2 conversion to methane [161].

Recent studies have further highlighted the dynamic behavior of catalytic interfaces. Lavroff *et al*. (2022) demonstrated how computational models could simulate realistic reaction conditions to predict surface restructuring and particle behavior [157]. This underscores the importance of incorporating dynamic models that account for structural changes in metals to enhance our understanding of catalytic behavior in real-world applications. Zhang *et al*. (2020) discussed how metastable states influence catalysis on dynamic interfaces [158]. Similarly, Li *et al*. (2021) emphasized the critical role of thermal fluctuations in CO binding on electrocatalytic metal surfaces [159].

Further studies, such as those by Grajciar *et al*. [160] and Wang *et al*. [163], underline the importance of including dynamics in *operando* computational mod-

eling to achieve a more realistic understanding of catalytic processes. This is also demonstrated in the work of Van Speybroeck *et al.*, who used first-principles molecular dynamics simulations to model zeolite-catalyzed reactions under *operando* conditions [162].

Despite these insightful examples, most computational studies still tend to treat metallic systems as static entities, focusing primarily on their chemical reactivity or physical properties without fully considering their dynamic nature[164–170]. Such approaches often overlook the fact that metal atoms and clusters are inherently dynamic, exhibiting significant mobility, reactivity, and the ability to undergo deformation, relaxation, and surface restructuring under realistic experimental conditions[161, 171–173].

However, computational calculations this field are frequently constrained by short timescales or by dynamically treating only a subset of the system's degrees of freedom. Traditionally, computational studies have focused on the chemistry of reactions within these systems while neglecting the dynamics of active sites, thus underestimating the importance of these dynamic changes on the overall behavior and performance of metals. A shift towards modeling that integrates the dynamic aspects of metallic systems can provide a more accurate representation of these materials under real-world conditions.

To accurately study the structural dynamics of these systems, it is crucial to treat atomic interactions with precision while also simulating these systems on large enough spatial and temporal scales. This approach is necessary to avoid finite-size effects and ensure that microscopic dynamic transitions, and not only vibrations, are sampled with sufficient statistical relevance. Yet, accurate DFT calculations are limited by their short timescales and small system sizes, while classical atomistic force fields may lack the accuracy needed to effectively capture the structural dynamics of these systems, as shown in Figure 1.5.

Addressing these limitations requires the development of more sophisticated models and methods capable of capturing the full range of interactions and behaviors in metallic systems over longer timescales. These advanced models should integrate the atomic degrees of freedom (the positions and movements of atoms) to provide a more comprehensive understanding of the structure-related properties. The continued integration of advanced techniques, such as machine learning and enhanced sampling, promises to overcome these challenges. In particular Machine learning

techniques offer the potential to predict complex interactions and behaviors with greater accuracy and can effectively bridge the gap between different length and time scales (Figure 1.5a). The integration of these advanced computational methods, particularly Machine Learning potentials, has fundamentally transformed the field of materials science[174–179]. ML potentials are capable of achieving the accuracy of Density Functional Theory calculations while simultaneously managing the larger system sizes that are typical of classical simulations[180–183, 176, 184, 179, 185]. These advancements facilitate a dynamic approach to studying metallic systems, taking into account their mobility, reactivity, and structural changes. By adopting these advanced techniques, it is possible to overcome the limitations of traditional methods, enabling more accurate and comprehensive simulations of metallic systems. Furthermore, by handling system sizes comparable to those in classical molecular dynamics simulations, these methods allow for proper sampling of the system, enabling the study of relevant properties with higher precision. The concept of neural network potentials was first introduced by Behler and Parinnello (2007)[186], who proposed a method to represent high-dimensional potential energy surfaces with machine learning models. Their approach marked a significant breakthrough in modeling complex atomic interactions with high accuracy, setting the foundation for subsequent research in the field. Nearly a decade later, Behler (2016) expanded on this foundational work, providing a comprehensive perspective on how ML models could achieve predictive reliability akin to quantum mechanical accuracy [187]. This further bridged the gap between empirical potentials and quantum mechanical methods, solidifying the potential of ML models in atomistic simulations. Smith *et al*. (2017) built upon these foundations by demonstrating the extensibility of neural network potentials; their work showed that these models could accurately capture complex chemical behaviors while reducing the computational cost [188].

Mueller *et al*. (2020) explored the rapid advancements in machine learning potentials, demonstrating the evolving ability to model material properties with unparalleled accuracy. The study emphasized that these advancements are reshaping the field of materials science, enabling the investigation and prediction of complex material behaviors that were previously challenging due to computational limitations [189]. These advancements underscore the transformative potential of ML potentials in revolutionizing atomic simulations, offering unparalleled opportunities for exploring material properties and behaviors and undertaking studies that were once limited by computational constraints [176, 190–196, 184, 197–199].

The advent of machine learning potentials has opened new doors to studying concepts introduced decades ago, like the dynamical steady state of catalyst surfaces, proposed by Spencer over forty years ago [200]. Dynamic changes observed, such as surface atom mobility and instability, challenge traditional static models of catalysis, prompting a shift toward dynamic modeling. These dynamic behaviors extend beyond catalysis, influencing a range of structural properties of metals, including mechanical, electronic, and magnetic characteristics. Confirming this, recent studies have underscored the crucial importance of considering the dynamic behavior of metal surfaces in catalysis through the use of machine learning potentials.

In his work, Bonati *et al.* [185], investigate the dynamic behavior of the Fe(111) surface and its impact on nitrogen decomposition. Utilizing ML potentials and advanced simulation techniques, the study reveals that under *operando* conditions, the surface undergoes significant dynamical changes. The step structure of the surface becomes destabilized, leading to the continuous formation and disappearance of catalytic sites. These findings highlight the dangers of extrapolating low-temperature results to *operando* conditions and underscore the necessity of incorporating dynamics into catalytic activity calculations. Then Perego *et al.* [201] demonstrate that the dynamic behavior of the iron surface at high temperatures significantly alters reaction pathways and the stability of catalytic sites. Both studies emphasize that the catalytic process is driven by the transition to highly fluctuating interfacial environments, highlighting that catalytic activity can only be accurately inferred from dynamic models. Incorporating the dynamic characteristics of metallic systems into computational methodologies is therefore essential. To this end, Machine learning offers advanced predictive tools that can unravel the complex interactions influencing metal behavior at the atomic level.

1.3 Aim of the thesis

The investigation of structural and microscopic dynamics within various materials, with a particular focus on metallic systems, forms the foundation of my research. This thesis aims to uncover the complex behaviors and intrinsic properties that characterize these materials under relevant conditions. This study integrates computational methods like Density Functional Theory, Molecular Dynamics, and Deep Potential Molecular Dynamics (DPMD) with advanced machine learning algorithms

to achieve this goal. Machine learning is employed in this research not only to train potentials that enable the study of microscopic dynamics of metal surfaces but also to optimize methods for encoding, classifying, and comparing different metallic systems. Specifically, descriptors of atomic environments, such as the Smooth Overlap of Atomic Positions (SOAP)[202] and the Local Environments and Neighbors Shuffling (LENS)[203], are used to transform raw simulation data into detailed fingerprints of atomic arrangements and dynamics. By decoding the intrinsic microscopic dynamics, we aim to provide a comprehensive depiction of the behaviors of metallic systems. This research broadens our understanding of metal behavior and lays the groundwork for designing systems with tailored characteristics for specific applications. Additionally, this thesis seeks to transcend traditional boundaries by adapting and developing analytical methodologies applicable to different materials. By taking inspiration from approaches from other fields, such as soft materials and self-assembling systems[204, 205], and applying them to metallic systems, we aim to deepen our understanding of metal properties. These approaches have demonstrated that by starting from and resolving the microscopic dynamics of a system, one can reconstruct the macroscopic properties of the system. Viewing metals as complex systems of interacting atoms and tracking their relative atomic motions within the lattice, we can reconstruct their dynamic complexity. This approach aspires to bring about a paradigm shift, moving the focus from static structural analysis and internal thermodynamics to the study of atomic dynamics. This new perspective reveals how microscopic dynamics and atomic interactions influence the macroscopic properties of metals, providing valuable insights that could transform the methodologies used in metal studies. By adopting a perspective focused on microscopic dynamics, this thesis not only provides a detailed depiction of the complex nature of metallic systems but also lays the groundwork for future research to explore these dynamic properties further. By linking these concepts, we present a comprehensive and cohesive approach to studying the dynamic behaviors of metallic systems, ultimately contributing to a more profound understanding of their properties and potential applications.

Chapter 2

Methods and theoretical backgrounds

This chapter explores the theoretical foundations of computational methods in materials science, with a particular focus on their application to metallic systems. It will concentrate on the specific techniques utilized in the manuscripts presented in this thesis, illustrating how these methods have been applied throughout my research activity.

These computational models offer invaluable insights into the atomic-scale processes governing the structural evolution, phase transitions, and mechanical properties of metals. Simulations allow us to unravel the intricate atomic and electronic structures, shedding light on phenomena that are impossible to observe directly through experimental means. Through a short review of these methods, we aim to provide a solid foundation for understanding their principles and applications in the context of materials science. The discussion begins with the concept of the Hamiltonian, essential in building the theoretical framework in computational simulations. We then move to Density Functional Theory, a quantum mechanical modeling method able to compute the electronic structure of metals, alloys, and other materials. DFT shifts the focus to electron density rather than wave functions, making it a more feasible approach for solving the Schrödinger equation in complex systems.

We then transition from the quantum mechanical framework of DFT to Classical Molecular Dynamics, marking a significant shift in perspective and scale. This transition moves from considering electrons and their interactions to a classical approach modeling atoms and molecules under classical mechanics laws. Our discussion follows with the introduction of Deep Potential Molecular Dynamics (DPMD), which integrates deep learning with molecular dynamics simulations for

higher accuracy and efficiency in modeling complex material systems. Thanks to ML potentials trained on quantum mechanical data, DPMD enhances traditional MD simulations, enabling the study of larger systems and longer timescales with quantum-level precision. We then discuss enhanced sampling techniques, which offer efficient strategies to explore the energy landscape and better understand rare events and phase transitions within metallic systems. Finally, the chapter introduces the role of descriptors in molecular dynamics simulations; descriptors are essential tools that aid in identifying and quantifying structural motifs and dynamic processes, bridging the gap between raw computational data and interpretable scientific insights.

2.1 The Hamiltonian

In theoretical physics and chemistry, understanding a system begins with defining the system itself and identifying the fundamental interactions within it. Central to this analysis is the Hamiltonian, which encapsulates the total energy of the system, allowing us to derive all physical and chemical properties.

When dealing with microscopic entities like electrons and atoms, quantum mechanics is the ideal framework. This field allows us to describe and predict the behavior and interactions of particles at atomic and subatomic levels. The key to these predictions lies in solving the Schrödinger equation, which provides insights into the quantum states and properties of systems composed of nuclei and electrons.

This microscopic understanding is vital as it forms the foundation for explaining and predicting the macroscopic properties of materials.

In our context, primary attention is given to electrostatic interactions. Relativistic effects are generally negligible, especially when considering only valence electrons, while magnetic effects, arising from the movement of electrons generating magnetic fields, are typically disregarded due to their minimal impact on the overall system properties.

Consequently, a system comprising nuclei and electrons is described by the nonrelativistic Schrödinger equation:

$$H\Psi = E\Psi, \tag{2.1}$$

where H is the Hamiltonian operator, Ψ is the wave function of the system, and E is the energy eigenvalue. The Hamiltonian operator H includes the kinetic energy of the electrons and nuclei, as well as the potential energy from electrostatic interactions. The wave function Ψ contains all the quantum mechanical information about the system, and solving this equation provides the possible energy levels E that the system can occupy.

The Hamiltonian H for the system is expressed as:

$$H = T_{\text{nuc}} + T_{\text{el}} + V_{\text{nuc-nuc}} + V_{\text{nuc-el}} + V_{\text{el-el}}, \tag{2.2}$$

In this expression, T_{nuc} and T_{el} are operators corresponding to the kinetic energy of the nuclei and electrons, respectively. The terms $V_{\text{nuc-nuc}}$, $V_{\text{nuc-el}}$, and $V_{\text{el-el}}$ are potential energy operators representing the nuclear-nuclear, nuclear-electron, and electron-electron interactions. These operators act on the wave function of the system, and their expectation values yield the corresponding energy contributions. Specifically, the operators encode the quantum mechanical rules for how these energies are derived from the wave function, translating the abstract wave function into physically observable quantities. For simplicity, spin effects are initially neglected in this formulation.

The explicit components of the Hamiltonian are outlined as follows:

$$T_{\text{nuc}} = \sum_{I=1}^{L} \frac{P_I^2}{2M_I}, \tag{2.3}$$

$$T_{\text{el}} = \sum_{i=1}^{N} \frac{p_i^2}{2m}, \tag{2.4}$$

$$V_{\text{nuc-nuc}} = \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J e^2}{|R_I - R_J|}, \tag{2.5}$$

$$V_{\text{nuc-el}} = -\sum_{i,I} \frac{Z_I e^2}{|r_i - R_I|}, \tag{2.6}$$

$$V_{\text{el-el}} = \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|r_i - r_j|}. \tag{2.7}$$

In these equations, P_I and M_I represent the momentum and mass of the I-th nucleus, p_i and m the momentum and mass of the i-th electron, and Z_I the charge

of the I-th nucleus. The positions of the nuclei and electrons are denoted by R_I and r_i, respectively, and e is the elementary charge. The factor of $\frac{1}{2}$ in $V_{\text{nuc-nuc}}$ and $V_{\text{el-el}}$ prevents the double counting of interactions between the same pairs of particles.

2.1.1 The Born–Oppenheimer approximation

The Born–Oppenheimer, or adiabatic, approximation is fundamental in theoretical descriptions due to the significant mass difference between electrons and nuclei. This difference allows nuclei to be treated as moving much slower than electrons, leading to the assumption that electrons can instantaneously adapt to the positions of the nuclei.

Within this framework, the system's Hamiltonian, as presented in eq 2.2, can be partitioned into electronic and nuclear components. The electronic Hamiltonian $H_{\text{el}}(\{R\})$ for a fixed configuration of nuclei is given by:

$$H_{\text{el}}(\{R\}) = T_{\text{el}} + V_{\text{nuc-el}} + V_{\text{el-el}} + V_{\text{nuc-nuc}}, \tag{2.8}$$

where T_{el} is the operator for the kinetic energy of the electrons, $V_{\text{nuc-el}}$ is the operator for the attractive interaction between nuclei and electrons, $V_{\text{el-el}}$ is the operator for the repulsive interaction between electrons, and $V_{\text{nuc-nuc}}$ represents the fixed repulsive interaction between nuclei.

Given this Hamiltonian, the electronic Schrödinger equation for a fixed nuclear configuration is:

$$H_{\text{el}}(\{R\})\Psi(\mathbf{r},\{R\}) = E_{\text{el}}(\{R\})\Psi(\mathbf{r},\{R\}), \tag{2.9}$$

where $E_{\text{el}}(\{R\})$ represents the electronic energy for that specific configuration of nuclei, which defines the potential energy surface (PES) upon which the nuclei move.

The total energy of the system within the Born–Oppenheimer approximation is then obtained by adding the kinetic energy of the nuclei T_{nuc} to the electronic energy $E_{\text{el}}(\{R\})$ and the fixed repulsive interaction between the nuclei $V_{\text{nuc-nuc}}$. Thus, the total energy is given by:

$$E_{\text{total}} = E_{\text{el}}(\{R\}) + T_{\text{nuc}} + V_{\text{nuc-nuc}}. \tag{2.10}$$

This approximation enables:

1. **Decoupling of Nuclear and Electronic Motion**: The assumption that nuclei move much more slowly than electrons allows the electronic Schrödinger equation to be solved for fixed positions of the nuclei, with the resulting electronic energy providing the PES for nuclear motion.

2. **Simplification of Computational Calculations**: By reducing the dimensionality of the problem, the approximation allows the electronic and nuclear wavefunctions to be treated separately, thus decreasing computational complexity and making it feasible to study larger and more complex systems.

3. **Understanding Molecular Dynamics**: The PES derived from electronic structure calculations is fundamental in studying nuclear dynamics, either using classical or quantum mechanical methods, and is crucial for simulating molecular dynamics and chemical reactions.

To compute the forces acting on the nuclei, the Hellmann–Feynman theorem provides a direct method based on the ground state wavefunction:

$$F_I = -\nabla_{R_I} E_{\mathrm{el}}(\{R\}) = \langle \Psi(\mathbf{r},\{R\}) \,|\nabla_{R_I} H_{\mathrm{el}}(\{R\})|\, \Psi(\mathbf{r},\{R\})\rangle \,. \tag{2.11}$$

This theorem, coupled with the Born–Oppenheimer approximation, not only simplifies the treatment of electronic and nuclear motions but also provides a practical way for calculating forces, which is fundamental in molecular dynamics simulations.

2.1.2 Bridging the gap to Density Functional Theory

The study of the Hamiltonian and the simplifications introduced by the Born–Oppenheimer approximation are essential for understanding atomic and molecular behavior. However, these concepts also underscore the challenges associated with solving the many-body Schrödinger equation, particularly due to the complex electron-electron interactions. The traditional wavefunction-based approach in quantum mechanics requires a detailed description of the system through a many-body wavefunction,

which depends on the coordinates and spins of all electrons and nuclei. This wavefunction is a high-dimensional object that, while complete in its description, is computationally intractable for systems with many particles.

In contrast, DFT offers a profound simplification by shifting the focus from the many-body wavefunction to the electron density—a physically observable quantity that depends only on three spatial coordinates, regardless of the number of electrons in the system. This transition from wavefunction to electron density reduces the computational complexity while providing a more direct connection to experimentally measurable quantities.

The foundation of DFT lies on the Hohenberg–Kohn theorems, which establish that all ground-state properties of a system are uniquely determined by its electron density $\rho(\mathbf{r})$. According to the first Hohenberg–Kohn theorem, the ground-state electron density $\rho(\mathbf{r})$ uniquely determines the external potential $V_{\text{ext}}(\mathbf{r})$ acting on the electrons, and consequently, the Hamiltonian and all other properties of the system. This insight, which earned Walter Kohn the Nobel Prize in Chemistry in 1998, underscores the central role of electron density in quantum systems.

According to the second Hohenberg–Kohn theorem, the ground-state energy of a system can be obtained by minimizing the total energy functional $E[\rho]$ with respect to the electron density $\rho(\mathbf{r})$. This energy functional is expressed as:

$$E[\rho] = T[\rho] + V_{\text{ext}}[\rho] + V_H[\rho] + E_{\text{xc}}[\rho], \tag{2.12}$$

where:

- $T[\rho]$ is the kinetic energy functional of the electrons,
- $V_{\text{ext}}[\rho]$ is the external potential energy functional,
- $V_H[\rho]$ is the Hartree energy functional, describing the classical electrostatic interaction between electrons, and
- $E_{\text{xc}}[\rho]$ is the exchange-correlation energy functional, accounting for complex quantum mechanical effects beyond the Hartree term.

The exact form of the exchange-correlation functional $E_{\text{xc}}[\rho]$ is not known, but it is universal, meaning it applies to any electronic system. The exchange-correlation functional can be decomposed as:

$$E_{\text{xc}}[\rho] = (T[\rho] - T_S[\rho]) + (E_{\text{ee}}[\rho] - J[\rho]), \tag{2.13}$$

where:

- $T[\rho]$ is the true kinetic energy functional of the interacting electrons,
- $T_S[\rho]$ is the kinetic energy functional of a non-interacting reference system with the same density $\rho(\mathbf{r})$,
- $E_{\text{ee}}[\rho]$ is the exact electron-electron interaction energy, and
- $J[\rho]$ is the classical Coulomb energy functional representing the electrostatic repulsion energy between the electrons.

Thus, the exchange-correlation functional $E_{\text{xc}}[\rho]$ consists of two key parts: the difference between the true and non-interacting kinetic energy (the correlation energy) and the difference between the exact and classical electron-electron interaction energy (the exchange energy).

2.1.3 The Kohn-Sham formulation

The Kohn-Sham formulation of DFT simplifies the many-body problem by introducing a set of fictitious non-interacting particles that reproduce the exact electron density of the interacting system. The Kohn-Sham equations, derived from the variational principle, describe the behavior of these non-interacting electrons in an effective potential $V_{\text{eff}}(\mathbf{r})$, and are given by:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{\text{eff}}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}), \tag{2.14}$$

where:

- $\phi_i(\mathbf{r})$ are the Kohn-Sham orbitals,
- ε_i are the corresponding orbital energies, and
- $V_{\text{eff}}(\mathbf{r})$ is the effective potential, which includes contributions from:
  - the external potential $V_{\text{ext}}[\rho]$,

- the Hartree potential $V_H[\rho]$, and
- the exchange-correlation potential $V_{\text{xc}}[\rho]$.

The electron density is then constructed as a sum over the square of the Kohn-Sham orbitals:

$$\rho(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2. \tag{2.15}$$

By minimizing the total energy functional $E[\rho]$ with respect to the electron density, subject to the normalization constraint, the Kohn-Sham equations provide a self-consistent solution to the electronic structure problem. These equations, along with the effective potential in the Kohn-Sham formalism, form the basis of practical DFT calculations, allowing for the determination of ground-state properties of complex systems at a relatively low computational cost [206].

Exchange-Correlation functionals

In DFT, the exchange-correlation functional $E_{\text{xc}}[\rho]$ is crucial for determining the accuracy of the calculations. This functional encapsulates the complex many-body interactions within a system, including both exchange interactions, which arise from the Pauli exclusion principle, and correlation interactions, which account for the dynamic electron-electron correlations.

A significant framework for categorizing exchange-correlation functionals was introduced by Perdew, commonly referred to as "Jacob's ladder" of exchange-correlation functionals. This ladder metaphorically represents the progression from simpler to more sophisticated functionals, starting from the Local-Density Approximation (LDA) at the base and moving up to more accurate methods such as hybrid functionals, which incorporate exact exchange.

1. **Local-Density Approximation (LDA)**: LDA is the simplest rung on Jacob's ladder. It assumes that the exchange-correlation energy at each point in space can be approximated by that of a uniform electron gas at the same density. Mathematically, this is expressed as:

$$E_{\text{xc}}^{\text{LDA}}[\rho] = \int \rho(\mathbf{r}) \varepsilon_{\text{xc}}^{\text{unif}}(\rho(\mathbf{r}))\, d\mathbf{r}, \tag{2.16}$$

where $\varepsilon_{\text{xc}}^{\text{unif}}(\rho)$ is the exchange-correlation energy per electron for a uniform electron gas at density $\rho(\mathbf{r})$. While LDA is effective for systems with slowly varying electron densities, it tends to be less accurate for systems with significant density gradients, such as those involving van der Waals forces or hydrogen bonding.

2. **Generalized Gradient Approximation (GGA)**: The next rung is the Generalized Gradient Approximation (GGA), which improves upon LDA by including the gradient of the electron density, $\nabla\rho(\mathbf{r})$, in the functional form. This allows GGA functionals to more accurately account for variations in electron density, making them particularly suitable for systems with non-uniform densities. The exchange-correlation energy in GGA is generally expressed as:

$$E_{\text{xc}}^{\text{GGA}}[\rho] = \int f_{\text{GGA}}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))\, d\mathbf{r}, \tag{2.17}$$

where f_{GGA} is a function that depends on both the local electron density and its gradient. Well-known GGA functionals include PBE (Perdew-Burke-Ernzerhof) and B88 (Becke 88).

3. **Meta-GGAs**: Meta-GGAs build upon GGAs by incorporating additional information, such as the second derivatives of the electron density and/or kinetic energy densities. This allows for a more accurate description of exchange-correlation effects, especially in cases where traditional GGAs might fall short, such as in transition states or weakly interacting systems. A typical Meta-GGA functional is expressed as:

$$E_{\text{xc}}^{\text{meta-GGA}}[\rho] = \int f_{\text{meta-GGA}}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}), \nabla^2\rho(\mathbf{r}), \tau(\mathbf{r}))\, d\mathbf{r}, \tag{2.18}$$

where $\tau(\mathbf{r})$ represents the kinetic energy density.

4. **Hybrid Functionals**: Hybrid functionals combine part of the exact exchange energy from Hartree-Fock theory with DFT approximations. Exact exchange accounts for the antisymmetry of the electronic wavefunction, which removes self-interaction errors found in pure DFT. By blending this with DFT exchange-correlation functionals, hybrid functionals improve accuracy in predicting electronic properties. A typical hybrid functional is written as:

$$E_{\mathrm{xc}}^{\mathrm{hybrid}}[\rho] = aE_{\mathrm{x}}^{\mathrm{exact}}[\rho] + (1-a)E_{\mathrm{x}}^{\mathrm{DFT}}[\rho] + E_{\mathrm{c}}^{\mathrm{DFT}}[\rho], \tag{2.19}$$

where a is a mixing parameter, $E_{\mathrm{x}}^{\mathrm{exact}}[\rho]$ is the exact exchange energy from Hartree-Fock, and $E_{\mathrm{x}}^{\mathrm{DFT}}[\rho]$ and $E_{\mathrm{c}}^{\mathrm{DFT}}[\rho]$ are the DFT exchange and correlation energies. Popular hybrid functionals like B3LYP and PBE0 are widely used for their ability to accurately predict properties of complex systems.

In contrast to the Hartree-Fock model, which approximates electronic structure by assuming the wavefunction can be described by a single Slater determinant, the Kohn-Sham approach within DFT maintains a formally exact framework. The approximation in DFT arises primarily from the specific form chosen for the exchange-correlation functional $E_{\mathrm{xc}}[\rho]$ and its corresponding potential $V_{\mathrm{xc}}[\rho]$.

2.1.4 Basis Sets in DFT Calculations

To solve the Kohn-Sham equations, the electron wavefunctions $\phi_i(\mathbf{r})$ are often expanded in terms of basis sets. In DFT calculations, two types of basis sets are commonly used: plane waves and localized Gaussian functions.

Plane Waves are particularly well-suited for periodic systems, such as crystals, because they naturally align with periodic boundary conditions. In this method, the wavefunction is expanded into a series of sinusoidal functions, which can efficiently describe the periodic nature of crystalline materials. However, representing the wavefunction near atomic nuclei requires a very large number of plane waves, especially due to the rapid variations of the wavefunction in these regions. This can make calculations computationally expensive.

Localized Gaussian Functions, on the other hand, are often used for molecules and finite systems. These basis sets consist of functions that are centered on atoms and decay rapidly with distance, making them highly effective for describing the local environment of atoms in non-periodic systems (such as isolated molecules or clusters). Gaussian basis sets are particularly efficient for systems where electron density is localized around atoms, especially in regions far from nuclei or in covalent bonds. However, while Gaussian functions are generally computationally efficient, the basis sets may become computationally demanding when trying to capture the detailed behavior of the wavefunction near atomic nuclei due to the rapid changes in

electron density in that region. In these cases, one often needs to include additional, specialized functions in the basis set to accurately describe the steep gradients near the nucleus[207].

2.1.5 DFT for material science

As discussed in the previous section, Density Functional Theory (DFT) has revolutionized materials science by providing a robust theoretical framework for predicting material properties with high accuracy. This section addresses the crucial role of pseudopotentials in simplifying complex interactions within materials, the strategic selection of k-points for accurate surface studies, and the importance of periodic boundary conditions in simulating the infinite nature of metallic solids and surfaces.

Pseudopotentials

Pseudopotentials are a fundamental component in Density Functional Theory (DFT) calculations, used to simplify the complex interactions between electrons and atomic nuclei. By replacing the true Coulomb potential of the nucleus and tightly bound core electrons with a smoother, effective potential, pseudopotentials significantly enhance computational efficiency. This approach focuses primarily on the valence electrons, which are most important for determining chemical bonding and material properties.

In the vicinity of atomic nuclei, the electron wavefunctions exhibit rapid oscillations due to the strong Coulomb potential and the need to satisfy orthogonality with core states. Accurately representing these oscillations requires a large number of basis functions, leading to high computational costs. Pseudopotentials address this challenge by smoothing out the potential in the core region, thereby eliminating the need to explicitly account for the core electrons and reducing the number of basis functions required. This allows computational resources to be concentrated on accurately describing the valence electrons.

There are three primary types of pseudopotentials, each tailored to specific computational needs:

- **Norm-conserving pseudopotentials** ensure that the pseudo wavefunction outside a certain cutoff radius matches the true all-electron wavefunction and

that the integral of the charge density within the cutoff radius is conserved. This type maintains physical accuracy while simplifying the wavefunction, making it effective in systems where precise charge distribution is crucial.

- **Ultrasoft pseudopotentials** further reduce the computational burden by allowing for even smoother pseudo wavefunctions that require fewer plane waves for accurate representation. Unlike norm-conserving pseudopotentials, ultrasoft pseudopotentials relax the constraint on charge conservation within the core region. This flexibility reduces the plane-wave cutoff energy needed for convergence, making it especially advantageous in systems with transition metals, heavy elements, or complex materials. For these reasons, as discussed in Chapter 3, we have selected ultrasoft pseudopotentials for our calculations to achieve a balance between computational efficiency and accuracy.

- The **Projector Augmented-Wave (PAW) method** provides a more refined approach by reconstructing the all-electron wavefunction in the core region. The PAW method combines the efficiency of pseudopotentials with the accuracy of all-electron calculations, making it ideal for high-precision simulations that require detailed electron density near the nucleus.

An alternative to plane-wave-based pseudopotential methods is the use of local Gaussian basis sets, as implemented in codes like CRYSTAL [77]. Gaussian-type orbitals can efficiently represent both core and valence electrons without the need for pseudopotentials, allowing for all-electron calculations with reduced computational cost compared to other all-electron methods.

The effectiveness of pseudopotentials depends on their construction and application to specific systems. A critical factor in their development is transferability, the ability of the pseudopotential to accurately represent the true potential across various atomic environments. This becomes particularly challenging in systems where atoms exhibit different oxidation states or coordination geometries [208–210].

Choosing k-points for surface calculations

In DFT, **k**-points are specific points in reciprocal space used to sample the electronic states of a periodic system. They are essential for integrating over the Brillouin zone, which represents the periodicity of the crystal lattice in reciprocal space. Accurate **k**-point sampling is crucial for obtaining reliable results in DFT calculations[77]

When performing surface calculations using slab models, the choice and distribution of **k**-points are particularly important to accurately describe both surface and bulk properties of the material. The Monkhorst-Pack scheme is a widely used method for generating a uniform and systematic grid of **k**-points over the Brillouin zone [211]. This scheme ensures that the integration over reciprocal space is performed efficiently and accurately.

In the context of slab models, the supercell is extended along the surface normal direction (usually the z-axis) due to the inclusion of the vacuum region. As a result, the corresponding reciprocal lattice vector $\mathbf{b}_3$ is much smaller in magnitude compared to $\mathbf{b}_1$ and $\mathbf{b}_2$, which are associated with the directions parallel to the surface. Consequently, the Brillouin zone is much narrower along the k_z direction, and fewer **k**-points are required in that direction for accurate sampling.

For example, in a supercell with lattice vectors:

$$\mathbf{a}_1 = a(1,0,0), \quad \mathbf{a}_2 = a(0,1,0), \quad \mathbf{a}_3 = a(0,0,L),$$

the corresponding reciprocal lattice vectors are:

$$\mathbf{b}_1 = \frac{2\pi}{a}(1,0,0), \quad \mathbf{b}_2 = \frac{2\pi}{a}(0,1,0), \quad \mathbf{b}_3 = \frac{2\pi}{L}(0,0,1).$$

Here, a is the lattice constant, and L is the length of the supercell along the z-axis, including both the slab and the vacuum region. Since $L \gg a$, $|\mathbf{b}_3| \ll |\mathbf{b}_1|, |\mathbf{b}_2|$.

Therefore, a typical approach is to use an $M \times N \times 1$ **k**-point mesh, where $M, N > 1$, to effectively sample the Brillouin zone in the k_x and k_y directions, while using only one **k**-point along k_z.

The selection of **k**-points is critical, as it directly affects the accuracy of DFT calculations. Insufficient sampling can lead to inaccurate results, such as incorrect total energies or densities of states. Conversely, using an excessively dense **k**-point mesh can significantly increase computational costs without proportional gains in accuracy. Therefore, choosing an appropriate **k**-point mesh that balances computational efficiency with the required precision is essential.

Periodicity in bulk and surface metal simulations

Periodicity is a fundamental concept in simulations of crystalline materials, including metallic surfaces. In bulk materials, atoms are arranged in a repeating lattice that extends infinitely in all three dimensions. This periodicity allows us to model extensive systems by considering a single unit cell that is repeated throughout space using periodic boundary conditions (PBC). In surface studies, we adjust this periodicity by introducing a vacuum gap along one dimension, while maintaining periodicity in the other two dimensions to simulate an infinite surface.

Applying PBC provides significant computational advantages by reducing the problem to a finite system while preserving the essential physics of an infinite crystal. A key theoretical foundation for this approach is Bloch's theorem, which simplifies the mathematical treatment of electrons in a periodic potential[77].

Bloch's theorem states that the wavefunction $\psi_{n\mathbf{k}}(\mathbf{r})$ of an electron in a periodic potential can be expressed as:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}),$$

where:

- $\mathbf{k}$ is the crystal wavevector (also called the Bloch wavevector),
- n is the band index,
- $u_{n\mathbf{k}}(\mathbf{r})$ is a function that has the same periodicity as the lattice, i.e., $u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r})$ for any lattice translation vector $\mathbf{R}$.

This theorem indicates that the electron wavefunction in a periodic potential can be written as a plane wave modulated by a function that shares the periodicity of the lattice. This formulation greatly simplifies the calculation of electronic states in crystals.[77]

When applying PBC to a finite-sized simulation cell, we impose that the wavefunction repeats itself after translations by the supercell lattice vectors:

$$\psi(\mathbf{r} + N_i\mathbf{a}_i) = \psi(\mathbf{r}),$$

where N_i is the number of unit cells along the i-th lattice vector $\mathbf{a}_i$.

Using Bloch's theorem, this condition becomes:

$$\psi(\mathbf{r}+N_i\mathbf{a}_i) = e^{i\mathbf{k}\cdot N_i\mathbf{a}_i}\psi(\mathbf{r}).$$

Combining the two expressions, we find that:

$$e^{i\mathbf{k}\cdot N_i\mathbf{a}_i} = 1,$$

which implies that:

$$\mathbf{k}\cdot N_i\mathbf{a}_i = 2\pi m_i,$$

where m_i is an integer. This leads to the quantization of the allowed $\mathbf{k}$-values:

$$k_i = \frac{2\pi m_i}{N_i a_i},$$

where $a_i = |\mathbf{a}_i|$. The quantization of $\mathbf{k}$-points allows us to sample the Brillouin zone with a finite set of wavevectors, which is essential for practical calculations of electronic properties.

In surface simulations using the slab model, a supercell is constructed that contains several layers of the material and a vacuum region to separate periodic images along the surface normal direction. PBCs are applied in all three dimensions, but the vacuum gap ensures that there is minimal interaction between the slabs along the direction perpendicular to the surface.

For accurate modeling, the vacuum region must be large enough so that the electron density decays to nearly zero, preventing interactions between periodic images of the slab. The thickness of the slab and the size of the vacuum gap are chosen to balance computational efficiency with the need for physical accuracy.

In reciprocal space, the periodicity of the supercell leads to a set of reciprocal lattice vectors $\mathbf{b}_i$ defined by:

$$\mathbf{b}_i = 2\pi\frac{\mathbf{a}_j \times \mathbf{a}_k}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)},$$

where (i, j, k) is a cyclic permutation of $(1, 2, 3)$. The quantization of $\mathbf{k}$-points in the supercell corresponds to discrete sampling points in reciprocal space.

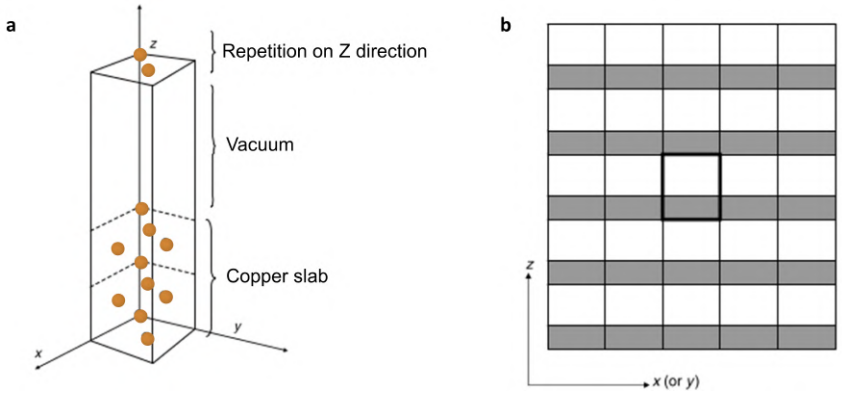


Fig. 2.1 **The slab model.** (a) Representation of a supercell modeling a solid surface with periodic boundary conditions applied in the x, y, and z directions to emulate an infinite material. (b) Two-dimensional schematic of the supercell, showing repeated units demarcated by bold lines. Shaded and white areas represent atomic positions and vacuum, respectively. (c) Perspective of a slab used to model a surface within a fully periodic simulation framework.

In the context of slab calculations, the supercell is typically much larger along the surface normal direction (which we can take as the z-axis) due to the inclusion of the vacuum region. As a result, the corresponding reciprocal lattice vector $\mathbf{b}_3$ is much smaller in magnitude compared to $\mathbf{b}_1$ and $\mathbf{b}_2$. Therefore, fewer $\mathbf{k}$-points are needed along the k_z direction for accurate Brillouin zone integration.

For example, in FCC metals like copper, we can define the supercell lattice vectors as:

$$\mathbf{a}_1 = a(1,0,0), \quad \mathbf{a}_2 = a(0,1,0), \quad \mathbf{a}_3 = a(0,0,L),$$

where a is the lattice constant, and L is the length of the supercell along the z-axis, including both the slab and the vacuum region.

The corresponding reciprocal lattice vectors are:

$$\mathbf{b}_1 = \frac{2\pi}{a}(1,0,0), \quad \mathbf{b}_2 = \frac{2\pi}{a}(0,1,0), \quad \mathbf{b}_3 = \frac{2\pi}{L}(0,0,1).$$

Since $L \gg a$, the magnitude of $\mathbf{b}_3$ is much smaller than that of $\mathbf{b}_1$ and $\mathbf{b}_2$, indicating that the Brillouin zone is much narrower along k_z than along k_x and k_y.

Therefore, when choosing a $\mathbf{k}$-point mesh for the slab calculations, it is common to use a mesh that is dense in the k_x and k_y directions but sparse along k_z, such as an

$M \times N \times 1$ mesh. This approach ensures sufficient sampling of the Brillouin zone while maintaining computational efficiency.

By carefully selecting the number of layers in the slab, the size of the vacuum region, and the **k**-point sampling, we can obtain reliable results that accurately represent the physical properties of the surface under study.

2.1.6 Calculation of surface energies

In the study of metallic surfaces, a fundamental quantity is the surface energy, which indicates the energy required to create a new surface. The basic principle is that the energy needed to split a bulk material into two surfaces is equivalent to the surface energy. This relationship can be mathematically expressed as follows:

$$\gamma = \frac{1}{2A}\left(E_{\text{slab}} - nE_{\text{bulk}}\right) \tag{2.20}$$

where E_{slab} is the total energy of the slab model representing the surface, E_{bulk} is the energy per atom or formula unit in the bulk material, n is the number of atoms or formula units in the slab, and A is the total area of the surfaces (both top and bottom) in the slab model. The calculation of surface energy in DFT involves comparing the energy of the surface to that of the bulk. The surface calculation often uses a larger supercell with a vacuum gap and fewer k-points, while the bulk calculation employs a denser k-point grid and potentially smaller supercells. To minimize discrepancies due to different theoretical treatments, it is crucial to ensure that both energies are well-converged with respect to parameters such as slab layers, k-points, energy cutoff, and supercell size. Notably, when comparing DFT-calculated surface energies for different copper surfaces, it is observed that the surface energy of Cu(111) is lower than that of Cu(100). This suggests greater stability or a more "bulk-like" nature for Cu(111), as shown in Chapter3, aligning with the trend that surfaces with denser packing of surface atoms are usually more stable in simple materials.

2.2 From Density Functional Theory to Molecular Dynamics

DFT provides profound insights into the electronic properties of materials, while Molecular Dynamics offers a dynamic and temporal perspective, enabling us to simulate and understand molecular systems' behavior over time. By complementing DFT, MD shifts focus from electronic structures to the thermodynamic and kinetic properties that govern the interactions and movements of molecules. This section provides a short overview of the theoretical background behind molecular modeling and simulations, laying the groundwork for the presented research. By starting with key concepts in statistical mechanics, we will explore molecular mechanics, force fields, and the fundamental methods of MD. Emphasis will be placed on advanced sampling techniques such as Metadynamics, which are crucial for overcoming the limitations of traditional simulations and offering a more comprehensive understanding of the systems under study.

2.2.1 Molecular Dynamics: theoretical foundations and applications

Classical Molecular Dynamics is a classical simulation approach where electrons are not modeled explicitly, and their presence is approximated. This allows for the simulation of larger time and spatial scales compared to the quantum methods explained previously;

In this context, let's consider a system consisting of N particles in three dimensions, each characterized by their positions $\{\mathbf{r}_1(t),\ldots,\mathbf{r}_N(t)\}$ and momenta $\{\mathbf{p}_1(t),\ldots,\mathbf{p}_N(t)\}$ at any given time t. The behavior of this system is governed by Newton's second law, which, in conjunction with the Hamiltonian formulation of classical mechanics, provides a comprehensive description of its dynamics. The Hamiltonian H of the system is defined as the sum of kinetic and potential energies:

$$H(\{\mathbf{r}_1,\ldots,\mathbf{r}_N,\mathbf{p}_1,\ldots,\mathbf{p}_N\}) = K(\{\mathbf{p}_1,\ldots,\mathbf{p}_N\}) + U(\{\mathbf{r}_1,\ldots,\mathbf{r}_N\}), \tag{2.21}$$

where the kinetic energy K is given by:

$$K(\{\mathbf{p}_1,\ldots,\mathbf{p}_N\}) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i}, \tag{2.22}$$

and $U(\{\mathbf{r}_1,\ldots,\mathbf{r}_N\})$ represents the potential energy of the system, which is a function of the positions of all the particles. The potential energy encompasses all the system's interactions, including intermolecular and intramolecular forces. The equations of motion for the particles are derived from the Hamiltonian:

$$\dot{\mathbf{r}}_i = \frac{\partial H}{\partial \mathbf{p}_i}, \quad \dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{r}_i}. \tag{2.23}$$

These equations describe how the positions and momenta of the particles evolve over time, allowing us to simulate the dynamical behavior of the molecular system under study.

MD solves equations of motion numerically and generates trajectories for a system of N interacting particles. According to Newton's second law, the acceleration of a particle of mass m_i is given by:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i}, \tag{2.24}$$

where the force $\mathbf{F}_i$ acting on particle i is a function of the positions of all N particles.

The solution to the equations of motion requires numerical integration schemes, as analytical solutions are not feasible for general N-particle systems. Integration is done in small steps separated by a fixed time parameter Δt. At each step, the total force on each particle at time t is calculated, and from this, acceleration, velocity, and position are updated to obtain the values at $t+\Delta t$.

One of the common integration methods used in MD simulations is the Verlet algorithm. The Verlet algorithm calculates the position of particles at time $t+\Delta t$ using the following formula:

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \frac{\Delta t^2}{m_i}\mathbf{F}_i(t). \tag{2.25}$$

This method is particularly valued for its numerical stability and simplicity, as it does not explicitly require velocity for the integration of positions, making it highly suitable for systems where only positional data is needed.

The choice of the time step Δt is crucial, as it must be small enough to capture the fastest motions in the system while being large enough to conserve computational efficiency. A general guideline is to select Δt as one-tenth of the period of the fastest motion in the system.

In statistical mechanics, the behavior of a system is analyzed using the concept of *statistical ensembles*, which represent groups of microstates sharing the same thermodynamic properties. These ensembles include the *Micro-Canonical (NVE)*, *Canonical (NVT)*, *Isothermal-Isobaric (NPT)*, and *Grand-Canonical (μVT)* ensembles. The introduction of these ensembles highlighted that understanding the macroscopic properties of a system doesn't require tracking every particle's precise motion. Instead, it relies on averaging over a multitude of microscopic configurations. This is done through the *ensemble average* of a physical observable $A(\mathbf{r},\mathbf{p})$, which is computed as an average over all possible microstates of the system.

$$\langle A \rangle = \iint A(\mathbf{r},\mathbf{p}) P(\mathbf{r},\mathbf{p})\, d\mathbf{r}\, d\mathbf{p}, \tag{2.26}$$

where

$$P(\mathbf{r},\mathbf{p}) = \frac{1}{Z} e^{-\beta H(\mathbf{r},\mathbf{p})}, \tag{2.27}$$

and

$$Z = \iint e^{-\beta H(\mathbf{r},\mathbf{p})}\, d\mathbf{r}\, d\mathbf{p}. \tag{2.28}$$

A key principle in MD is the ergodic hypothesis, which states that, given enough time, a system's dynamics will cover all possible states in phase space equally. This means we can use time averages from a single long simulation to represent ensemble averages, accurately reflecting the system's behavior over time.

$$\langle A \rangle = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} A(\mathbf{r}(t),\mathbf{p}(t))\, dt, \tag{2.29}$$

where

$$\langle A \rangle = \frac{1}{M} \sum_{n=1}^{M} A(\mathbf{r}(t_n), \mathbf{p}(t_n)). \tag{2.30}$$

The choice of the integration domain in simulation time, known as the integration domain, is crucial. It's essential to select an appropriate integration domain relevant to our purpose. The finite timescale that can be explored in MD simulations has significant implications for both the accuracy and interpretation of the results. Limited sampling can introduce errors or biases in the estimation of thermodynamic and kinetic properties, such as free energies, entropies, and reaction rates. Additionally, short simulation times may limit the relevance of MD simulations for systems or phenomena that require longer timescales to be properly captured. To overcome these challenges, various techniques have been developed to accelerate the exploration of phase space and extend the timescales accessible in MD simulations, as discussed in the following sections.

2.2.2 *Ab Initio* Molecular Dynamics

Ab initio molecular dynamics (AIMD) is a computational technique that combines the principles of quantum mechanics with classical molecular dynamics to study the behavior of materials at an atomic level. This method allows for the simulation of systems where electronic structure calculations are essential for understanding their properties and behavior, especially in the context of surface phenomena.

AIMD is based on solving the electronic Schrödinger equation within the framework of DFT alongside the classical Newtonian equations of motion for the nuclei. The system's electronic structure is determined by solving the Kohn-Sham equations, reported previously in equation 2.14

The forces acting on the nuclei are obtained from the Hellmann-Feynman theorem:

$$\mathbf{F}_I = -\nabla_I E_{\text{tot}} = -\nabla_I \left[E_{\text{elec}} + E_{\text{nuc-nuc}} \right], \tag{2.31}$$

where E_{tot} is the total energy of the system, E_{elec} is the electronic energy, and $E_{\text{nuc-nuc}}$ is the nuclear-nuclear repulsion energy.

These forces are then used to integrate the equations of motion for the nuclei, typically using Verlet or other numerical integration algorithms. The combined

quantum-mechanical and classical approach of AIMD makes it particularly well-suited for studying surfaces, where the interaction between adsorbates and the substrate, surface reconstructions, and reaction mechanisms are of interest.

2.2.3 Molecular mechanics and classical force fields

Molecular mechanics, a classical approach within molecular modeling, describes molecular systems using classical mechanics. Molecules are modeled as point masses with charges, connected by springs representing covalent bonds, simulating vibrational, stretching, and angular motions. Non-bonded interactions, like van der Waals forces and electrostatic interactions, are captured using potential functions such as the Lennard-Jones potential and Coulomb's law. The Lennard-Jones potential approximates van der Waals forces, while Coulomb's law models electrostatic interactions between charged atoms. This approach efficiently simulates large systems over long timescales, providing insights into their structural, thermodynamic, and kinetic properties.

In classical MD, the potential energy can be described using a general functional form that depends on the atomic coordinates $\mathbf{r}^N$. Let V be this functional form, also known as the force field, it can be expressed as:

$$V = V_{\text{bonded}} + V_{\text{non-bonded}} \tag{2.32}$$

In Fig. 2.2 it is reported the schematic representation of the main contributions in molecular mechanics force fields, where the following summations give the two terms:

$$V_{\text{bonded}} = V_{\text{bonds}} + V_{\text{angles}} + V_{\text{dihedrals}} \tag{2.33}$$

$$V_{\text{non-bonded}} = V_{\text{electrostatics}} + V_{\text{van der Waals}} \tag{2.34}$$

The terms in the above equations can be modeled using different approaches, with various functional forms and parameter sets. Numerous molecular modeling force fields in the literature adopt a functional form that characterizes the energy variations arising from deviations in bond lengths or angles, bond rotations, and interactions among non-bonded atoms:

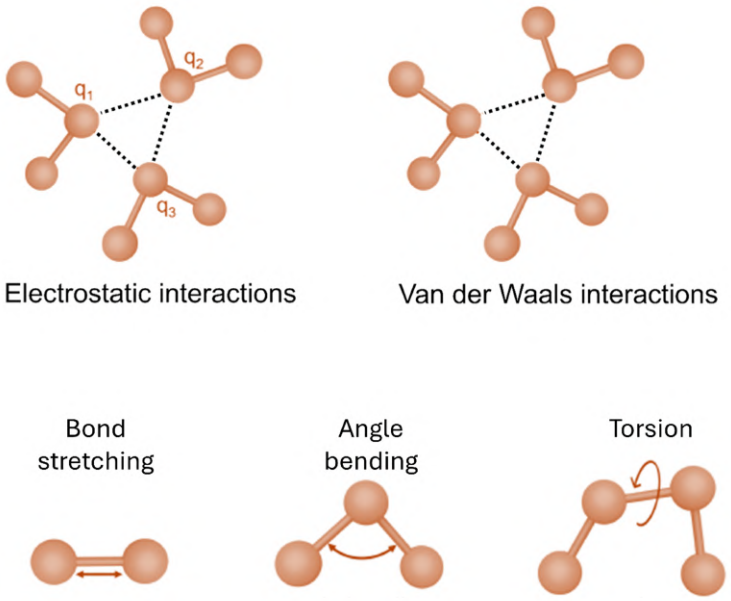


Fig. 2.2 **Representation of Bonded and Non-Bonded Interactions**: Schematic representation of key molecular interactions in molecular mechanics. The top row illustrates non-bonded interactions: electrostatic interactions (left) and van der Waals interactions (right). The bottom row depicts bonded interactions: bond stretching (left), angle bending (middle), and torsion motions (right). These interactions are fundamental in defining the potential energy components in molecular mechanics.

$$
\begin{aligned}
V(\mathbf{r}^N) = & \sum_{\text{bonds}} \frac{k_{b_i}}{2}(l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_{\theta_i}}{2}(\theta_i - \theta_{i,0})^2 \\
& + \sum_{\text{dihedrals}} \frac{V_{n_i}}{2}\left(1 + \cos\left(n_i \omega_i - \gamma_i\right)\right) \\
& + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)
\end{aligned}
\tag{2.35}
$$

In this general form of FF, the initial term models the interaction between bonded atoms using a harmonic potential, where k_{b_i} represents the bond stretching constant, $l_{i,0}$ is the equilibrium bond length (the bond length when all other force field terms are zero), and l_i is the actual bond length. The second term accounts for angle deviations from their equilibrium values through a harmonic potential, where k_{θ_i} is the force constant, $\theta_{i,0}$ is the equilibrium angle, and θ_i is the actual angle. The third term represents a torsional potential, depicting the energy changes due to bond

rotations: V_{n_i} corresponds to the rotation barrier, n_i indicates the periodicity (the number of minima in the function), ω_i is the dihedral angle, and γ_i is the phase angle. The final term describes non-bonded interactions, typically modeled using van der Waals forces and Coulombic interactions, with the Lennard-Jones potential approximating the van der Waals forces and Coulomb's law governing electrostatic interactions.

Employing internal coordinates, such as bond lengths, angles, and dihedrals, rather than absolute atomic positions, offers significant advantages. Internal coordinates lead to a more intuitive potential energy function, directly linked to structural changes in the molecule, and provide a clearer physical interpretation of each term. This approach also simplifies the parametrization process, enhancing the transferability of the force field across different molecular systems.

Van der Waals interactions

The van der Waals forces are commonly described using the Lennard-Jones 12-6 pair potential:

$$V_{\mathrm{LJ}} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

The Lennard-Jones potential function, illustrated in Fig. 2.3, includes a short-range repulsive term that scales with $1/r^{12}$, modeling the Pauli exclusion principle's repulsion between electron clouds, and a long-range attractive term that scales with $1/r^6$, representing van der Waals attraction. This potential is advantageous because it only requires two parameters: σ_{ij}, the collision diameter, indicating the distance where the potential energy is zero, and ε_{ij}, the depth of the potential well, indicating the strength of the interaction.

The parameters ε and σ are typically provided for pure substances. Cross-interaction parameters σ_{ij} and ε_{ij} for different species can be determined using mixing rules. The Lorentz-Berthelot rule [212] is commonly used, where σ_{ij} is the arithmetic mean and ε_{ij} is the geometric mean of the pure species parameters:

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}), \quad \varepsilon_{ij} = \sqrt{\varepsilon_{ii} \cdot \varepsilon_{jj}}$$

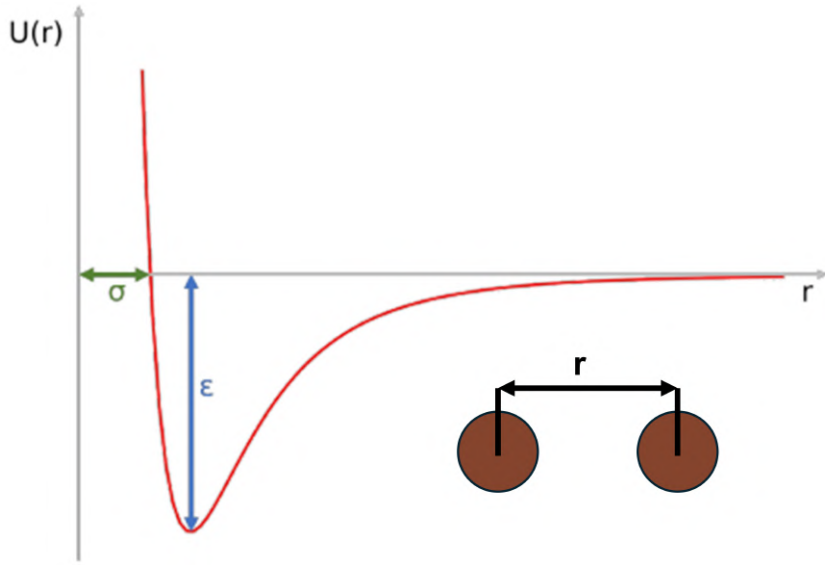


Fig. 2.3 Graphical representation of the Lennard-Jones potential, $U(r)$, showing the relationship between potential energy and interatomic distance r. Key features include the equilibrium distance σ where the potential is zero, and the well depth ε, indicating the strength of the attractive forces.

The computational cost of evaluating these non-bonded interactions can be substantial, as it scales with the square of the number of particles. However, because van der Waals interactions are short-range and decay rapidly with distance, they can be efficiently calculated using truncation methods, where interactions are only considered within a specified cutoff distance. Beyond this cutoff, the potential can be ignored or approximated using tail-correction methods.

While the Lennard-Jones potential is widely used, there are other methods to model van der Waals interactions in molecular dynamics. For example, the Buckingham potential, which uses an exponential term for repulsion, and the Morse potential, which provides a more accurate description of bond interactions, are alternative approaches[213–215]. Additionally, the use of polarizable force fields can provide a more accurate depiction of van der Waals interactions by allowing the electron cloud to respond to the local environment dynamically.

Modeling Electrostatic Interactions

The electrostatic interaction between two atoms is determined using Coulomb's law:

$$V_{\text{electrostatic}} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

where q_i and q_j are the partial charges assigned to atoms i and j, r_{ij} is the distance separating these atoms, and ε_0 is the vacuum permittivity. To efficiently calculate these long-range interactions in systems with periodic boundaries, and to avoid significant inaccuracies that simple cutoff methods might introduce, advanced techniques such as the Ewald summation[216] and fast multipole methods[217] are often employed.

2.2.4 Force Field parametrization

Creating and parameterizing a classical force field for non-standard molecules involves several challenges. The initial steps include constructing the molecular structure and assigning atom types, which encode the atom's hybridization and surrounding environment, both critical for the force field.

A key aspect of force field parametrization is determining the molecular charge distribution, often represented by partial atomic charges centered on the atoms. These charges can be derived from the electrostatic potential, typically by fitting them to quantum mechanical electrostatic potential data around the molecule. However, such quantum mechanical calculations can be computationally prohibitive for large systems. In these cases, semi-empirical methods like Austin Model 1 (AM1) are employed. AM1 simplifies the calculation by using empirical parameters combined with quantum mechanics[218].

For biological systems, experimental data from sources such as X-ray or neutron diffraction studies, and databases like the Protein Data Bank or the Cambridge Structural Database, are essential for parameter derivation. In the absence of experimental data, quantum mechanical calculations provide necessary parameters.

Force fields are typically optimized for specific molecular classes and may perform poorly when applied to different systems. For instance, force fields like AMBER [219], CHARMM [220], and GROMOS [221] are specifically tuned for biomolecules such as proteins, lipids, and nucleic acids. The concept of transferability is critical, enabling the same parameter set to be used across various molecules. However, if a force field lacks parameters for the molecules under study, its predictive accuracy may be compromised. To address such gaps, the General Amber

Force Field (GAFF)[222] was developed, extending AMBER to include parameters suitable for a broad range of organic molecules.

2.2.5 Generalized AMBER Force Field (GAFF)

GAFF, considered one of the most comprehensive and versatile force fields has a structure similar to other AMBER force fields and is expressed as follows:

$$V(\mathbf{r}^N) = \sum_{\text{bonds}} k_{b_i}(l_i - l_{i,0})^2 + \sum_{\text{angles}} k_{\theta_i}(\theta_i - \theta_{i,0})^2 + \sum_{\text{dihedrals}} \frac{V_{n_i}}{2}(1 + \cos(n_i\omega_i - \gamma_i)) + \sum_{i<j}^{N} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right] \tag{2.36}$$

For calculating partial atomic charges in GAFF, the restrained electrostatic potential (RESP) algorithm is commonly employed [223]. This method derives from the electrostatic potential technique (ESP), incorporating restraints on non-hydrogen atoms to enhance accuracy. The derivation of RESP charges involves ab initio calculations, generally using the Hartree-Fock method with the *6-31G** basis set.

2.3 Classical force field for metal systems

Applying classical force fields, such as pair potentials, to simulate metallic materials poses significant challenges due to the peculiarities of the metallic bond, as described in Chapter 1. The classical force field methods, shown in previous sections, based on pairwise interactions between atoms, are effective in modeling the ionic and covalent bonds typical of non-metallic materials. However, accurately describing the motion of atoms in a metal requires a force field that can account for the complex interactions characteristic of metallic bonds while also being computationally accessible for large system sizes and long simulation times.

To achieve this balance, multibody force fields are introduced. These force fields consider not just pairwise interactions but also the collective effect of multiple neighboring atoms on each atom's behavior, thereby providing a more accurate representation of atomic interactions in metals. The necessity of this approach

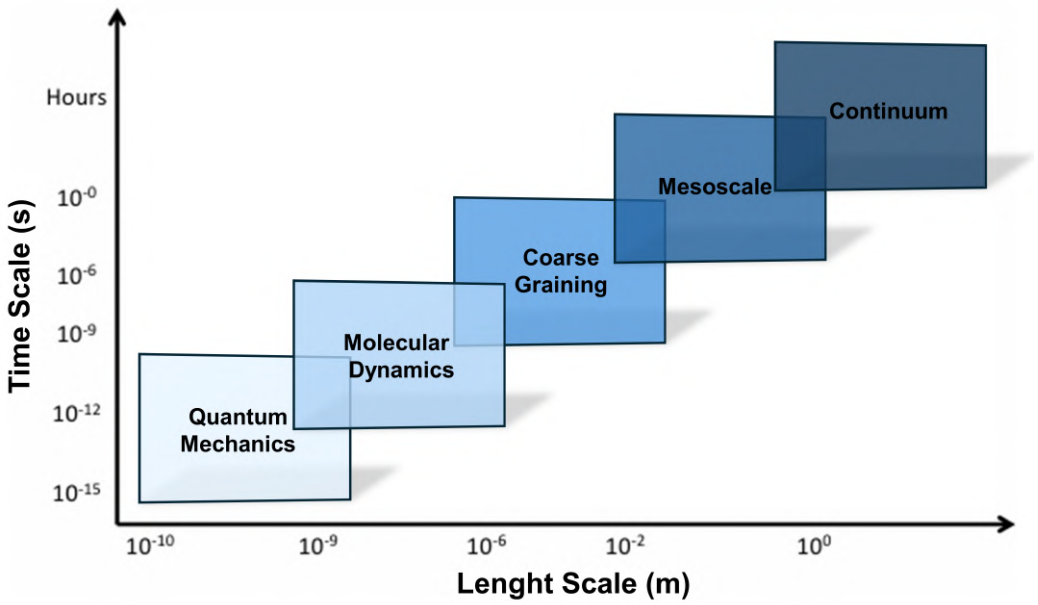


Fig. 2.4 **Illustration of the concept of multiscale modeling.** This figure highlights how spatial resolution correlates with observable temporal scales. The x-axis spans from quantum mechanical to macroscopic distances, while the y-axis ranges from femtoseconds to hours, demonstrating the adaptability of modeling techniques to capture phenomena at varying scales.

arises from the unique nature of metallic bonds, where electrons are not localized between specific atoms but are instead delocalized over the entire metallic structure. This delocalization leads to a complex interplay of forces that cannot be accurately captured by simple pair potentials.

In metallic systems, the properties and behavior of each atom are significantly influenced by its local environment, including the positions and types of neighboring atoms. This means that the energy of a given atom is not just a function of its distance to another single atom but is also affected by the cumulative effect of the surrounding atoms.

The Embedded-Atom Method (EAM), pioneered by M. S. Daw and M. I. Baskes in 1984 [140], addresses these complexities by incorporating the influence of the entire local atomic environment. Unlike classical force fields that treat atomic interactions as simple pairwise forces, EAM accounts for the effect of local electron density on each atom. This method has proven particularly effective in capturing atomic phenomena in metallic systems, including impurities, surfaces, and defects, providing a more realistic and accurate description of metallic bonding.

EAM is based on the concept that the energy of an atom in a metallic environment can be described as a function of the local electron density contributed by its neighboring atoms. This approach allows for the simulation of more realistic metallic behavior, enabling the study of complex phenomena such as plastic deformation, phase transitions, and surface reconstructions with higher accuracy and computational efficiency. In this context, we will explore how the EAM method represents an effective alternative to classical force fields for modeling metallic systems, providing insights into their atomic-scale behavior with greater fidelity. Mathematically, the total energy E of a metallic system is given by:

$$E = \sum_{i=1}^{N} F_i(\rho_i) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i} \phi_{ij}(R_{ij}) \tag{2.37}$$

where:

- $F_i(\rho_i)$ represents the embedding energy function, which describes the energy of atom i as a function of the local electron density ρ_i. This term reflects the cohesive energy the surrounding atoms contribute and captures the effects of metallic bonding.
- $\phi_{ij}(R_{ij})$ is the pair potential function that represents the direct interaction between two atoms i and j, separated by distance R_{ij}. This term usually models electrostatic and van der Waals forces and often has a repulsive core and an attractive tail.

The local electron density ρ_i at atom i is computed as:

$$\rho_i = \sum_{j \neq i} \rho_j(R_{ij}) \tag{2.38}$$

where $\rho_j(R_{ij})$ represents the electron density contribution from atom j at a distance R_{ij} from atom i. The function $\rho_j(R_{ij})$ is often derived empirically or from first-principles calculations to accurately model the physical behavior of metals. EAM has been particularly useful in predicting properties like surface energy, lattice constants, and vacancy formation energy, formation of fractures and defects in metallic systems.[224–226]

While the EAM is known for its accuracy in modeling many-body interactions within metallic systems, it is computationally intensive. The Gupta potential, how-

ever, employs a more computationally efficient method [227]. It is rooted in the tight-binding theory, specifically using the second-moment approximation of the density of states to capture metallic bonding. This approximation simplifies metallic bonding representations, which is particularly advantageous for studies involving complex metallic systems such as alloys and nanoparticles.

The mathematical expression for the Gupta potential is:

$$E = \sum_{i=1}^{N} \left(A \sqrt{\sum_{j \neq i} e^{-2q\left(\frac{R_{ij}}{r_0}-1\right)}} - \sum_{j \neq i} B e^{-p\left(\frac{R_{ij}}{r_0}-1\right)} \right) \tag{2.39}$$

where:

- A and B are empirical constants tuned through experimental or theoretical data to define the interaction strengths.
- p and q adjust the decay characteristics of the potential.
- r_0 is typically set at the equilibrium bond length found in the bulk metallic lattice.

The second-moment approximation relies on the concept that the energy of an atom in a metallic system is influenced by the local electron density, which is related to the square root of the sum of the squared bond lengths (hence, the "second moment" of the bond length distribution). This allows the Gupta potential to capture many-body effects indirectly by focusing on the statistical distribution of bond lengths around each atom, rather than explicitly modeling all interactions.

This approximation allows the Gupta potential to distinguish itself by effectively approximating the many-body interactions found in metallic bonding while reducing the computational complexity. It focuses on the local density effects around each atom, thereby enhancing its performance and accuracy in predicting the properties of nanostructures.

Gupta's formulation addresses a critical limitation in classical potentials like Morse and Lennard-Jones, which typically predict an expansion in interlayer separations that contradicts experimental observations of contraction in many metallic surfaces. These features make the Gupta potential especially suitable for simulations of metallic nanoparticles. Its efficient computation of energy states based on local

electron densities enables accurate predictions of structural properties in nanoparticles, notably gold, where traditional methods may falter due to computational constraints. For these reasons, we chose this potential to simulate these systems, as discussed in Chapters 4 and 5.

2.4 Machine Learning potentials and DeePMD simulations

In the field of molecular dynamics, machine learning potentials—particularly those implemented in the Deep Potential Molecular Dynamics (DeePMD) framework [180]—have significantly advanced the study of metallic systems. These potentials combine the accuracy characteristic of *ab initio* methods with the computational efficiency of classical force fields.

DeePMD is a novel approach that utilizes deep learning algorithms to accurately predict the potential energy and forces in molecular systems. This method has proven to be especially effective in studying systems where traditional computational approaches may be inadequate due to computational cost or inability to capture complex interactions.

Potential Energy Representation

At the core of the DeePMD representation is a deep neural network that models the potential energy of a system based on atomic positions. The total potential energy $V(\mathbf{R})$ is expressed as the cumulative sum of individual atomic contributions:

$$V(\mathbf{R}) = \sum_{i=1}^{N} V_i(\mathbf{R}), \tag{2.40}$$

where V_i denotes the potential energy contribution from the i-th atom, $\mathbf{R}$ represents the set of atomic positions, and N is the total number of atoms.

A foundational concept of neural network potentials, including DeePMD, is the incorporation of local atomic environments. This approach considers not only the positions of individual atoms but also the relative positions of neighboring atoms within a specified cutoff radius. DeePMD enhances this methodology by introducing

a local coordinate frame for each atom. This ensures that the potential energy model remains extensive and adheres to symmetry invariance properties, such as translational, rotational, and permutational invariance. By assigning a local reference frame, DeePMD accurately captures the intricate interactions within the system.

More precisely, the local atomic environments are processed using two key components: the *embedding network* and the *fitting network*. The **embedding network** transforms the relative positions of neighboring atoms into invariant descriptors. It processes the distances and angular information between the central atom and its neighbors, applying smooth and continuous functions to ensure differentiability. This transformation results in descriptors that are invariant under translations, rotations, and permutations of identical atoms, preserving the essential physical symmetries of the system.

These invariant descriptors are then fed into the **fitting network**, a deep neural network that predicts the atomic energy contributions V_i. The fitting network maps the complex, non-linear relationships captured by the descriptors to the potential energy of each atom. Each local environment is described using a smooth and adaptive embedding network that transforms the coordinates of neighboring atoms into a lower-dimensional space in a continuous and differentiable manner. This approach addresses the limitations of non-smooth models, which can introduce discontinuities in the potential energy surface, leading to inaccuracies in force calculations and unstable dynamics.

DeePMD's potential energy representation is extensive—scaling linearly with system size—and invariant under system translations and rotations. This consistency ensures that the energy remains the same regardless of the system's absolute position or orientation, preserving the physical realism of simulations.

Network architecture and optimization

DeePMD employs an advanced multi-layer neural network architecture to accurately model complex interatomic interactions. The core of this approach lies in the ability of neural networks to learn intricate relationships within data, which is crucial when dealing with the non-linear and highly variable nature of interatomic potentials.

The neural network in DeePMD processes data through multiple hidden layers, each consisting of neurons. In each layer l, the input to a neuron is a linear combination of the outputs (or activations) from the neurons in the previous layer $l-1$. This

linear combination is computed as a weighted sum of these activations, along with a bias term. Mathematically, for a neuron j in layer l, this transformation is expressed as:

$$z_j^{(l)} = \sum_{i=1}^{n_{l-1}} w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)},$$

where:

- $w_{ij}^{(l)}$ is the weight connecting the i-th neuron in layer $l-1$ to the j-th neuron in layer l,
- $a_i^{(l-1)}$ is the activation of the i-th neuron in layer $l-1$,
- $b_j^{(l)}$ is the bias term for neuron j in layer l,
- n_{l-1} is the number of neurons in layer $l-1$.

The neuron then applies a non-linear activation function ϕ to this linear combination to produce its output:

$$a_j^{(l)} = \phi(z_j^{(l)}).$$

The role of this non-linearity is to enable the network to capture and model the complex, non-linear relationships inherent in interatomic potentials.

In DeePMD, the hyperbolic tangent function (tanh) is commonly used as the activation function. The tanh function is defined as:

$$\phi(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}},$$

which outputs values in the range between -1 and 1. Applying this function to $z_j^{(l)}$ normalizes the neuron outputs and helps prevent issues such as vanishing gradients during training.

To optimize the network's parameters, DeePMD utilizes the ADAM optimizer, an adaptive variant of stochastic gradient descent. The ADAM optimizer is well-suited for training deep networks on large datasets, as it dynamically adjusts the learning rates of individual parameters based on estimates of the first and second moments of

the gradients. This helps achieve faster convergence and improves the stability of the training process. The update rule for ADAM is given by:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \varepsilon} \hat{m}_t,$$

where:

- θ represents the network parameters (including weights $w_{ij}^{(l)}$ and biases $b_j^{(l)}$),
- η is the learning rate,
- $\hat{m}_t$ and $\hat{v}_t$ are estimates of the first and second moments of the gradients,
- ε is a small constant added to prevent division by zero.

The training process involves minimizing a specially designed loss function, which quantifies the difference between the network's predictions and the reference data derived from quantum mechanical calculations. The loss function used in DeePMD is a weighted sum of the errors in predicting the energy, forces, and virial tensor, which are critical quantities in molecular simulations. The loss function is defined as:

$$L(p_\varepsilon, p_f, p_\xi) = p_\varepsilon \Delta\varepsilon^2 + \frac{p_f}{3N} \sum_{i=1}^{N} |\Delta F_i|^2 + \frac{p_\xi}{9} \|\Delta\xi\|^2,$$

where $\Delta\varepsilon$ denotes the difference between the predicted and reference energy per atom, ΔF_i represents the difference in forces on atom i, and $\Delta\xi$ is the difference in the virial tensor. The prefactors p_ε, p_f, and p_ξ are tunable parameters that control the relative importance of each term in the loss function, allowing for balanced training across different physical quantities. This flexibility is crucial because it enables the network to learn the most relevant features of the system under study, whether that be energy, forces, or stress.

One of the most significant advantages of DeePMD is its scalability. The architecture is designed to support efficient parallelization, which is essential for handling large-scale molecular simulations; the computational cost of DeePMD scales linearly with the number of atoms, making it suitable for studying complex systems, such as metallic materials, where traditional methods might be computationally too expensive.

2.4.1 DeePMD-kit training procedure

The DeePMD framework works as a bridge between AIMD and classical MD. Its application to metallic systems has consistently demonstrated its capability to accurately capture complex behaviors, enabling the investigation of larger systems over extended timescales. As described in section 3, we have chosen DeePMD to train a ML potential for copper surfaces.

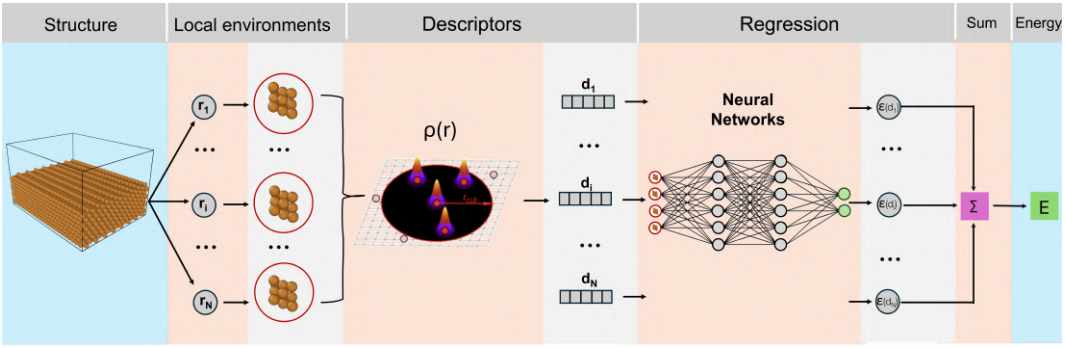


Fig. 2.5 **Schematic representation of the development process for a machine learning potential using deep neural networks.** The workflow is divided into several key stages: **(1) Structure** shows the atomic structure of a metallic slab. **(2) Local Environments** illustrates individual atoms and their nearest neighbors within the structure. **(3) Descriptors** calculate a probability density function, $\rho(r)$, capturing the spatial distribution of atoms around a central atom, which encodes geometric and topological local information. **(4) Regression *via* Neural Networks** involves feeding the computed descriptors into a neural network trained to approximate the energy function based on these descriptors. **(5) Sum** aggregates the energy contributions from each local environment predicted by the neural network to compute the total system energy, E.

To better understand the process of developing a machine learning force field using DeePMD-kit for such complex systems, we outline the primary steps:

1. **Data Collection and Preprocessing**: A comprehensive dataset was compiled, including atomic coordinates, potential energies, and forces, all computed with DFT-level accuracy. The dataset was meticulously preprocessed to ensure it was fully optimized for use in training the ML model

2. **Neural Network Architecture Setup**: We employed a deep, feedforward network structure, selecting the appropriate number of hidden layers and neurons per layer, choosing the hyperbolic tangent as activation function.

3. **Hyperparameter Configuration**: Critical hyperparameters, including the learning rate, batch size, and training epochs, were carefully configured.

4. **Model Training Process**: Using the prepared dataset, we trained the neural network, continuously monitoring and adjusting hyperparameters based on the loss function to optimize model performance.

5. **Validation and Testing for Generalization**: The model's generalization capabilities and efficacy were assessed using separate validation and test datasets, ensuring an accurate representation of the system's physical properties.

6. **Analysis and Interpretation of Results**: Post-training, detailed analysis and interpretation of the model's predictions were conducted against reference data to confirm the accuracy and reliability of the ML force field.

We will discuss the validation and application of this potential in the dedicated section 3.2.2 in Chapter 3.

2.5 Enhanced Sampling Molecular Dynamics: Metadynamics

As discussed before, MD simulations are indispensable for understanding the dynamic behavior of molecular systems. However, conventional MD often struggles to efficiently explore phase space, especially in complex energy landscapes with multiple metastable states—local minima in the potential energy surface where the system can become trapped. This trapping limits the sampling of configurations, leading to incomplete and potentially biased insights into the system's behavior. Insufficient sampling can result in inaccurate free energy surfaces, which are crucial for understanding molecular interactions, reaction mechanisms, and conformational changes.

To overcome these limitations, enhanced sampling methods such as Metadynamics (MetaD) have been developed [228, 229]. MetaD introduces a history-dependent bias potential that encourages the system to escape from local minima and explore a broader region of phase space, ultimately providing a more accurate estimation of the Free Energy Surface (FES).

The key concept in MetaD is the introduction of collective variables (CVs), which are functions of the system's atomic coordinates $\mathbf{R}$. Each CV is defined as:

$$s_i = \Pi_i(\mathbf{R}),$$

where Π_i maps the high-dimensional coordinate space $\mathbf{R}$ to a lower-dimensional space representing essential degrees of freedom. The CVs capture crucial motions within the system, such as transitions between metastable states and reaction pathways. Selecting appropriate CVs is vital for the efficiency and accuracy of the sampling process.

In MetaD, the potential energy function of the system is modified to include a time-dependent bias:

$$\widetilde{U}_{\text{pot}}(\mathbf{R},t) = U_{\text{pot}}(\mathbf{R}) + V_G(\mathbf{s}(\mathbf{R}),t),$$

where $V_G(\mathbf{s}(\mathbf{R}),t)$ is the bias potential applied to the CVs, and $\mathbf{s}(\mathbf{R}) = \{s_1(\mathbf{R}),\ldots,s_N(\mathbf{R})\}$.

The bias potential is introduced as a sum of Gaussian functions added incrementally to the CV space based on the system's sampling history:

$$V_G(\mathbf{s},t) = \sum_{t'=\tau_G,2\tau_G,\ldots}^{t} W \exp\left(-\sum_{i=1}^{N} \frac{(s_i - s_i(t'))^2}{2\sigma_i^2}\right),$$

where:

- $\mathbf{s} = \{s_1, s_2, \ldots, s_N\}$ represents the set of CVs,
- σ_i is the width of the Gaussian function associated with the i-th CV,
- W is the height of the Gaussian functions,
- τ_G is the time interval between successive Gaussian depositions,
- $s_i(t')$ is the value of the i-th CV at time t'.

As these Gaussians accumulate, the bias potential $V_G(\mathbf{s},t)$ modifies the underlying FES, allowing the system to overcome energy barriers and explore new regions of phase space. The cumulative effect flattens the energy landscape, enhancing the sampling of different states. The inverse of the deposited potential can then be used as an estimate of the free energy $F(\mathbf{s})$.

The probability distribution function associated with the collective variables $P(\mathbf{s})$, is directly related to the Boltzmann distribution of the system's atomic coordinates $\mathbf{R}$:

$$P(\mathbf{s}) = \int d\mathbf{R}\, P(\mathbf{R})\, \delta\big(\mathbf{s} - \Pi(\mathbf{R})\big),$$

where:

- $P(\mathbf{R}) = \dfrac{1}{Z} \exp\big(-\beta U_{\text{pot}}(\mathbf{R})\big)$ is the Boltzmann probability distribution, with $\beta = \dfrac{1}{k_B T}$, with k_B being the Boltzmann constant and T the temperature, and Z is the partition function,
- $\delta\big(\mathbf{s} - \Pi(\mathbf{R})\big)$ is the multidimensional Dirac delta function,
- $\Pi(\mathbf{R}) = \{\Pi_1(\mathbf{R}), \Pi_2(\mathbf{R}), \ldots, \Pi_N(\mathbf{R})\}$ represents the set of collective variables as functions of the atomic coordinates.

The free energy as a function of the CVs $\mathbf{s}$ is then defined by:

$$F(\mathbf{s}) = -\frac{1}{\beta} \ln P(\mathbf{s}).$$

This expression links the probability distribution $P(\mathbf{s})$ to the free energy landscape, allowing the reconstruction of $F(\mathbf{s})$ from the sampled distribution.

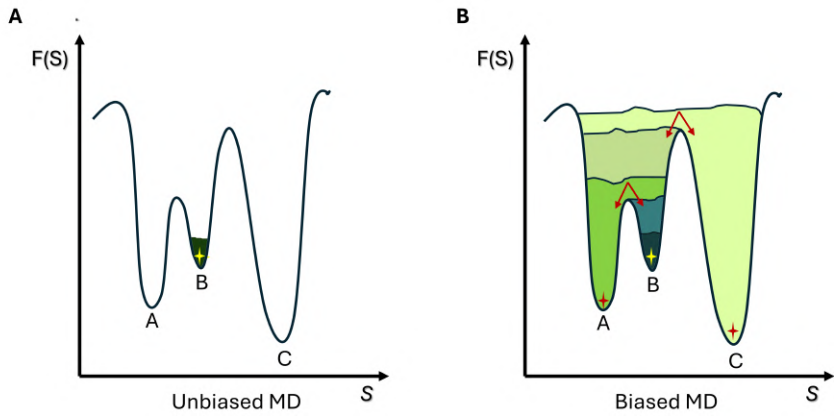


Fig. 2.6 **Schematic comparison of MD simulations demonstrating the effects of potential energy surface exploration with and without bias**. **(A)** In an unbiased MD simulation, the system explores accessible configurations solely based on thermal fluctuations, often becoming confined to local minima, such as at point B. **(B)** In a biased MD simulation, an energy bias is introduced, which alters the free energy landscape, depicted by the shaded area. This modification enables the system to explore previously inaccessible regions of the phase space, allowing a comprehensive exploration from point A through B to C, thus overcoming the barrier limitations present in the unbiased scenario.

However, MetaD also presents challenges. Determining when a simulation has converged is non-trivial. Continuous bias deposition can introduce artifacts in the energy landscape, potentially pushing the system into physically unrealistic regions. Therefore, careful monitoring is required to avoid over-biasing and to ensure that the reconstructed FES is accurate.

Different versions of MetaD have been developed to improve the method's efficiency and accuracy, particularly in how the bias $V_G(s,t)$ is added and how the original (unbiased) FES profile is retrieved.

Well-Tempered Metadynamics

Well-Tempered Metadynamics (WT-MetaD) is a variation of MetaD that addresses the issue of bias over-deposition by gradually decreasing the rate of bias addition over time. This ensures a more controlled and systematic exploration of the phase space [230]. In WT-MetaD, the height of each Gaussian added to the potential is rescaled based on the time already spent at that point in the CV space. The potential in WT-MetaD is given by:

$$V(s,t) = k_B \Delta T \ln\left(1 + \frac{\omega N(s,t)}{k_B \Delta T}\right),$$

where:

- $N(s,t)$ is the histogram of the variable s sampled in the biased simulation,
- ΔT is related to the **bias factor** $\gamma = \frac{T+\Delta T}{T}$, controlling how much bias is added compared to the system temperature T,
- ω is a frequency parameter controlling how often bias is added.

In this context, ΔT helps control the amount of bias applied relative to the physical temperature. A higher bias factor γ leads to smoother exploration of the free energy surface, while a lower bias factor speeds up the system's convergence.

As WT-MetaD evolves, it converges to a modified free energy expression:

$$V(s,t \to \infty) = -\frac{1}{\gamma} F(s) + C,$$

where C is a constant and $F(s)$ is the free energy. This shows that WT-MetaD leads to a modified free energy surface, with the bias factor γ controlling the exploration of the free energy landscape.

WT-MetaD has been successfully applied to study various conformational changes in molecular systems. For instance, it is particularly useful in investigating isomerization processes, as discussed in Chapter 7.

Infrequent Metadynamics

Infrequent Metadynamics, an extension of the WT-MetaD framework, is utilized to estimate kinetic rates for rare events.

The main idea behind Infrequent Metadynamics is to deposit bias infrequently enough so that the system has sufficient time to naturally explore the configuration space without being artificially driven over energy barriers. By doing so, it is possible to collect unbiased transition times and construct an empirical distribution of these escape times. This method helps in understanding the kinetics of rare events, which are typically difficult to capture due to their low occurrence probability within conventional simulation timescales.

In practice, multiple independent simulations are conducted to gather a robust set of escape times. Each simulation starts from the same initial state but evolves independently, providing a diverse sampling of transition events. The collected escape times are used to build an empirical distribution, which is compared to the ideal Poisson distribution expected for rare events. This comparison is often performed using statistical tests such as the Kolmogorov-Smirnov test, which assesses the goodness-of-fit between the empirical distribution and the Poisson distribution.

Unlike conventional MetaD, which focuses on achieving convergence of the FES, Infrequent Metadynamics aims to cross energy barriers to observe transition events. Therefore, full convergence of the FES is not necessary; instead, the goal is to collect accurate transition times reflecting the system's natural kinetics. The unbiased transition time t_{un} can be obtained using:

$$t_{\text{un}} = \sum_{i=1}^{n_{\text{MD}}} \Delta t \, e^{\beta V(s(t_i),t_i)}, \tag{2.41}$$

where t_{un} is the unbiased transition time, n_{MD} is the total number of MD steps, Δt is the MD time step, $V(s(t_i),t_i)$ is the Metadynamics bias at time t_i, and $s(t_i)$ is the collective variable at time t_i.

The reliability of the dynamics reconstructed from Metad can be assessed by performing a statistical test that evaluates how well the computed distribution fits with the ideal Poisson distribution expected for rare events:

$$P_{n\geq 1} = 1 - P_0 = 1 - e^{-t/\tau} \tag{2.42}$$

where τ is the characteristic timescale of the transition.

Convergence criteria and limitations

Convergence in Metadynamics is typically assessed by monitoring how effectively the system explores the collective variable space and how consistent the free energy profiles remain over time. A common method for quantifying convergence is by observing the time-dependent function $c(t)$, which estimates the reversible work done by the bias potential during the simulation. This function can be used as an indicator of whether the system has sufficiently explored the phase space. To quantitatively measure this, block-analysis techniques are often employed, wherein the simulation

data is divided into time blocks to evaluate the stability and consistency of the free energy estimates over time.

While WT-MetaD enhances convergence by tempering the bias potential, it presents certain challenges. One notable limitation is the requirement for prior knowledge of the free-energy surface to properly select the bias factor, which governs how quickly the bias is applied. Additionally, accurately identifying a small number of effective CVs for complex processes is essential for efficient sampling, but this can be difficult for high-dimensional systems.

Reweighting Metadynamics bias

To obtain the unbiased probability distribution from a biased simulation, reweighting techniques are employed. This approach, pioneered by Tiwary and Parrinello, takes into account the time dependence of the bias potential[231]. The unbiased average of any observable O dependent on atomic positions is calculated as follows:

$$\langle O(\mathbf{r})\rangle = \frac{\langle O(\mathbf{r})e^{\beta[V(s,t)-c(t)]}\rangle_{\text{MetaD}}}{\langle e^{\beta[V(s,t)-c(t)]}\rangle_{\text{MetaD}}}, \tag{2.43}$$

where $c(t)$ is a time-dependent constant related to the Metadynamics bias.

2.5.1 Limitations in the selection of CVs for Metadynamics

One of the primary challenges in metadynamics is selecting CVs that accurately represent the physical phenomena under study. CVs, functions of the atomic coordinates, are usually chosen based on physical and chemical insights, providing a low-dimensional projection of the crucial degrees of freedom in the original conformational space. The selection of CVs should meet two criteria: different metastable states should be clearly separated in the projected phase space, and the CVs should efficiently sample the transition states. Poorly chosen CVs can lead to incorrect free-energy estimates, fail to accelerate the dynamics, and offer no improvement in sampling. If the CVs do not capture the essential dynamics and transitions within the system, the FES may fail to converge correctly, causing distinct states to merge indistinguishably.

Moreover, the number of CVs should be kept low, as exploring a high-dimensional space requires more computational time and makes convergence difficult. The sub-

jective nature of selecting CVs often relies on intuition and prior knowledge, which can lead to variables that do not capture the system's dynamics effectively.

Choosing a CV is similar to selecting a descriptor of the atomic environment. Both processes involve identifying key features that govern the system's behavior. Just as descriptors quantitatively characterize the local environment around atoms, providing a systematic way to capture relevant features, CVs must encapsulate the essential degrees of freedom and transitions. Both should highlight significant movements and changes within the system, ensuring that important dynamics and barriers are not overlooked. Using atomic environment descriptors can thus enhance the selection of CVs, offering a more structured and quantitative approach. In the following section, we will explore how atomic environment descriptors can be employed to achieve a deeper understanding of the system's characteristics and behavior.

2.6 Descriptors of Atomic Environments

Understanding atomic environments is crucial for grasping the physical properties of metallic systems and the fundamental processes that govern them. Processes such as phase transformations and mechanical deformations span multiple scales, influencing both atomic and macroscopic behaviors. Atomistic simulations serve as a bridge between these scales, offering insights into the dynamic evolution of these systems. However, the complexity inherent in these behaviors necessitates advanced analytical techniques to fully capture and comprehend them. In this context, machine learning has significantly enhanced our ability to detect and predict molecular patterns, providing powerful tools that complement traditional simulation methods.

For effective application of machine learning, a detailed representation of atomic and molecular environments is essential. Molecular dynamics simulations typically produce data in the form of tuples $A = \{\mathbf{r}_i, \alpha_i\}$ for $i \in \mathbb{N}$, where $\mathbf{r}_i$ are the Cartesian coordinates in $\mathbb{R}^3$ and α_i represents the identity of the i-th particle. While these coordinates contain all the necessary structural information, they are not sufficient for a complete description of the system due to their non-unique nature and sensitivity to transformations such as permutation, rotation, and translation.

To address this issue, descriptors of atomic environments transform Cartesian coordinates into a unique descriptor space, providing clear fingerprints of atomic neighborhoods. This approach is analogous to selecting collective variables, as described in the previous section, focusing on the most significant degrees of freedom. Traditionally, structural descriptors, also known as order parameters [232], have been used to identify phases in condensed matter and liquid crystals. These descriptors are based on human understanding and prior knowledge of physical systems and are essential for providing insights into local atomic environments, particularly in metallic systems.

2.6.1 Human-Based Descriptors

For metallic systems, notable descriptors include the centrosymmetry parameter (CSP), adaptive common neighbor analysis (a-CNA), and Steinhardt order parameters such as the widely used Q_4, Q_6, and Q_{12} parameters.

The CSP quantifies the degree of local symmetry around an atom, making it particularly useful for detecting defects in crystalline structures such as dislocations and grain boundaries. It is calculated by evaluating the vector sum of the forces experienced by an atom due to its neighbors, which measures how much an atom's local environment deviates from perfect centrosymmetry.

Adaptive common neighbor analysis enhances structural classification accuracy in systems with varying atomic densities by adapting to the local atomic environment. This method extends the common neighbor analysis by classifying atomic environments based on the connectivity and types of bonds between neighboring atoms. The adaptability of a-CNA to specific local densities makes it more versatile and accurate, particularly for complex and heterogeneous systems.

Steinhardt order parameters, introduced by Steinhardt *et al.* in 1983 [232], are crucial for quantifying the local orientational order around an atom. These parameters, including Q_4, Q_6, and Q_{12}, are particularly effective in studying phase transitions and nucleation processes. The indices 4, 6, and 12 correspond to the specific orders of spherical harmonics used in their calculation. For example, the Q_6 parameter employs spherical harmonics of order 6 to measure the degree of local orientational order.

Mathematically, the Steinhardt order parameter Q_l for a given atom is defined as:

$$Q_l = \left(\frac{4\pi}{2l+1} \sum_{m=-l}^{l} \left| \bar{Q}_{lm} \right|^2 \right)^{1/2},$$

where $\bar{Q}_{lm}$ represents the average of the spherical harmonics Y_{lm} calculated over the nearest neighbors of the atom. These parameters are sensitive to variations in the local structural arrangement, enabling the identification of different phases and local structural motifs within a metallic system [233].

However, these descriptors have limitations. They heavily rely on human expertise and detailed prior knowledge of the system, which can introduce biases and limit their generality across different systems and conditions. Additionally, the selection and design of these descriptors are often based on a trial-and-error approach, which may not capture all the relevant features and subtleties of the atomic environments, potentially overlooking important details.

The advent of machine learning has driven the development of new, more sophisticated, data-driven descriptors. These modern descriptors are designed to automatically learn the most relevant features from data without requiring explicit human intervention or detailed prior knowledge of the system. In the next section, we will explore such modern descriptors, focusing on the Smooth Overlap of Atomic Positions (SOAP) [202], which plays a primary role in the research presented in this thesis.

2.6.2 Smooth Overlap of Atomic Positions (SOAP)

The Smooth Overlap of Atomic Positions[202] descriptor offers a sophisticated method to describe atomic environments. A key aspect of the SOAP descriptor is its ability to capture the local order around an atom, describing how the surrounding atoms are spatially arranged. Each unique spatial arrangement corresponds to a distinct subspectrum. Mathematically, SOAP vectors provide a high-dimensional embedding that represents the local order and disorder surrounding each atom within the system. This high-dimensional representation acts as a fingerprint—a data-driven local order parameter that characterizes the atomic environment. The function employed by SOAP captures rotational and translational invariant many-body density correlation features within a defined cutoff sphere.

Originally developed for approximating potential energy surfaces and interatomic potentials, SOAP effectively represents chemical environments. The vectorized SOAP representation for the i-th local environment is constructed by considering the local density, formulated as a sum of Gaussian contributions:

$$\rho_i^{(\alpha)}(r) = \sum_j \exp\left(-\frac{|r - r_{ij}|^2}{2\sigma^2}\right) f_{\text{rcut}}(|r - r_{ij}|) \tag{2.44}$$

Here, f_{rcut} is a cutoff function ensuring the inclusion of particles within a certain radius.

The density function undergoes an expansion using orthonormal radial basis functions and spherical harmonics:

$$\rho_i^{(\alpha)}(r) = \sum_n \sum_l \sum_{m=-l}^{l} c_{nlm}^{\alpha} g_n(r) Y_{lm}(\theta, \phi) \tag{2.45}$$

where c_{nlm}^{α} represents the expansion coefficients. The SOAP representation's uniqueness comes from forming the partial power spectrum of these coefficients:

$$p_{nn'l}^{\alpha\alpha'} = \sqrt{\frac{8\pi^2}{2l+1}} \sum_{m=-l}^{l} c_{nlm}^{\alpha,i} \cdot c_{n'lm}^{\alpha',i} \tag{2.46}$$

In this context, a molecular dynamics snapshot is represented as a collection of these vectors:

$$\Gamma = \{\mathbf{p}_i, \alpha_i\}_{i \in \mathbb{N}} \tag{2.47}$$

The dimensionality of these vectors is determined by the radial and angular basis expansions and the number of atomic species considered. The number of components of the vectors in this representation depends on the extent of the radial and angular basis expansions, and on the number of different atomic species taken into account. The dimensionality is given by the formula:

$$N_{\text{features}} = N_\alpha \cdot n_{\max} \cdot \frac{(N_\alpha \cdot n_{\max} + 1) \cdot (l_{\max} + 1)}{2} \tag{2.48}$$

Here, N_{features} represents the total number of features. The involved parameters, N_α (the number of atomic species), $n_{\max}$ (the maximum radial basis), and $l_{\max}$ (the maximum angular momentum), can be tuned during the analysis. Due to the

quadratic dependence on the number of elements and on the radial basis number, the number of features reaches a very high value. Therefore, the SOAP vectors associated with an MD trajectory constitute a substantially high-dimensional dataset, which can be interpreted only after further processing, such as dimensionality reduction techniques, as shown later in our works.

2.6.3 Similarity between Atomic Environments

In the analysis of molecular structures and atomic environments, it is essential to quantify the similarity between different configurations to facilitate the classification and comparison of materials. Descriptors, such as SOAP vectors, provide a way to represent complex atomic environments in a high-dimensional feature space. To compare these environments effectively, we employ similarity measures that are capable of handling the high-dimensional and nonlinear nature of the data.

The similarity between SOAP vectors, denoted as $\mathbf{p}_i \in \mathbb{R}^D$, derived from molecular dynamics simulations, is evaluated using a kernel function, $\kappa(\cdot,\cdot)$. This function takes two vectors as input and provides a measure of their similarity, which is bounded and typically calculated as the dot product of the two vectors:

$$\kappa(a,b) = a \cdot b, \quad \forall a,b \in \mathbb{R}^D. \tag{2.49}$$

From this, the kernel distance can be derived, providing a way to quantify the difference between two points in the feature space:

$$D_\kappa(a,b) = \sqrt{\kappa(a,a) + \kappa(b,b) - 2\kappa(a,b)} = \sqrt{2(1-\kappa(a,b))}. \tag{2.50}$$

In simpler terms, the kernel distance between two points in the feature space is determined by their self-similarities and mutual similarity, $\kappa(a,b)$. This approach is fundamental for comparing various SOAP vectors, as will be demonstrated in the results discussed in Chapters 3, 4, and 5 of this thesis.

2.6.4 The Local Environments and Neighbors Shuffling (LENS) Descriptor

Understanding the dynamic behavior of particles in a molecular system requires more than just static structural descriptors. The Local Environments and Neighbors Shuffling (LENS) descriptor provides a time-dependent method for monitoring changes in the local environments of particles over time, offering deeper insights into system dynamics that traditional methods may overlook.

LENS tracks how the local environment of a particle evolves throughout its trajectory by monitoring changes in the identities of its neighboring particles. Specifically, it measures the extent to which neighbors are added, removed, or reshuffled between consecutive time intervals. The LENS descriptor, $\delta_i(t+\Delta t)$, for a particle i at time $t+\Delta t$ is defined as:

$$\delta_i^{t+\Delta t} = \frac{\left| C_i^t \cup C_i^{t+\Delta t} - C_i^t \cap C_i^{t+\Delta t} \right|}{\left| C_i^t + C_i^{t+\Delta t} \right|} \tag{2.51}$$

where C_i^t represents the set of neighbor identities (IDs) around particle i at time t, and $\Delta C_i^{t+\Delta t}$ denotes the symmetric difference between the neighbor sets at times t and $t+\Delta t$.

LENS has proven to be particularly effective in detecting rare local fluctuations that are often missed by average-based analyses. This capability is highlighted in the study of various metallic systems, where LENS identified dynamic regions coexisting with more static areas, illustrating how local fluctuations can drive larger collective rearrangements.

To apply LENS effectively, a sufficiently sampled trajectory dataset is required, containing detailed sequential information on the relative positions and identities of neighboring particles across different time intervals. This time-series nature allows LENS to monitor changes in the local environment of each particle over time, making it possible to detect both common and rare dynamic events.

The abstract and flexible nature of LENS makes it well-suited for analyzing a wide range of complex systems beyond molecular structures. Its versatility has been demonstrated in various test cases, from soft materials to crystalline molecular and

atomic systems, and across different molecular dynamics simulation models and force fields.

The application of LENS has enriched the analysis presented in this thesis by revealing dynamic behaviors and structural rearrangements of single or few atoms, phenomena typically missed by standard descriptors like SOAP. These insights are discussed in detail in Chapters 5 and 6.

2.6.5 Combining LENS and SOAP descriptors

The macroscopic properties of complex systems often depend on microscopic dynamic events, which are challenging to detect due to their fleeting and intricate nature. To address this, combining the structural insights from SOAP with the dynamic sensitivity of LENS offers a comprehensive analysis of molecular systems.

Starting, for example, at time t_1, a SOAP spectrum is computed for each particle i in the system. We also calculate its LENS value for the immediately subsequent time interval Δt. By including the LENS term as an extra component into each SOAP power spectrum, we thus obtain a new vector $\chi_i^{t_1}$ containing information on the structural properties in the neighboring environment surrounding atom i at time t_1 and its evolution in the subsequent time interval $t_1 + \Delta t$.

The combination is expressed as:

$$\chi_i^{t_1} = (p_i^{t_1}, \delta_i^{t_1+\Delta t})$$

The SOAP spectrum and LENS scalar component are normalized to have equal weight in the dataset. Despite their different forms—a high-dimensional vector for the SOAP spectrum and a scalar for LENS—this normalization ensures they have the same 'statistical weight' within the dataset. By applying this procedure across the entire trajectory, we generate a new dataset (SOAP&LENS dataset) containing N vectors, each with a dimension of $n+1$, where n represents the SOAP spectrum dimension (structural information) and 1 corresponds to the LENS (dynamic) component. This updated dataset effectively captures information on the instantaneous environments surrounding each particle i and their propensity to change over time at the analysis's resolution Δt. The decision to combine the structural insights of SOAP and the dynamic sensitivity of LENS into the comprehensive SOAP&LENS descriptor was informed by the distinct strengths observed from each individual

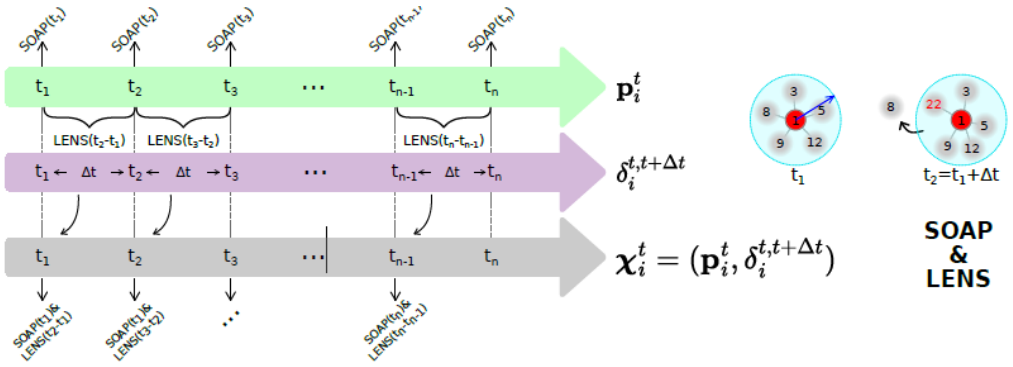


Fig. 2.7 **Scheme of the SOAP&LENS combined dataset.** The SOAP power spectrum of each particle at every time step (p_i^t) is combined with the LENS scalar value calculated at the subsequent time interval ($\delta_i^{t+\Delta t}$), obtaining a new dataset $\chi_i^t = (p_i^t, \delta_i^{t+\Delta t})$. *Adapted with permission from* [234]

descriptor. The SOAP descriptor captures rich structural information by representing the local atomic environment with high precision, while the LENS descriptor highlights the dynamic changes in particle neighborhoods over time, providing a clear picture of the microscopic events that drive macroscopic behaviors. This combined approach allows for the detection of rare but significant dynamic events that might be missed using only structural descriptors. It also improves classification accuracy by distinguishing between different dynamical behaviors within structurally similar environments. Consequently, the SOAP&LENS descriptor offers a more detailed and accurate characterization of molecular systems, capturing the intricate interplay between structure and dynamics that governs their macroscopic properties.

In Chapter 6, we demonstrate the application of this new descriptor on a metallic system, such as Cu(211) [234], revealing detailed structure-dynamics relationships that improve our predictive understanding of these intricate systems.

Chapter 3

Innate dynamics and identity crisis of a Copper surface

Despite their crystalline order, which typically suggests rigidity and structural regularity, metals experience significant structural changes at temperatures much lower than their melting point[75, 76, 200]. These changes are the results of the complex dynamic behavior within the crystalline lattice that profoundly impacts material properties. Recognizing and understanding such structural dynamics is essential, as they play a crucial role in determining the functionality and performance of metallic surfaces in various applications, from catalysis to the design of functional materials. However, traditional studies often treat metal surfaces as static[164–170], overlooking the complex atomic interactions that significantly influence material functionality.[235–239, 185] This chapter will demonstrate how advanced computational techniques and ML are essential tools for comprehending these dynamic behaviors at the atomic scale, providing deeper insights into the intricate nature of metallic surfaces. Building on these considerations, this chapter focuses on the dynamics of copper surfaces at different temperatures, providing new insights into their behavior using deep-potential molecular dynamics simulations (DPMD), previously introduced in Chapter 2. By training a deep neural network potential on DFT data, we obtained a dynamically accurate force field for Cu surfaces. This allows us to simulate surfaces composed by more than 2000 atoms for 150 ns, scales on which interesting dynamical events occur on these surfaces, even at temperatures in which they are assumed to be solid. Then utilizing high-dimensional structural descriptors and unsupervised machine learning, we could classify and track all atomic environ-

ments that populate these surfaces in relevant conditions. This method reveals the transient nature of atomic environments that compose such Cu surfaces, where non-native atomic states continuously appear and disappear in dynamic equilibrium with native ones. This also allows us to introduce the concept of a "*statistical identity*" of metal surfaces, a concept that we believe will help to rationalize their behavior and properties under relevant conditions. Describing the findings reported in the paper "*Innate Dynamics and Identity Crisis of a Metal Surface Unveiled by Machine Learning of Atomic Environments*"[172], this discussion elucidates the complex behavior of different copper slabs that transform yet partially retain their identity under specific conditions. The insights gained here are of great theoretical significance for the understanding of metallic dynamics and have practical applications where surface properties are critical. The following section will delve deeper into the methodology and findings of this study, showcasing how these advanced techniques provide a comprehensive understanding of the dynamic behavior of copper surfaces at the atomic level.

Full bibliographic reference: **Cioni, M**., Polino, D., Rapetti, D., Pesce, L., Delle Piane, M., & Pavan, G. M. (2023). "Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments." *Journal of Chemical Physics*, 158(12), 124701.[1]

3.1 Introduction

As highlighted in Chapter 1, metallic surfaces present an intriguing paradox in materials science. Despite their crystalline order, which typically suggests rigidity and structural regularity, they may exhibit dynamic properties, even below their melting point, more commonly associated with non-crystalline states. This has significant practical implications, ranging from catalysis, sensing, nanotechnology to the design of functional materials. As temperatures approach the Hüttig temperature—about one-third of their melting point—metallic surfaces can undergo transformative behaviors such as "deconstruction," "preroughening," faceting, roughening, and surface premelting[240–243]. While these transformations indicate that

[1] My contribution to this article, as the first author, involved the development of the deep neural network potential trained on density functional theory calculations. I conducted all the atomistic simulations and analyses, contributed to the interpretation of the results, and contributed to the writing of the manuscript.

dynamics is present on these surfaces well below the melting temperature, these complex structural transformations, observed experimentally and theoretically, are not yet fully understood.

To fully grasp the behavior of metallic systems, it is essential to move beyond their static properties and examine in detail the dynamics that emerge within them in response to thermal and mechanical stimuli. Recent advances in computational modeling and simulation have enabled the dynamic representation of these systems, which is crucial for accurately describing their behaviors at the atomic scale[244–246, 235–239, 185, 159].

The dynamic behavior of metallic surfaces is crucial because it directly influences their properties and performance in various applications. For instance, Gazzarrini *et al.* demonstrated how atom mobility in copper nanoparticles can produce variations in the number of vertex, edge, and face atoms, affecting the nanoparticles' efficiency in catalyzing CO_2 conversion to methane[161]. Similarly, Nelli et al. studied the dynamic diffusion of atomic impurities in copper-cobalt icosahedral nanoparticles via metadynamics simulations.[247]

These studies provide preliminary evidence that metal lattices are complex systems where atoms are in dynamic equilibrium and continuous exchange, similar to what has been observed in soft self-assembling systems.[204, 205, 248, 249]

The appearance and disappearance of specific atomic environments and their dynamic evolution on a metal surface are crucial in determining properties directly related to the structural features of the surface[250, 251] In this context, machine learning plays a primary role in disentangling complex structural dynamics. ML potentials trained on quantum mechanical data provide accurate force fields that enable the simulation of metals on relevant spatiotemporal scales, as discussed in the second chapter.

Since the pioneering work of Behler and Parrinello introduced high-dimensional neural network potentials[186], various approaches have been developed over the past decade,[190, 252–254, 179, 255, 180, 256, 257] allowing for investigations with ab initio accuracy in increasingly complex systems.

For example, unsupervised clustering of SOAP data extracted from molecular dynamics simulations has recently allowed the reconstruction of the structural and dynamical complexity of a variety of molecular materials and systems,[202, 258, 259, 204, 260, 248] leading to the development of robust data-driven metrics for

their classification.[261, 259, 248] Similar data-driven approaches have also been used to explore the mechanism of gold nanoparticle melting.[262]

Building on these considerations, we will study four slabs of copper with different surface orientations: (111), (110), (211), and (210), each comprising 2400 atoms (2304 for the (210) surface). These slabs will be simulated at three temperatures: 500 K, 600 K, and 700 K employing a machine learning potential trained on DFT calculations, enabling us to capture the dynamic atomic interactions accurately.

The analysis of these simulations will be conducted using two complementary approaches, which we define as *bottom-up* and *top-down* analysis. The *bottom-up* approach involves unsupervised clustering of SOAP data extracted from molecular dynamics simulations to reconstruct the structural and dynamical complexity of the systems. The *top-down* approach uses a pre-constructed dictionary of SOAP atomic environments to classify and track the dynamic evolution of these environments on the metal surfaces. Through these methods, we aim to provide a comprehensive understanding of the dynamic behaviors of metallic surfaces and their implications for material properties and functionalities, as detailed in the subsequent sections.

3.2 Results

3.2.1 Modeling the dynamics of large Cu surfaces using a NN potential

In the present work, we use Cu as an example of a metal widely used for various applications, that express a non-trivial surface dynamics even well below the melting temperatures (see Results below)[263–267] 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),[268, 269] where the structural/dynamical features of the Cu surface are important.

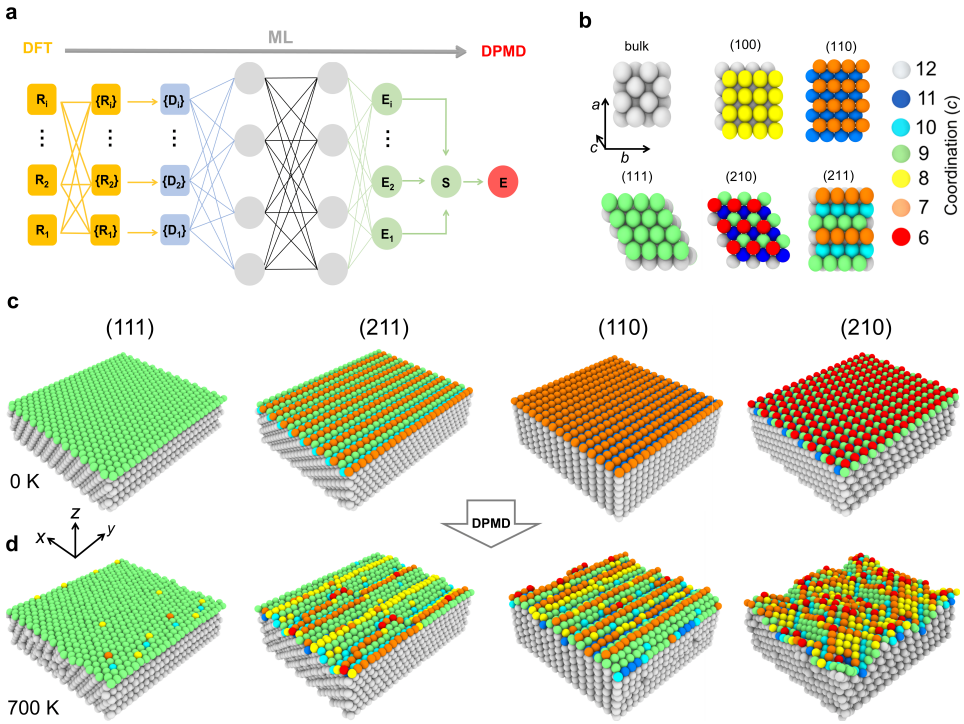


Fig. 3.1 **Atomistic DPMD simulations of Cu surfaces near the Hüttig temperature. a** Scheme of the DeepMD[180] model developed in this work: Atomic configurations and energies obtained from DFT calculations have been used to train an inter-atomic *NN*-potential for the Cu surfaces. **b** Atomistic models 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 range of temperature 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). *Reproduced with permission from Ref* [172].

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[132] calculations are limited to timescales and sizes that are too short/small to this end, whereas classical atomistic forcefields[141, 227, 270–272, 225] 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 (see Methods section for more details) adopting the Deep Potential Molecular Dynamics (DeePMD) scheme developed by Zhang *et al.*[273, 180] (Figure 3.1a), following the procedure described in the dedicated section 2.4 of Chapter 2

An extensive campaign of AIMD simulations of small Cu FCC bulk, (100), (110), (111), (211), and (210) surface environments conducted at temperatures between 500 and 700 K (Figure 3.1b) provided a rich dataset of atomic configurations, forces and energies used to train a first guess *NN*-potential (see first subsection of *Methods* section for more details) An active (iterative) learning strategy[274, 275] 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 DFT level and that are used for the training (Figure 3.1b: ~ 100 atoms, depending on the surface type – sufficient to account up to the $\sim 4-5$ atom neighbors) are smaller than the surfaces that are then simulated at atomistic level (Figures 3.1c,d: containing ~ 2400 atoms), the iterative nature of the approach and our tests demonstrate that the obtained potential is robust and reliable. 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* deep-potential molecular dynamics (DPMD) simulations using such incomplete *NN*-potential are then re-simulated at DFT-level and added to the dataset. Such a process is conducted iteratively, and at each iteration, the 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).

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 of Figure 3.1b (~ 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 the 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 data-set 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 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. The *NN*-potential was finally validated against some copper bulk and surface properties, as reported in the following section.

3.2.2 Details about the *NN*-potential validation

To validate the *NN*-potential, we calculated the lattice constant, the vacancy and interstitial formation energies, and the surface energies at 0 K for the unreconstructed surfaces. The computed values, reported in Table 3.1, are in agreement with those computed at the DFT level, with the embedded atom model (EAM) of Mendelev *et al.*[225], and with available experimental values.[276–278]

First, the lattice parameter was calculated by optimizing the atomic positions and cell dimensions of the supercell. The vacancy formation energy was calculated by removing a copper atom from the bulk and allowing the remaining atoms to relax, using the formula:

$$E_{\mathrm{f}} = E_{\mathrm{vac}} - \left(\frac{N-1}{N}\right) E_{\mathrm{bulk}} \tag{3.1}$$

where E_{f} is the vacancy formation energy, E_{bulk} is the energy of the bulk, N is the number of atoms in the bulk, and E_{vac} is the energy of the slab with one copper atom removed.

The interstitial formation energy (IFE) was calculated by first relaxing the crystal cell structure to obtain the initial energy E_0. An atom was then inserted into the relaxed structure, and the system was relaxed again to obtain the final energy E_{f}^{i}. The IFE was calculated using the formula:

$$E_{\mathrm{f}}^{i} = E_{\mathrm{f}} - \left(\frac{N+1}{N}\right) \times E_0 \tag{3.2}$$

where E_f^i is the interstitial formation energy, E_f is the total energy of the system with the interstitial atom, E_0 is the total energy of the perfect crystal, and N is the number of atoms in the unit cell.

Surface energies for each slab, including the (110) missing-row (1×2) reconstructed surface, were calculated using the formula:

$$\gamma = \frac{E_{\text{slab}} - N_{\text{slab}} E_{\text{bulk}}}{2A} \tag{3.3}$$

where γ is the surface energy, E_{slab} is the total energy of the slab, E_{bulk} is the total energy of bulk copper per atom, N_{slab} is the number of atoms in the slab, and A is the surface area of the slab.

Additionally, 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 finding aligns with experimental evidence showing that the Cu (110) surface at low temperatures does not undergo the (1×2) missing-row reconstruction typical of other noble metals, such as Au and Ag.

Lastly, the diffusion barriers of single adatoms on the (100), (110), and (111) surfaces were computed. The results, reported in Table 3.2, are in good agreement with available theoretical and experimental data.[279–282, 224, 283–287] This test validates the potential's ability to reproduce the dynamic properties of the surface atoms accurately.

	NN-potential	DFT	EAM	Expt.
lattice parameter (Å)	3.626	3.630	3.639	3.615
Vacancy formation (eV)	1.14	1.10	1.07	1.27
Interstitial formation (eV)	3.28	3.49	3.87	2.8-4.2
Surface energies (eV Å^{-2})				
(111)	0.99	1.07	1.04	1.79[a]
(100)	1.10	1.19	1.21	1.79[a]
(110)	1.29	1.36	1.31	1.79[a]
(110)(1×2)	1.30	1.38	1.30	1.79[a]
(211)	1.21	1.28	1.24	1.79[a]
(210)	1.36	1.50	1.39	1.79[a]

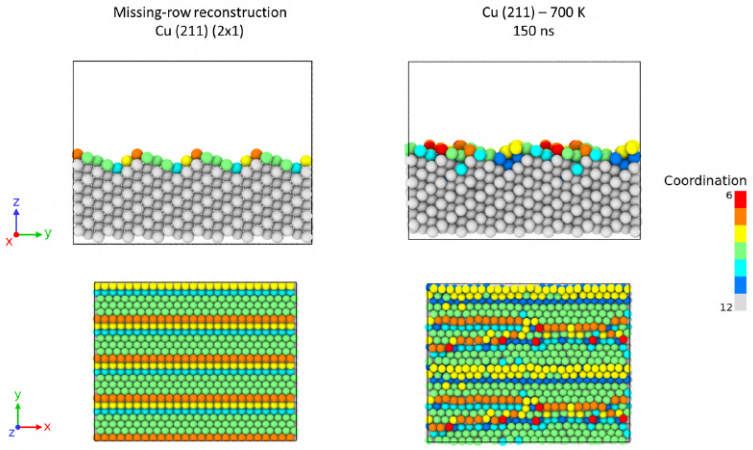
Table 3.1 Lattice parameter, vacancy and interstitial formation energies, and surface energies computed with our trained *NN*-potential, DFT and the embedded atom model (EAM) of Mendelev *et al.*[225] and compared with available experimental data[276–278]. [a]For average orientation.

These values were derived by sampling approximately 80 diffusion events at 500, 600, and 700 K. The diffusion times were fitted to a Poisson distribution, providing a characteristic time for each event. The inverse of these times was then fitted to an Arrhenius plot to calculate the diffusion energy barriers.

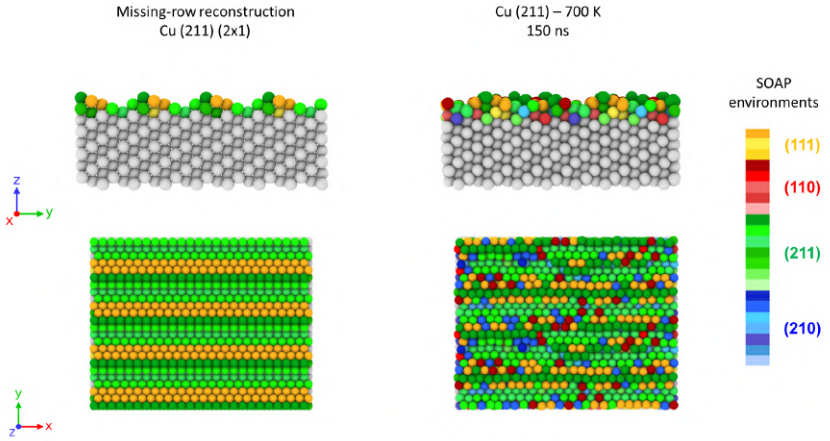
Surface	NN-potential	Theory	Expt.
(100)	0.43	0.39-0.53	0.28-0.40
(110) in channel	0.28	0.23-0.53	
(110) cross channel	0.31	0.26-0.49	
(111)	0.11	0.1	0.1-0.15

Table 3.2 Adatom diffusion energy barriers (eV) on Cu (100), (110), and (111) computed with our trained *NN*-potential and compared with available theoretical and experimental literature data.[279–282, 224, 283–287]

Figures 3.2 and 3.3 illustrate the structural and environmental changes in Cu (211) and Cu (110) surfaces with missing-row reconstructions. The figures show side and top views of the surfaces initially and after 150 ns at 700 K, highlighting significant atomic rearrangements due to thermal effects. Coordination number and SOAP environment coloring emphasize changes in atomic coordination and local atomic environments, showing the emergence of new atomic configurations and structural transformations.

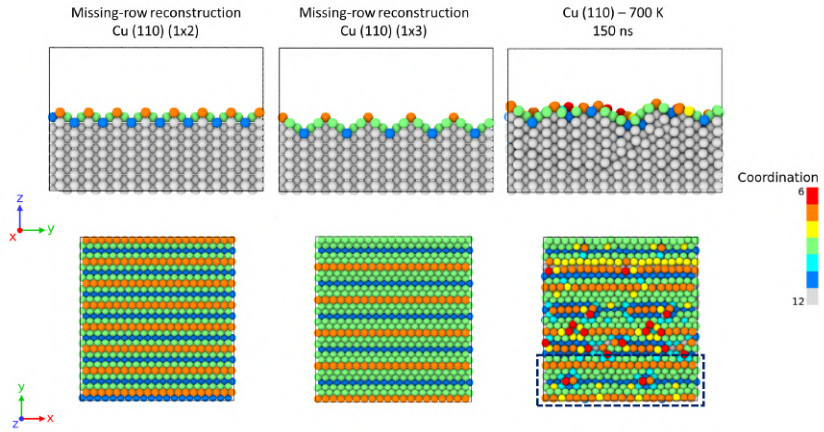


(a) Side and top views of the Cu (211) surface with missing-row reconstruction (2x1), colored by coordination number. Left: initial structure, Right: after 150 ns at 700 K.

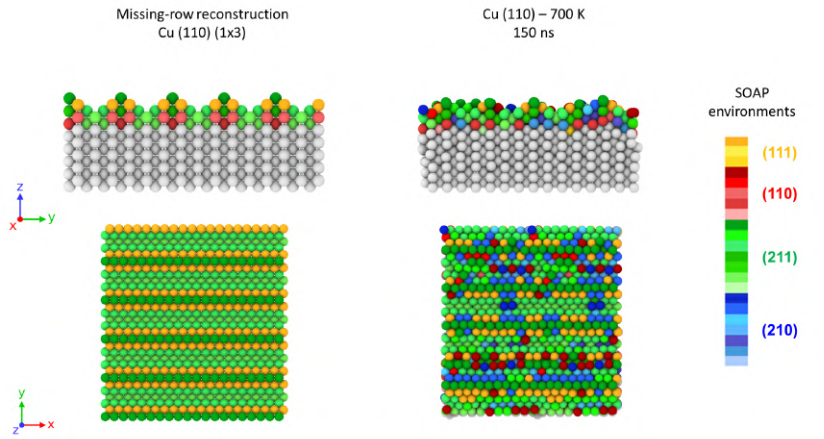


(b) Side and top views of the Cu (211) surface with missing-row reconstruction (2x1), colored by SOAP environments. Left: initial structure, Right: after 150 ns at 700 K.

Fig. 3.2 Structural and environmental changes in Cu(211) surface with missing-row reconstruction (2x1). *Reproduced with permission from Ref* [172].



(a) Side and top views of the Cu (110) surface with missing-row reconstructions (1x2 and 1x3), colored by coordination number. Left and middle: initial structures, Right: after 150 ns at 700 K.



(b) Side and top views of the Cu (110) surface with missing-row reconstruction (1x3), colored by SOAP environments. Left: initial structure, Right: after 150 ns at 700 K.

Fig. 3.3 Structural and environmental changes in Cu (110) surface with missing-row reconstructions (1x2 and 1x3). *Reproduced with permission from Ref* [172].

Figure 3.4 shows the radial distribution functions (RDF) of surface atoms at different temperatures for Cu(111), Cu(110), Cu(211), and Cu(210) surfaces. The RDF peaks are sharper and more defined at 0 K, indicating well-ordered atomic arrangements. As temperature increases, peaks broaden and decrease in height, reflecting increased atomic vibrations and disorder. Cu(111) shows the most distinct

peaks, suggesting higher stability compared to other surfaces, aligning with the known stability of the (111) surface in FCC metals like copper.

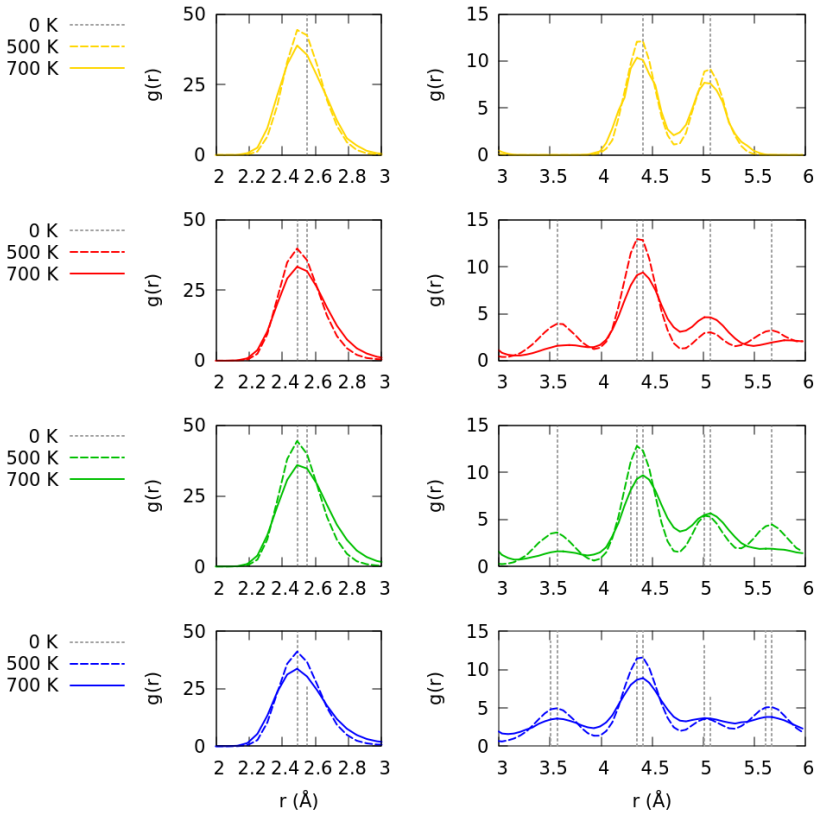


Fig. 3.4 Radial distribution functions of surface atoms at short (left) and longer distances (right) calculated at 0 K (with grey short-dashed impulses), at 500 (dashed lines) and 700 (solid lines) K for Cu(111) (yellow), Cu(110) (red), Cu(211) (green) and Cu(210) (blue). *Reproduced with permission from Ref* [172].

3.2.3 Investigating Cu surface dynamics and structures with DPMD simulations

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* deep potential molecular dynamics (DPMD) simulations. All Cu surface models have depth > 15Å and replicate on the *xy* plane through periodic boundary conditions, effectively modeling a portion of the bulk of infinite surfaces (Figure 3.1c). During the DPMD simulations, the 2 bottom layers in the surface models are kept fixed, while all other

atoms are free to move. Such setup and system sizes prevent finite-size effects and guarantee reliable modeling of the structural dynamics of these Cu surfaces.

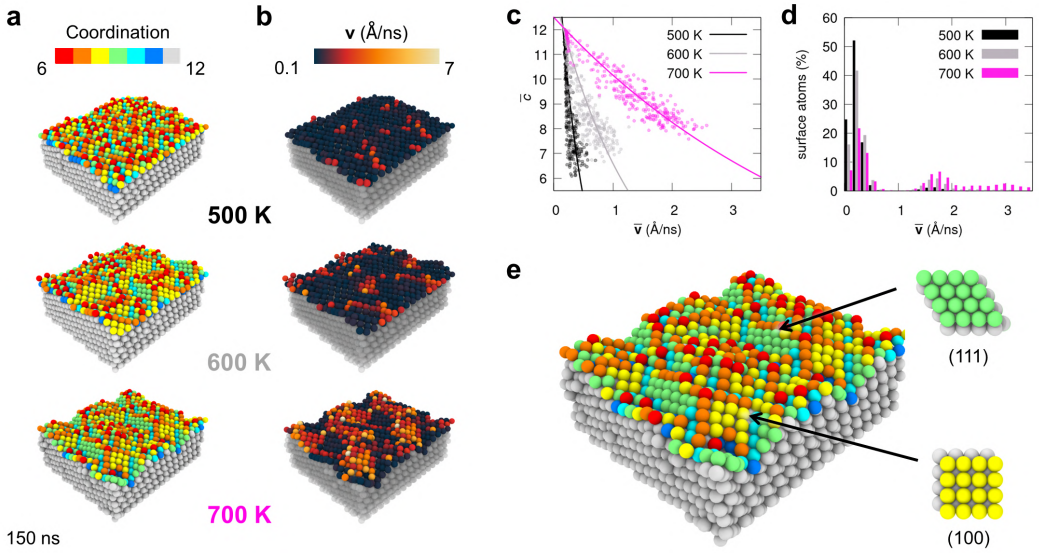


Fig. 3.5 **Dynamic diversity in a metal surface.** **a** The Cu(210) surface after 150 ns of DPMD simulations at 500, 600 and 700 K of temperature (top-to-down). Atoms are colored based on their coordination. **b** Relative diffusion velocity in the (210) surface at the different temperatures: the Cu 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 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). *Reproduced with permission from Ref* [172]

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. Changes in the atoms' colors in Figure 3.1d 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 surface energies, and more importantly with experimental evidences. 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.[288, 289]

For what concerns the (110) surface, it has been demonstrated that this orientation is characterized by an increasing disorder induced by thermal energy already at T > 550 K[290–292]. 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"[290]. Tosatti and coworkers attributed this phenomenon to a missing-row type reconstruction, given the small difference between the surface energies of the reconstructed and unreconstructed surfaces[293, 240]. Our simulations confirm this interpretation as can be observed in Figure 3.3. 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. It is immediate to notice the formation of a (1×3) missing-row type reconstruction in the section highlighted by the dashed box.

More recently, also Cu(211) has been the object of several studies, given its reactivity toward the methanol synthesis.[268]. Witte *et al.* have found that upon oxygen adsorption already below room temperature also the (211) surface undergoes reconstruction forming a (211)(2×1) surface[294]. The same type of reconstruction is observed in our simulations at 600 and 700 K. In figure 3.2 we compare the ideal (211)(2x1) reconstructed surface with a snapshot of the (211) surface after 150 ns at 700 K. As for the (110), also in this case it easily detectable a reconstruction of the surface toward the (211)(2×1) surface.

From a dynamic perspective, the (210) surface is found dynamic at all temperatures and its dynamics increases with temperature (Figure 3.5a). 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 simulations, 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 in Figure 3.5b move by tens of nanometers during the DPMD simulations, while black atoms just vibrate around their lattice position.

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 of Figure 3.5b 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 time-scales, and are also in continuous dynamic exchange with each other as it will be characterized in detail in Figures 3.6, 3.8, and 3.9. The plots of Figure 3.5c 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 (Figure 3.5d) 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 3.5a 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 which corresponds to (111) facets (in green) and (100) facets (in yellow). A zoom onto the (210) surface at 700 K (Figure 3.5e) 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 note that similar results have been observed in the experimental work of Kirby *et al*.[295] In particular, the authors 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 Figures 3.5e *vs.* 3.5b, which provides an insight into the dynamic diversity of the surface, revealing how non-native (111) domains (Figure 3.5e: green) are more static while (100) ones (in yellow) correspond to more dynamic regions in Figure 3.5b. Noteworthy, along the DPMD simulations the atoms change color dynamically in these Cu surfaces, demonstrating the dynamic equilibrium present within them.

3.2.4 Unsupervised machine learning of microscopic 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 struc-

tural/dynamical complexity of various types of molecular systems.[248, 260, 204, 259, 261] This descriptor was introduced and explained in Chapter 2 of this thesis (for detailed information, please refer to the dedicated section 2.6). We use SOAP vectors as high-dimensional descriptors of the local environments surrounding each atom on 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.[248, 260, 261, 204] Key advantages of such analysis are that the SOAP descriptor is agnostic[202] and the analysis is unsupervised 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).[248, 260, 261, 204] We perform our SOAP analysis on the DPMD trajectories of the various simulated Cu surfaces[172]. For all systems, the SOAP descriptors were calculated under periodic boundary conditions along the *xy* dimensions using nmax = 8 radial basis function and lmax = 8 maximum number of spherical harmonics. The choice of the cut-off radius (rcut) determines the size and the shape of the neighborhood considered in characterizing the atomic environment for each SOAP center; in this work, we opted for rcut= 6.0 Å, the same cutoff adopted for the training of the Cu *NN*-potential. After 75 ns of DPMD, all surfaces reach a microscopic equilibrium where the atomic environments populating the systems do not change anymore [172] The last 75 ns of DPMD, representing 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 topmost 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. Unsupervised clustering of the SOAP data *via* the Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN*) algorithm [296, 297] identifies the main SOAP clusters (states/environments) populating the equilibrium DPMD trajectories (Figure 3.6a,b).

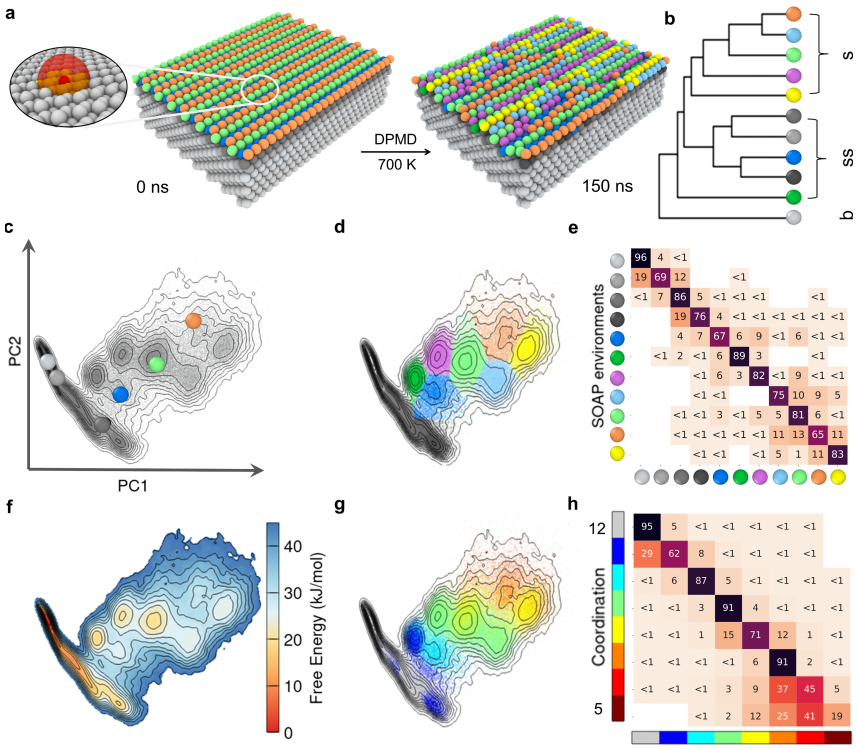


Fig. 3.6 **ML of atomic environments in the Cu(211) surface and of their dynamics. a** Cu atoms on the (211) surface colored based on the SOAP environments emerging along the equilibrium (last 75 ns) DPMD simulation at 700 K. **b** Hierarchical dendrogram connecting the 11 detected SOAP clusters, corresponding to 6 internal atomic environments – 5 sub-surface (ss) and 1 bulk (b) – plus 5 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** Unsupervised 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 a 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). *Reproduced with permission from* Ref [172].

As an example, HDBSCAN* identifies 11 main SOAP clusters in the Cu(211) surface at 700 K (Figure 3.6a,b). These correspond to the density peaks seen in the PCA of the SOAP data in Figure 3.6c,d, namely, to the most "visited" atomic environments during the DPMD simulation. The dendrogram of Figure 3.6b 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 Figures 3.6a,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 Figure 3.6c indicate which SOAP states are present in the ideal (211) surface at 0 K (at DPMD start). Comparing Figure 3.6c and Figure 3.6d 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 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 Figure 3.6e 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 [172]. In general, the higher are the numbers on the diagonal of the matrix, the more persistent is the surface. *Vice versa*, the higher are the off-diagonal probabilities, the more probable are the atomic transition between the SOAP states in Δt and the more dynamic is thus the surface. The fact that the transition matrix of Figure 3.6e 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 $\sim 6\%$ 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.[298–300] All

entries $< 1\%$ should be taken as qualitative, as they refer to events observed only sparsely along the DPMD.

This analysis shows that, in the timescales that can be exlpored in this simulation, 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 nanoseconds timescale. This demonstrates how the (211) surface has a "discrete" dynamics.

Such a "discrete" dynamics is observed also for the (110) surface. Also here the atoms of the first layers move via jumps across different crystallographic positions. This type of motion is correlated to 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 of the peak intensities and a more diffuse background, as observed in Figure 3.4. This type of atomic motion could explain the anomalous reduction of the intensity peaks of Cu (110) at T>550 K obtained by diffraction experiments.[241, 240, 243] 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. This behavior is also mirrored by a quasi-liquid like g(r) at 700 K (see Figure 3.4). On the other hand, (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 Free Energy Surface profile (FES) for the Cu(211) surface model at 700 K (Figure 3.6f). Such FES shows how at 700 K all surface SOAP states are separated from each other by relatively low free energy barriers within $\sim 10-15$ 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[172]. The SOAP analysis of Figures 3.6d-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 3.6g 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 coordination number between the atomic environments in the Cu surface, while at the same time a simple coordination analysis is less sensitive – see, *e.g.*, green domains in the PCA having the same coordination number (~ 9) but corresponding to different SOAP density peaks (Figure 3.6g).

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.

3.2.5 A dictionary of SOAP atomic environments

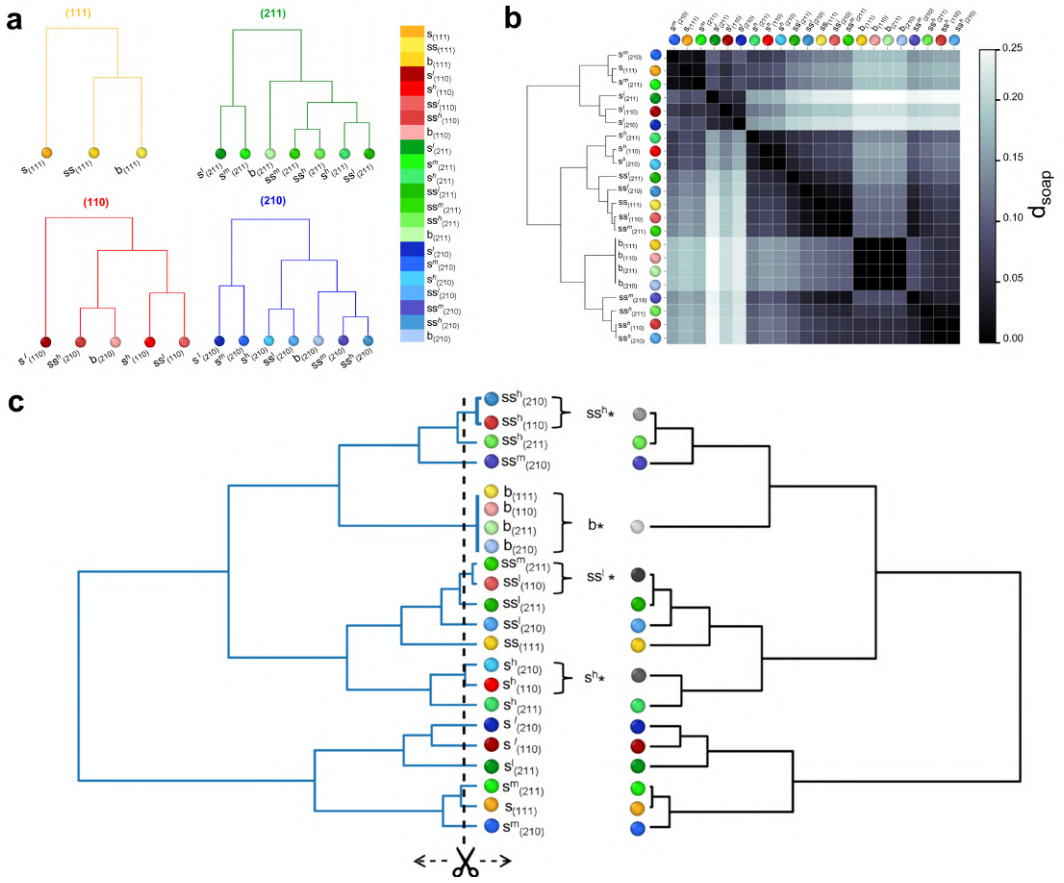


Fig. 3.7 **A SOAP dictionary for classifying atomic environments in Cu surfaces. a** Hierarchical dendrograms for the SOAP environments detected in the different ideal Cu surfaces (at 0 K): Cu(111) environments are shown in yellow, (211) in green, (110) in red, and (210) in blue. **b** Distance matrix: 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*), surface (s*). *Reproduced with permission from* Ref [172]

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 3 SOAP atomic environments in the ideal (111) surface – Figure 3.7a (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 5 main SOAP atomic environments (in red), while the ideal (210) and (211) are characterized by 7 SOAP environments (Figure 3.7a: 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 data-set 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.[248, 261, 259] 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. 2.46 of Chapter 2. The SOAP distance between two SOAP spectra $\vec{a}$ and $\vec{b}$ can be calculated by:

$$d_{SOAP}\left(\vec{a},\vec{b}\right) = \sqrt{2 - 2\mathscr{K}\left(\vec{a},\vec{b}\right)} \tag{3.4}$$

where, with the SOAP power spectrum representation that we are using,

$$\mathscr{K}\left(\vec{a},\vec{b}\right) = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\left\|\vec{b}\right\|} \tag{3.5}$$

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 Figure 3.7b. The colors of the matrix cells represent the SOAP distance (d_{soap}) between the various SOAP atomic environments: dark colors indicate very similar environments ($d_{soap} \sim 0$), light colors identify structurally different atomic environments. The dendrogram adjacent to the matrix (Figure 3.7b: 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, some high-coordination sub-surface ones (ss), etc. 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.*, $\mathrm{s}_{(111)}$ *vs.* $\mathrm{s}^{\mathrm{m}}_{(211)}$ and $\mathrm{s}^{\mathrm{m}}_{(210)}$, or $\mathrm{s}^{\mathrm{l}}_{(110)}$ *vs.* $\mathrm{s}^{\mathrm{l}}_{(210)}$.

Figure 3.7c 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, Figure 3.7c (right) shows what the dendrogram becomes when considering only $d_{soap} \geq 0.01$ (in black), *i.e.* when considering different AEs only above a certain d_{SOAP} 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 the next section, 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 condition and those included in the SOAP dictionary (*top-down* classification).

3.2.6 Atomic dynamic, structural reconstructions and statistical equivalent identities of Cu surfaces

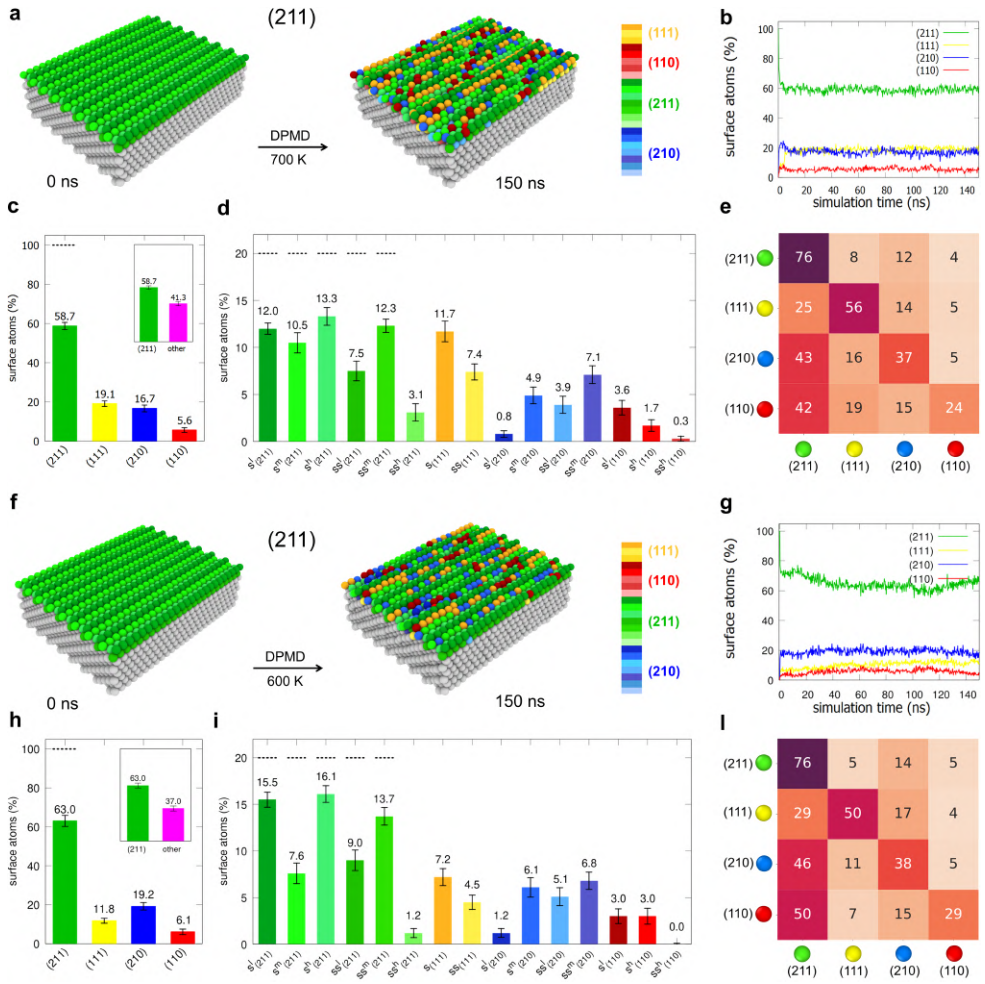


Fig. 3.8 **Dynamic reconstructions and equivalent identity of a Cu(211) surface. a** Cu(211) at 0 ns (left) and after 150 ns of DPMD at 700 K (right): SOAP environments native of the ideal (211) surface are colored in green. Red, blue and yellow colors identify non-native atomic domains, proper of (110), (210) and (111) surfaces. **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, red, combined in pink in the inset). **d** Breakdown of (211) composition at 700 K. Dashed lines indicate the composition at DPMD start. **e** Transition matrix showing the probabilities for atomic transitions in (211) between native and non-native environments at 700 K (within $\Delta t = 300$ ps). **f-l** Same analyses for the (211) surface at 600 K. *Reproduced with permission from* Ref [172]

Starting from the Cu(211) surface at 700 K, at each timestep along the DPMD simulation we measure the d_{soap} distance between the SOAP spectrum of each atom and all SOAP spectra characteristic environments present in the dictionary of Figure 3.7. At each DPMD timestep, 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 microscopic surface reconstructions and atomics dynamics, in this phase we focus only on the five top-most layers of the simulated surfaces. Figure 3.8a shows how the (211), ideally composed only of green native SOAP sites at 0 K (left), convert into local non-native domains, proper of (111), (110) and (210) ideal surfaces (right: yellow, red and blue respectively), along 150 ns of DPMD at 700 K. The environment populations of Figure 3.8b show that the surface reaches a microscopic equilibrium along the DPMD, being populated of native and non-native domains. Along the DPMD, $\sim 40\%$ of native domains disappear converting into non-native domains (Figures 3.8b,c). On a statistical level, the (211) surface thus preserves its own identity only by $\sim 60\%$ at 700 K (Figure 3.8c: inset). In particular, $\sim 19\%$ of the emerging domains correspond to (111) environments, $\sim 17\%$ to (210), and $\sim 6\%$ to (110) ones (Figure 3.8c). Such surface reconfiguration is rather fast in this system (Figure 3.8b). Figure 3.8d 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, $\mathrm{s}_{(111)}$ is the predominant one ($\sim 12\%$), followed by $\mathrm{ss}_{(111)}$ and $\mathrm{ss}^{\mathrm{m}}_{(210)}$, both constituting $\sim 7\%$ of the (211) surface at 700 K. While such analysis provides detailed quantitative information on the composition, structural diversity, and reconfiguration in the metal surface, it is interesting to interrogate on the dynamical features of such phenomena.

The transition matrix of Figure 3.8e 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 $\sim 42-43\%$) than to remain there in $\Delta t = 300$ ps. This provides a picture reminiscent to what seen in Figure 3.5 for the (210) surface, where above the Hüttig temperature such dynamically diverse surface appears as composed of stable domains coexisting in equilibrium with more dynamic ones [172].

Decreasing the temperature to 600 K, the behavior of (211) does not change substantially (Figures 3.8f-l). In general, the transformation of the (211) surface is rather similar, while it occurs slightly slower at 600 K than at 700 K (Figure 3.8g). The appearance of (111) domains is just slightly reduced than at 700 K (Figures 3.8h,i: $\sim 12\%$ at 600 K *vs.* $\sim 19\%$ at 700 K). The transition matrix of Figure 3.8l is also very similar to that of Figure 3.8e. 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 3.1).Conversely, the same analysis demonstrated how the (211) surface appears as substantially static [172] at 500 K (Hüttig temperature of Cu: 447 K) in the same DPMD timescales. In such a regime the thermal bath is insufficient to trigger the reconfiguration and the surface remains trapped in the ideal (211) configuration (Supplementary Figure 3.2).

Previously, we highlighted the similarity between the final state of the simulated Cu (211) surface and the corresponding missing-row (211)(2×1) reconstruction. We deemed thus interesting to analyze with our SOAP-based approach the ideal (211)(2×1) reconstructed surface, and noticeably this surface, as reported in Figure 3.2 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. 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 which undergo frequent changes in their surrounding environment.

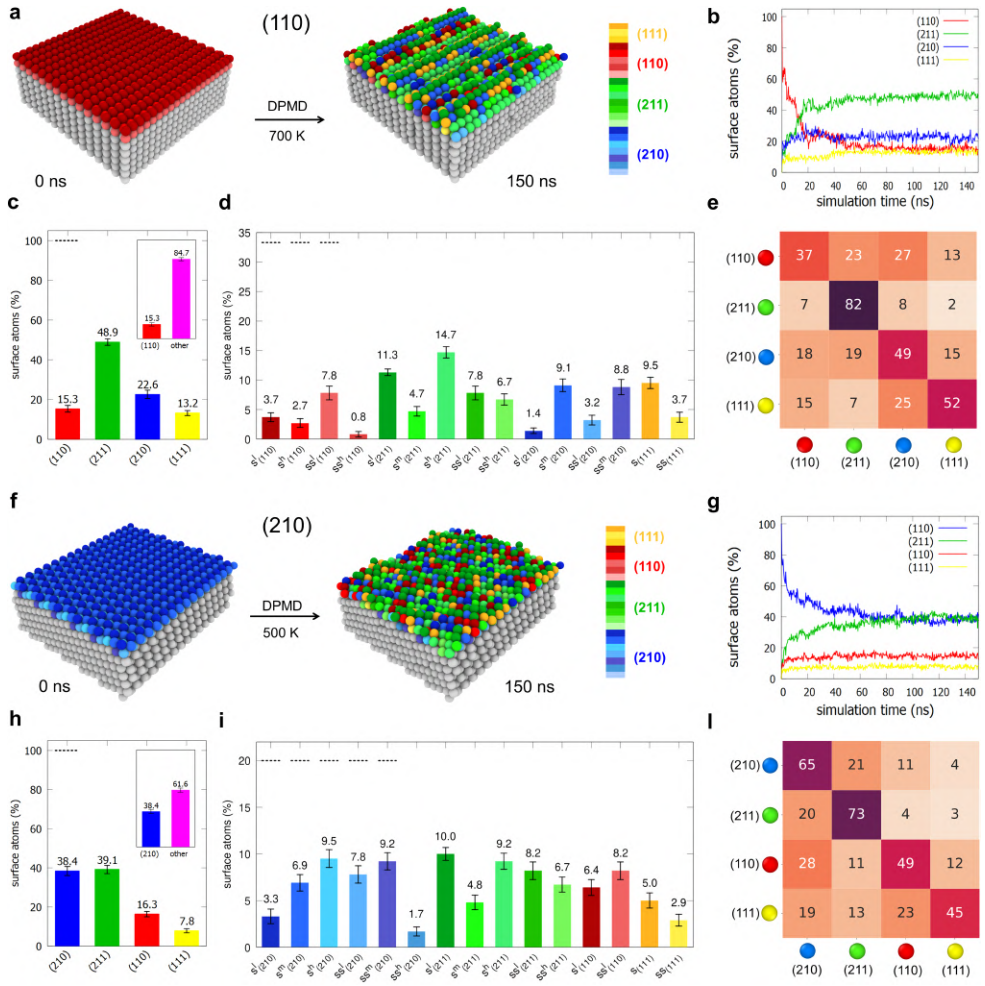


Fig. 3.9 **Dynamic reconstructions and equivalent identity of other Cu surfaces. a** Cu(110) at 0 ns (left) and after 150 ns of DPMD at 700 K (right): SOAP environments native of ideal (110) are colored in red. Green, blue and yellow colors identify non-native environments proper of (211), (210) and (111) surfaces. **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, non-native in pink. **d** Breakdown of (110) composition at 700 K. Dashed lines indicate the composition at DPMD start. **e** Transition matrix showing the probabilities for atomic transitions in (110) at 700 K (within $\Delta t = 300$ ps). **f-l** Same analyses for the (210) surface at 500 K. *Reproduced with permission from Ref* [172]

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 (Figures 3.9a-e). The snapshots of Figure 3.9a show how during 150 ns of DPMD, the surface becomes largely populated of non-red colors, and mainly of green (211) domains[172]. Figure 3.9b indicates a substantial reconfiguration of the (110) sur-

face. The instability of (110) at 700 K manifests in the fact that the surface is reconstructed by $\sim 85\%$. The native red environments drop to $< 20\%$ of the surface, reconstructing in large part (211) green domains (Figure 3.9c: rising to $\sim 50\%$). Starting from an ideal (110) configuration, such surface evolves towards reconstructing a different, more stable surface. Similar to what seen for (210) at 700 K, the transition matrix of Figure 3.9e reveals how the residual red native domains are also highly dynamic (survival probability $< 40\%$), which fits well with their relatively high surface energy,[301, 302] while the persistent domains in this surface are non-native environments. As seen experimentally for other metals,[303–305] also the Cu(110) surface appears unstable at 700 K and reconstructs large surface domains structurally/dynamically similar to the (211) surface ones (Figure 3.8). Also 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. Like for (211), at 500 K also the (110) surface is substantially immobile and preserves its identity: the thermal bath is insufficient to trigger the reconstruction in the timescales accessible *via* these DPMD simulations[172]

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 of the (110)(1×3) reconstructed surface are actually very close in terms of soap distance to those of the (211) surface (See Figure 3.3).

The Cu(210) surface is found way more dynamic than both (211) and (110) surfaces. Figures 3.9f-l show the analysis for (210) at 500 K. Even so close to the Hüttig temperature, this surface undergoes considerable reconstruction[172]. The (210) reconstructs non-native domains by $> 60\%$, preserving its identity only by $< 40\%$ (Figures 3.9g,h). Increasing the temperature to 600 K or 700 K has the unique effect of accelerating such reconfiguration, while the equilibrium populations remain substantially preserved[172]. This fits well with the higher energy of this Cu surface.[301, 302] One difference is in the dynamics of such atomic environments. The transition matrix of Figure 3.9l shows diagonal entries very close 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-persistent solid-like domains – native blue (210) and non-native

green (211) domains – coexisting with dynamic domains. This confirms that, also in this case, increasing the temperature does not generate a uniform increase of 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 Figure 3.1 clearly shows that this surface is very stable[301, 302] and does not undergo any considerable reconstructions in such regimes. Even at 700 K surface atoms with coordination $\neq 9$ emerge only sparsely and as statistical local fluctuations, indicating vibrations rather than reconstructions.

3.3 Conclusions

This chapter reports a data-driven approach that allows resolving at 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, as we present in the two following chapters. 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.*, Figure 3.6). The development of a dictionary of SOAP atomic Cu surface environments (Figure 3.7) 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 (Figures 3.8,3.9c,d,h,i). This provides an exquisitely statistical picture of these metal surfaces, and a data-driven estimation of their "*statistical identity*" in dynamic regimes (Figures 3.8,3.9). Knowing what local environments emerge, how often, 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 2400 atoms (replicating on *xy* through periodic boundary conditions). Noteworthy, the transition matrices of Figures 3.8,3.9 show transition probabilities ranging $\sim 5-40\%$ 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 Figures 3.6, 3.8, 3.9 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.[204, 248, 249]

In perspective, our data-driven approach offers remarkable opportunities to relate the innate structural dynamics of metals to their properties. While here we are interested in resolving the complex structural dynamics of metal surfaces *per se*, we envisage that this will have a considerable impact in various fields intricately controlled by the AEs or structural sites available over time. This is especially important in applications for metal surfaces such as heterogeneous catalysis and sensing, where the effectiveness and activity of the surfaces depend on the dynamic atomic environments present. These surface dynamics influence interactions with the surrounding environment, thereby affecting overall performance[306, 307, 269, 308, 268, 309, 161, 310]. This detailed control over surface and bulk dynamics can lead to significant advancements in areas such as electrochemical energy storage, corrosion resistance, and nanotechnology [311, 312]

In conclusion, this work reveals the intricate and dynamic nature of metals across various regimes, demonstrating the limitations of studying them solely as static structures, even at low temperatures. Our approach highlights the importance of examining microscopic dynamics for a deeper and more comprehensive understanding compared to traditional methods focused only on the reconstruction of internal thermodynamics. By analyzing the atomic dynamics, we can determine energy barriers, reconstruct the free energy surface, and establish the probability distributions and lifetimes of various states. Consequently, directly addressing the microscopic dynamics, we gain the ability to accurately determine macroscopic properties, ensuring a deeper and more comprehensive understanding of metallic behaviors under relevant conditions, in contrast with methods that focus solely on internal thermodynamics that do not possess the broad applicability and potential of our approach, as they are

unable to independently evaluate the system's dynamics. Thus, our work represents a paradigm shift in the study of metals, transitioning the focus from static structures to system dynamics. This shift enables a more thorough understanding of the behaviors and properties of metallic systems, uncovering complexities that structural analyses fail to capture.

3.4 Methods

3.4.1 DFT Database generation for training Cu *NN*-potential

The DFT database for training the Cu *NN*-potential was created using configurations from *ab initio* molecular dynamics trajectories of small Cu systems. Simulations were conducted with the PWscf code of Quantum ESPRESSO.[313] PBE exchange-correlation functional[314] was used to propagate nuclear dynamics, chosen for its balance between accuracy and computational cost. This functional has also proven reliable for other metals like gold and silver.[315] Ultrasoft RRKJ pseudopotentials[316] were employed to replace explicit core-valence electron interactions. Electron density and wavefunctions were expanded in plane-waves with energy cutoffs of 220 and 50 Ry, respectively. Occupations were treated using the cold smearing technique of Marzari[317] *et al.* with a Gaussian spreading of 0.01 Ry. The Brillouin zone was sampled with a $2 \times 2 \times 2$ Monkhorst–Pack k-point grid[211] for bulk structures and $2 \times 2 \times 1$ for slab models. Convergence tests for cutoff energy, k-point sampling, and occupation led to the chosen setup. AIMD simulations used a 1.0 fs time step in an NVT ensemble with the stochastic velocity rescaling thermostat.[318] Temperatures ranged between 500 and 700 K to explore a broad configurational space. The AIMD systems included both bulk and surface structures. Bulk was modeled by a periodic super-cell of $7.1638 \times 7.1638 \times 7.1638$ Å containing 32 atoms. Surface calculations involved four slab models for the (100), (110), (111), (211), and (210) surfaces (90-128 atoms). Slabs varied in atomic layers (6-8 layers), with a vacuum layer in the z-direction(16-20 Å), and the bottom two Cu layers were fixed during simulations.

3.4.2 Training the Cu *NN*-potential using DeePMD-kit

As discussed in Chapter 2, We trained the Cu *NN*-potential using the DeePMD-kit package.[273, 180] The smooth version of the deep potential model was used with a cutoff radius of 6.0 Å. The network construction included a cosine shape function to smoothly switch off the 1/r term from 1.0 to 6.0 Å. The filter network comprised three layers with (25, 50, 100) nodes per layer, and the fitting network had three layers with 240 nodes each. Training was performed with the ADAM optimizer,[319] using an exponentially decaying learning rate from 1.0×10^{-3} to 5.0×10^{-8}. The batch size was set to 4. Energy and force terms in the loss function changed during optimization from 1 to 10 and from 1000 to 1, respectively. The final model used for production runs was trained for 10.0×10^{6} steps.

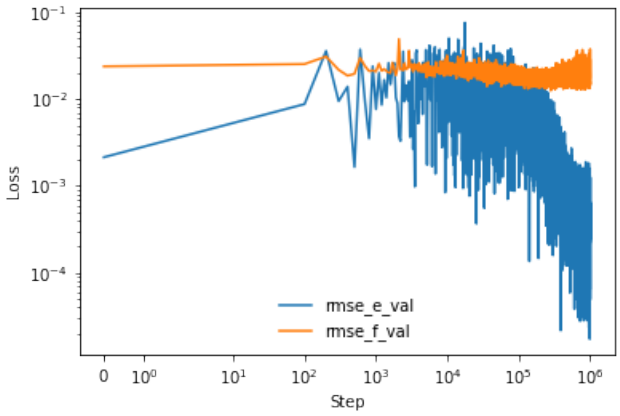


Fig. 3.10 **Root mean square (RMS) testing error of energy and forces during the training process.** The blue line represents the RMS error for energy (*mse_e_val*), while the orange line represents the RMS error for forces (*mse_f_val*). The training process involves 10 million steps, during which the errors are minimized using an exponentially decaying learning rate. *Reproduced with permission from Ref* [172]

In figure 3.10 it's possible to observe that, as training progresses, both errors decrease, demonstrating the model's improved accuracy in predicting both energy and force values. The fluctuations observed in the error values are typical during the training of complex neural network potentials, reflecting the ongoing optimization and adjustment of the network parameters.

The selection of the training dataset is critical for *NN*-potential training. Configurations from DFT MD simulations were used to train an initial Cu *NN*-potential. This potential was then employed in several DPMD simulations for different sur-

faces and temperatures to explore a broader configurational space (500-700 K) and extract new configurations for the training set. This active learning protocol follows the procedure by Deringer and Cs'anyi[274] and was implemented in DeepMD by Zhang *et al*.[275] The criterion for selecting new configurations was based on the agreement of force predictions by an ensemble of four *NN*-potentials trained on the same dataset but with different initial weights. Model deviation was measured as the maximum standard deviation of the force components. Configurations with deviations in the range of $[27 - 240 \times 10^{-3}]$ $eV/\text{Å}$ were added to the training dataset. The process continued until no new configurations were encountered, ensuring the *NN*-potential was complete for the sampled conditions. A total of 10,000 configurations were used for training, achieving root mean square errors of 1.0 meV/atom for energy and 40 meV/Å for force. To confirm the robustness and completeness of the *NN*-potential, additional tests with larger surface patches (about 600 atoms) were conducted. The deviations in forces and energies were within the training-testing errors, demonstrating the reliability of the approach and the obtained *NN*-potential. The fine sampling of atomic configurations and associated energies from DFT MD calculations allowed the inclusion of both local minima and transition states in the training dataset. This ensured the *NN*-potential could accurately represent energy differences, transition barriers, and kinetics with DFT precision.

Chapter 4

Miscroscopic analysis of atomic dynamics and statistical identities in Gold nanoparticles

In the previous chapter, we explored the dynamic behaviors of copper surfaces, emphasizing the necessity of considering these dynamics to fully understand the properties and functionalities of metallic materials. That work introduced a novel perspective by addressing and quantifying the impact of atomic dynamics on the structural properties of metallic systems. By thoroughly examining the dynamic behavior of metal atoms, we provided significant insights into how these dynamic behaviors influence the overall structure and functionality of the material[172]. This innovative approach resolved complex dynamic phenomena and elucidated their implications at a fundamental level, and, essentially proposed a new paradigm in studying metallic systems, moving the attention from structure to dynamics. Given that we already showed explicitly how the dynamics of metal surfaces can be strongly different (and more pronounced) than that of bulk, such dynamics becomes more and more prominent in metallic systems that are dominated by surface (over bulk) properties-*i.e.* small metal clusters and nanoparticles. Building on that analytical framework, we now extend our investigation to gold nanoparticles, taken as a case study of prototypical metal nanoparticles in our study "Machine Learning of Atomic Dynamics and Statistical Surface Identities in Gold Nanoparticles" (Rapetti *et al*., 2023)[171]. Here, we focus on the intricate dynamics of gold nanoparticles, which

exhibit even more complex behaviors than copper slabs at a very low temperature due to their reduced size and increased surface area.

As in previous work on copper slabs[172], our methodology combines *bottom-up* and *top-down* SOAP-based data-driven analyses to reveal the atomic environments that statistically populate Au NPs during MD simulations at various temperatures. This method allows us to track native and non-native atomic environments—those typical of other types of NPs—that continuously emerge and resorb on the NPs. By quantifying these properties, our analysis elucidates the complex dynamic behavior that characterizes such systems at temperatures well below their melting point.

Our approach achieves an atomistic-level understanding of Au NP behavior across various temperatures, a detail rarely achieved in experimental studies due to the challenges of tracking individual atomic motions over time. This in-depth understanding of the dynamic properties of atomic sites on Au NP surfaces is of practical significance, promising to enhance the design of more effective NP-based heterogeneous catalysts and improve industrial process efficiencies.

Furthermore, this research demonstrates that the methodologies developed for copper slabs can be effectively applied to other systems, highlighting the versatility and adaptability of the analytical techniques used. The approach presented here is not limited to gold nanoparticles but can be extended to other metal NPs, enabling a comprehensive understanding of their dynamic properties and enhancing their application in various fields. This cross-applicability underscores the broad potential impact of our methodologies, suggesting wide-ranging implications for the study and utilization of metal nanoparticles in diverse applications.

The following sections will delve deeper into the temperature-dependent properties of Au NPs, the application of machine learning in modeling these dynamic systems, and the broader implications of our findings for the design and optimization of nanoparticle-based catalysts and other functional materials.

Full bibliographic reference: Rapetti, D., Delle Piane, M., **Cioni, M.**, Polino, D., Ferrando, R., & Pavan, G. M. (2023). "Machine learning of atomic dynamics and statistical surface identities in gold nanoparticles." *Communications Chemistry*, 6(1), 143. Nature Publishing Group UK London. [1]

[1] In this work, my contributions included preparing the analytical framework, analyzing the results and contributing to the interpretation of the results

4.1 Introduction

Gold nanoparticles (Au NPs), with their unique electronic, catalytic, and optical properties, represent a significant area of interest in nanotechnology and materials science. Despite bulk gold often being considered an inert catalyst, Au NPs exhibit high catalytic activity[320–323], capable of oxidizing CO into CO_2[324, 320] and facilitating various other oxidative transformations. Additionally, Au NPs feature surface plasmon resonance (SPR)[325] and other distinct physical and chemical attributes, making them suitable for sensor devices and biomedical applications[325, 326].

Understanding the dynamic properties of atomic sites on the surfaces of nanoparticles and their evolution over time is crucial for the rational design of more effective NP-based heterogeneous catalysts[327, 328]. Seminal work in the field has highlighted the importance of understanding the dynamic formation of transient active sites on the surface of Au NPs[163]. These sites, which form only under specific reaction conditions, play a crucial role in catalytic processes, underscoring the need for methods capable of accurately predicting and studying these transient sites and their dynamics.

Moreover, as described in Chapter 2, recent advancements in experimental techniques have increased resolution up to single particles[329] or even individual atoms[330, 322]. High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) experiments of supported Au NPs have provided direct evidence that atoms move within the NPs at finite temperatures[331]. However, unraveling such atomic motion and obtaining quantitative insights remains challenging due to structural dispersion, the variable distribution of atomic surface sites, and the limitations of ensemble-averaged characterizations. Additionally, experimentally reconstructed NP models often lack information on the identity of individual atoms, making it impractical to track their dynamics over time.

Computational modeling, widely used for studying metals and metal clusters, holds considerable potential in addressing these challenges[244, 245, 236–238, 172, 235, 79, 239, 246]. Indeed atomistic simulations can capture individual atomic motions and track them over time, providing valuable insights into the internal atomic dynamics of Au NPs[332, 262, 322, 333–335]. Additionally, recent advancements in machine learning have enhanced the analysis of molecular dynamics trajectories

of various complex molecular systems, including Au NPs[204, 260, 248, 259, 261, 172, 262], thereby enriching the understanding of their behavior.

Building on these considerations, we investigate various shapes of Au NPs to understand how their atomic environments evolve at different temperatures. By simulating these nanoparticles We track the emergence, annihilation, lifetime, and dynamic interconversion of the AEs, allowing us to estimate a "*statistical identity*" for the nanoparticles, as already shown in previous work on Cu slabs[172](See Chapter 3). Using classical MD simulations, we analyze the atomic dynamics of Au NPs of different shapes and sizes, focusing on three primary geometries: icosahedral (Ih309, containing 309 atoms), decahedral (Dh348, containing 348 atoms), and truncated-octahedral (To309, containing 309 atoms) nanoparticles. The simulations were performed at various temperatures (300 K, 400 K, and 500 K) to capture the temperature-dependent behavior of these nanoparticles.

4.2 Results

4.2.1 Characterizing the innate dynamics of a gold NP *via* machine learning of atomic environments

As a first representative example of ideal Au NPs, we investigate, analyze and reconstruct the innate atomic dynamics of a 309-atoms icosahedron (Ih_{309}) at various temperatures. An icosahedral NP can be imagined as a series of concentric shells that envelop a single central atom. The first shell that resembles an icosahedron is constituted by the first 12 atoms surrounding the central one. Larger icosahedra can be generated by adding further surrounding atomic shells. In contrast, at each new larger shell, the NP resembles more and more the ideal platonic solid with 20 equilateral triangles as faces and 12 vertexes. Ideal atomic icosahedral NPs can be thus obtained as composed of 13, 55, 147, 309, 561, 923, *etc.*, atoms) – the so-called "magic atomic numbers" for icosahedral NPs. As a relevant example, here we start by studying the behavior of an ideal Au icosahedral NP composed of 309 atoms (Fig.4.1: Ih_{309}) at different temperatures *via* classical molecular dynamics simulations.

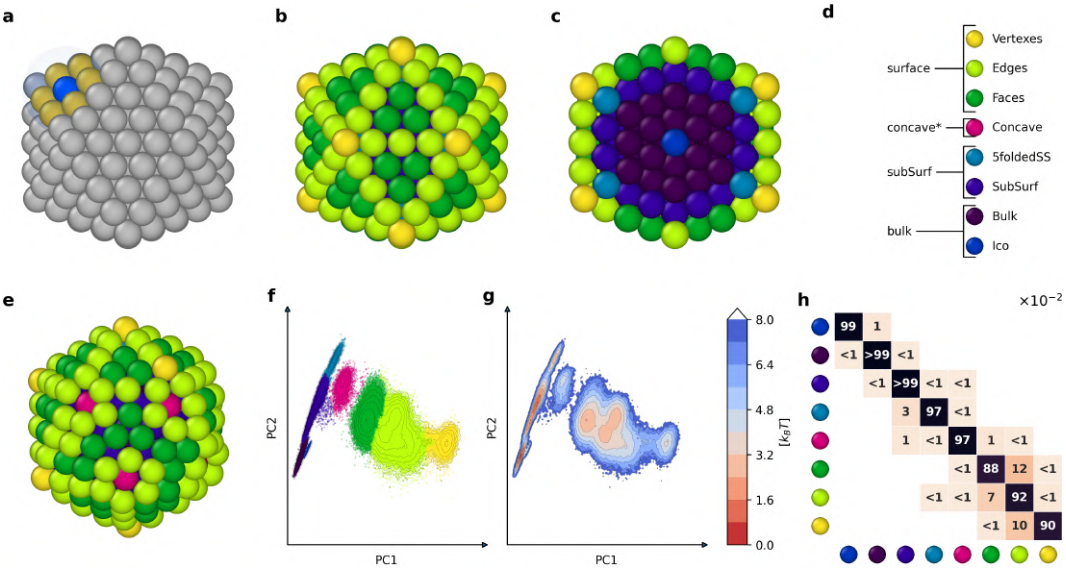


Fig. 4.1 ***Bottom-up*** **machine learning of atomic environments (AEs) and AEs' dynamics in Ih_{309} at 300 K.** (**a**) A SOAP vector is centered in each atom of the Au NP (in blue), obtaining a SOAP spectrum which is a characteristic fingerprint of the level of order/disorder in the displacement of the neighbor Au atoms (in gold) within a cutoff (shown as a transparent sphere). The SOAP spectra of all atoms in the NP (309) are calculated on 1000 frames taken every 1 ns along the last 1 μs of MD, obtaining a SOAP dataset containing 309'000 SOAP spectra in total. The main AEs that populate the Ih_{309} NP are identified *via* unsupervised clustering using the HDBSCAN* algorithm. (**b,c**) Main AEs present on the surface (**b**) and in the interior (**c**) of the ideal Ih_{309} NP before simulation start (at 0 K). (**d**) Color legend showing a structural interpretation of the AEs detected by the SOAP clusters. (**e**) Snapshot of Ih_{309} taken from the MD simulation (at 2 μs) at 300 K, with Au atoms colored based on the detected SOAP clusters. (**f**) PCA projection on the two first principal components of the SOAP dataset. The different colors identify the various main clusters detected by HDBSCAN*. (**g**) The inverse logarithm of the density of the SOAP dataset, identifying the SOAP clusters (AEs) as local energy minima. (**h**) Normalized transition matrix reporting the probabilities for an atom in a given AE at time t to remain in that AE (p_{ii}) or to undergo a transition to a different AE ($p_{i\rightarrow j}$) in dt (*i.e.*, at $t+dt$, with $dt = 1$ ns in our analyses). *Reproduced with permission from Ref* [171].

As a first step, we built an atomistic model of Ih_{309} (Fig.4.1a-c) that we simulated *via* 2 μs of MD at 300 K (see Methods section for complete computational details). From the resulting MD trajectories, we extracted 1000 frames sampled every 1 ns from the last 1 μs of MD (during which the population of the detected AEs plateau). Recently proven useful to reconstruct the structural and dynamical complexity of various types of self-organizing molecular/atomic systems[204, 248, 260, 261, 172], we used the smooth overlap of atomic positions (SOAP) vectors [202, 258] as abstract

high-dimensional descriptors capable of retaining rich information on order/disorder in the AEs – *i.e.*, in the local atoms' displacement (Fig.4.1**a**: in yellow) around each atom in the NP (in blue) within a cutoff. In particular, our SOAP analysis is performed using a cutoff of $rcut = 4.48$ Å, which was found as the best compromise between the cost of the calculation and the highest retained information in this case (see subsection 4.4.2 of Methods for more details.) We used SOAP rather than other geometrical analysis tools (such as polyhedral template matching[336] or common neighbors analysis[337]), because it is a non discreet descriptor that can capture the local atomic environment more accurately and robustly than conventional methods that tend to ignore the surface details of the system. SOAP also has the advantage of being easier to extract data from it out of the box. These features make SOAP suitable for studying dynamic materials with diverse structures and properties. At each sampled frame (1000) during the MD simulation, we calculated the SOAP [202, 258] power spectrum for each atom (309) at that frame in the Ih_{309} NP, obtaining a global SOAP dataset composed of 309'000 SOAP spectra in total. From this SOAP dataset, we then identified the main AEs that populate the Ih_{309} NP at 300 K *via* unsupervised clustering using the HDBSCAN* algorithm.[338]

This analysis finds 8 different clusters (AEs) emerging in the Ih_{309} at 300 K (Fig.4.1 (**b**-**f**). Particularly evident in Fig.4.1**b**-**c**, the identified SOAP clusters correspond to different structural AEs on the Ih_{309} NP. In detail (4.1**d**), we obtain an "*Ico*" AE, corresponding to the central atom of the icosahedral Ih_{309} NP (in blue). Shown in Fig.4.1**c**, such AE is different from the "*Bulk*" AEs (in violet) surrounding it in the bulk of the NP (this 13 atoms AE is non-crystalline, *i.e.* it is not possible to cover the whole space by units of this AE). The analysis detects "*SubSurf*" and "*5foldedSS*" AEs, identifying the atoms in the first layer below the NP surface (deep blue) and vertexes (characterized by a 5-folded symmetry axis: in light blue). On the surface of the (ideal) Ih_{309} NP (Fig.4.1**b**), our analysis detects a "*Faces*" (dark green) – close compact FCC(111) facet environments –, "*Edges*" (light green) and "*Vertexes*" AEs (the lowest coordinated atoms in the NP, in yellow). Shown in Fig.4.1**d**, the analysis also detects an additional pink cluster, namely, a "*Concave*" AE identifying the centers of the so-called rosettes (Fig.4.1**e**).[335] Noteworthy, such concave AEs are not present in the ideal Ih_{309} NP (Fig.4.1**b**) but they emerge along the MD, while their formation is known to be an energetically favored event on the surface of icosahedral NPs.[335, 247]. Fig.4.1**e** shows a representative MD snapshot of the Ih_{309} at 300 K, where a rosette triplet – formation of rosettes pertaining to

three neighbor vertexes, known experimentally – is clearly visible. (notably, each rosette center (pink) has six atom neighbors while classic vertexes (yellow) have five in a Ih_{309} NP). In particular, once a rosette triplet is formed, this configuration is found stable during the MD simulation, while even at 300 K, this is accompanied by continuous collective atoms motions that do not change the overall shape of the surface of the icosahedral NP.

Fig.4.1**f** shows the PCA (projection on the two first components PC1 and PC2) of the SOAP power spectra dataset, colored based on the SOAP clusters detected *via* unsupervised HDBSCAN* clustering. From the inverse logarithm of the PCA density, it is also possible to obtain the corresponding free energy surface of Fig.4.1**g**. From these two plots, we can clearly distinguish three different zones on the PCA and obtain the first qualitative information on their interconnection. The clump of AEs on the left collects the bulk and sub–surface environments (dark blue, violet, and light blue). These SOAP environments correspond to quite dense areas in the PCA, which indicates substantially low mobility at 300 K of the atoms that belong to these AEs in the Ih_{309} NP. The right part of the PCA is much less compact, indicating that the surface AEs (dark-, light-green, and yellow AEs: faces, edges, and vertexes) are in comparison much more dynamic at 300 K. Between the bulk and surface areas, there is a smaller zone in the PCA connecting them. The vast majority of these environments are classified as "*Concave*" (pink), which suggests that at room temperature the interior and exterior of these NPs communicate essentially *via* the creation of local “point defects” created on their surface – if we think of rosettes in this sense, as they are not present in the ideal icosahedral NP. The free energy surface of Fig.4.1**g** is derived from the density of points in each cluster, but the population of each cluster does not weight it. This means it does not fully represent the probability of an individual atom visiting the landscape. Nonetheless, this shows (i) how the various detected SOAP clusters (AEs) correspond to local density maxima and energy minima and (ii) that the barriers separating the surface states are relatively low, which allows for the considerable atomic exchange between these AEs.

Quantitative information on the internal dynamics of the NP can be obtained by tracking the SOAP spectra of all atoms at each sampled MD snapshot and monitoring their change. In particular, this allows us to analyze to what SOAP AE each atom belongs to at time t and at each successive timestep (*i.e.*, at $t + dt$, with $dt = 1$ ns in our analyses). Fig.4.1**h** shows a normalized transition matrix for the Ih_{309} NP at 300 K. This contains all probabilities (all values reported in the matrix are to be

intended as multiplied by $\times 10^{-2}$) for an atom in a given AE i to remain in that AE (p_{ii}) or to undergo transition to a different AE ($p_{i\to j}$) in dt (the rows of the matrix sum to 1).

Fig.4.1**h** shows that at 300 K the deep/core AEs tend to be rather stable ($p_{ii} \sim 1$: atoms belonging to such AEs have a high probability of remaining in such state in $dt = 1$ ns). On the other hand, a significant inter-AE exchange is already observable ($p_{i\to j} \geq 1$) at this temperature on the NP surface. Indeed, the matrix shows that in such conditions most of the action takes place in the sub-square in the matrix connecting the '*Faces*', '*Edges*', and '*Vertexes*' AEs (dark-, light-green, and yellow AEs).

From the transition probabilities of Fig.4.1**h**, it is possible to estimate the average lifetime of the various AEs and the transition rates between them. In particular, the off-diagonal entries ($p_{i\to j}$) divided by dt give the transition rates between two AEs $i \to j$, $k_{i\to j}$, from which one can estimate the characteristic timescales for the various transitions as: $\tau_{i\to j} = k_{i\to j}^{-1}$. The number of times a given transition event $i \to j$ is registered in the system along the last 1000 ns can be estimated as: $n_{i\to j} = [i] \cdot 1000/\tau_{i\to j}$, where $[i]$ is the average number of atoms in the i^{th} AE. For example, in the Ih_{309} NP at 300 K an atom in the *Faces* AE (dark-green) has a transition probability to the *Edges* AE (light-green) of $p_{Faces\to Edges} \sim 0.12$ ($\sim 12 \times 10^{-2}$), indicating a transition rate of $k_{Faces\to Edges} \sim 0.12$ ns^{-1} and characteristic transition timescale $\tau_{Faces\to Edges} \sim 8.3$ ns Furthermore, since this is the fastest transition involving the *Faces* AE, this sets the bottom limit for the lifetime of an atom in the (111) faces of this NP at 300 K (minimum residence time) as $\tau_{Faces} \sim 8.3$ ns.

Similar estimations for other dynamic transitions between the AEs within the NP can be calculated from the transition matrix of Fig.4.1**h** in an analogous way. We note that given the time window used for the analysis we report herein ($dt =$ 1 ns), any observed communication/exchange between the AEs involves processes happening on the *ns* scale or slower, thus reducing the probabilities that the AE exchanges are related to thermal vibrations (values related to $p_{i\to j} < 0.01$ (*i.e.*, 1 $\times 10^{-2}$ in Fig.4.1**h**) should be considered as purely qualitative, as these pertain to events that are only sparsely observed along the MD simulation). We also underline that, while the exact estimated values for AEs' lifetimes, probabilities and transition rates may slightly change depending on the employed FF,[227, 271, 339, 334] tests conducted with different types of FFs[270, 315] provided very similar results in terms of NP dynamics, confirming the generality of our observations.

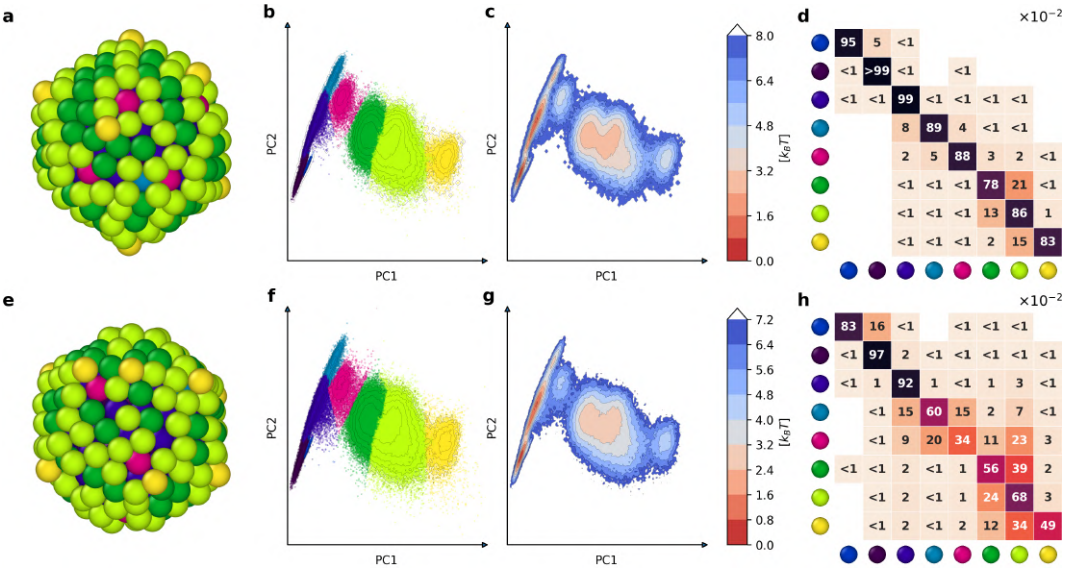


Fig. 4.2 **Effect of temperature on the Ih_{309} NP dynamics.** (**a**) MD frame of the Ih_{309} NP taken from the equilibrated-phase MD simulation at 400 K (atoms colored based on SOAP clusters of Figure 1). (**b**) PCA projection of the SOAP dataset obtained from the MD simulation of the Ih_{309} NP at 400 K. (**c**) Free energy surface (FES) obtained from SOAP PCA. (**d**) Normalized transition matrix indicating the residence (p_{ii}) and transition probabilities between the AEs in the Ih_{309} NP at 400 K of temperature in the time interval $dt = 1$ ns (all p_{ii} and $p_{i\to j}$ values are $\times 10^{-2}$). (**e**) MD frame of the Ih_{309} NP taken from the equilibrated-phase MD simulation at 500 K. (**f**) PCA projection of the SOAP dataset obtained from the MD simulation of the Ih_{309} NP at 500 K. (**g**) Associated free energy surface (FES). (**h**) Normalized transition matrix indicating the residence (p_{ii}) and transition $p_{i\to j}$) probabilities ($\times 10^{-2}$) between the AEs in the Ih_{309} NP at 500 K in the time interval $dt = 1$ ns. *Reproduced with permission from Ref* [171].

We simulated the Ih_{309} NP also at 400 K and 500 K by running 2 μs of MD. During this timescale, the NP attains an equilibrium state, with numerous dynamic events and transitions occurring on its surface [171]. We then extracted the SOAP spectra for all atoms from 1000 frames taken from the last 1 μs of MD following the same protocol used at 300 K. Fig.4.2 shows the results of these additional analyses. In particular, in these analyses, we used the simulation at 300 K as the training set for both the PCA computation and clustering (HDBSCAN*) analyses of the MD trajectories of the Ih_{309} NP at 400 K and 500 K. The PCA of Fig.4.2**b** and Fig.4.2**g** show how the clusters on the surface of the NP become more adjacent to each other at 400 K and 500 K than at 300 K. Moreover, the FES of Fig.4.2**c** and Fig.4.2**h** indicates that the minima corresponding to different surface AEs at 300 K tend to merge together when the temperature increases. In particular, at

500 K, the surface AEs constitute a unique large minimum, meaning that at such temperature, *e.g.*, *Faces* and *Edges* AEs are in continuous exchange with each other and that these effectively form a unique fuzzy surface state (*i.e.*, computing the PCA on the MD trajectory at 500 K would not find at all two distinct *Faces* and *Edges* AEs, but one single environment) The communication between *Faces*, *Edges*, and *Vertexes* AEs increases with increasing the temperature. This is even more evident in the normalized transition matrices of Fig.4.2**d** and Fig.4.2**i**. The sub-square in the matrices connecting the '*Edges*', '*Faces*', and '*Vertexes*' AEs shows that atoms belonging to these environments have larger probabilities to exchange with each other at 400 K and 500 K than at 300 K. In particular, at 400 K, such a surface atomic mobility is evident, but these atoms have still a higher probability of remaining in their environment than of jumping into another one in $dt = 1$ ns ($p_{ii} > 50\%$). On the other hand, at 500 K the residence probability for atoms in the surface AEs drops close to, and in some cases also below 50 %, suggesting that in such conditions the NP surface is pre-melting.[340] We note that, in good approximation, the number of atoms in each environment does not vary much during the simulations at all temperatures[171]. This suggests that, despite such rich atomic mobility, the Ih_{309} NP surface remains structurally that of an icosahedron at all analyzed temperatures. It is also interesting to note that the transition matrix of Fig.4.2**f** shows sparsely observed dynamic interconnections between the central atom of the NP and the surface AEs at 500 K. This does not mean that the central atom is diffusing to the NP surface, but rather that at such temperature internal voids may rarely form in the NP center, which makes the SOAP spectrum of deep bulk atoms change occasionally and become similar to that of surface AEs. This fits well with previous reports showing similar central vacancies in Ih_{309}.[173, 247]

While evidence of surface dynamics in Au NPs have been reported,[327, 322] obtaining clear insights as on the processes that characterize such dynamics, or on whether this is essentially due to, *e.g.*, local atomic reconfigurations or atomic diffusion is non-trivial. Tracking the motions and the fluctuations in the SOAP spectra of the individual atoms in the NP, our approach provide clear evidence that the dynamics of these NPs is not due only to oscillations between adjacent/similar AEs, but to real microscopic atomic diffusion. As a representative example, in Fig.4.3 we show the detail of the evolution during 1 μs of MD of an atom that is an Ih_{309} vertex at the start of the simulation. Fig.4.3**a**,**b** show respectively the temporal trajectory of the atom and the SOAP AEs that this visits in this time frame,

revealing how even at 300 K such atom visits a surprisingly large collection of different (surface and sub-surface) SOAP AEs (Fig.4.3**b**). In particular, such vertex atom diffuses first to surface AEs, and then also penetrates into the sub-surface. Its diffusion is described in detail also in the plot of Fig.4.3**c**, showing how such an atom also becomes at a certain point (∼820 ns of MD) a rosette center (pink).

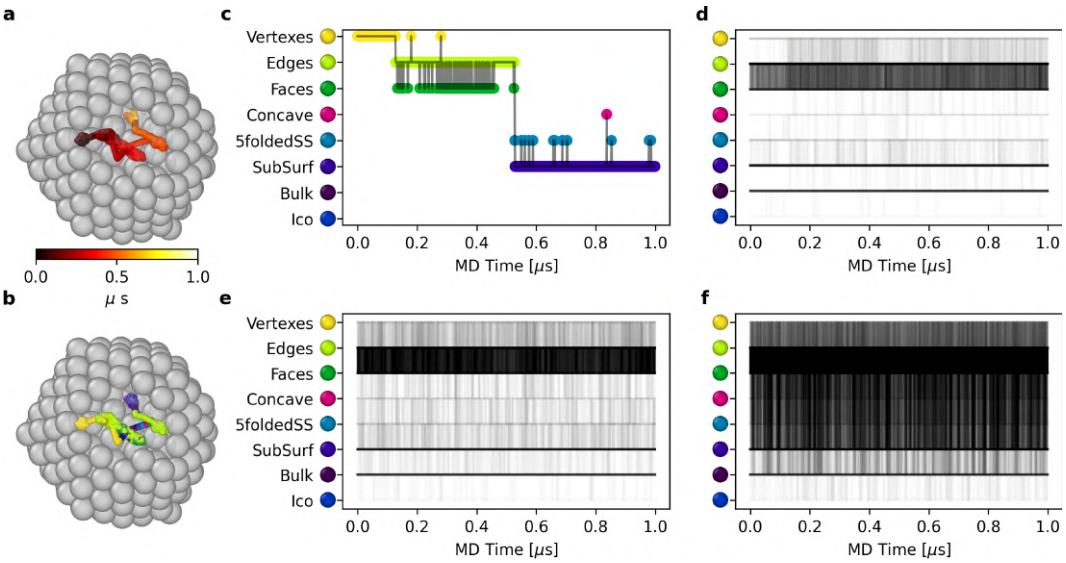


Fig. 4.3 **Atomic diffusion on the** Ih_{309} **NP.** (**a**) MD trajectory of an atom in the Ih_{309} NP at 300 K, colored based on simulation time. (**b**) MD trajectory of the same atom colored based on its SOAP AE (Fig.4.1): being initially a vertex, the atom diffuses on the NP surface visiting various surface and even sub-surface AEs. (**c**) AEs visited by the tracked atom during the MD of the Ih_{309} NP at 300 K. (**d**) AEs' transitions of all (309) atoms in the Ih_{309} NP at 300 K: at room temperature, only the surface of the NP appears as dynamic. (**e**) AEs' atomic transitions in the Ih_{309} NP at 400 K. (**f**) AEs' atomic transitions in the Ih_{309} NP at 500 K: surface pre-melting. *Reproduced with permission from Ref* [171].

Fig.4.3**d-e** show us AE visited by all the 309 atoms in the NP at 300 K (**d**), 400 K (**e**), and 500 K (**f**). These graphs reveal which environments are most prone to exchange in this NP. In particular, at 300 K only *Faces* and *Edges* surface AEs are dynamic. At 400 K the dynamics of the NP surface increases, but remains similar to that at 300 K (which fits well with the transition matrices of Fig.4.1**h** and Fig.4.2**d**). On the other hand, at 500 K the atoms exchange between all surface and sub-surface AEs (surface pre-melting).

We underline how all the analyses reported above are purely *bottom–up*, meaning that all information on the AEs, their similarity, classification, and dynamics are reconstructed only from the MD trajectories and in an unbiased data-driven way. At

the same time, such data-driven analyses are not always straightforward to interpret. For example, given the surface of such NPs is in continuous motion and new non-native states may also emerge in these NPs (*e.g.*, concave ones), a relevant question is whether such new non-native emerging AEs are closer to the native ones (proper of that type of ideal NP) or, *e.g.*, to other AEs native of different types of NPs. To answer such questions and obtain a more complete picture, we employed a different type of analysis.

4.2.2 A dictionary of Au NPs SOAP environments

To complement our study, we designed a different *top–down* analysis. We defined a "general" and transferable dictionary of SOAP environments analyzing ideal Au NPs (at 0 K) of different sizes and morphologies. We then used it to identify the native and non-native AEs that emerge in the simulated NPs and to analyze their dynamics at different temperatures. As described in the work on Cu slabs, at this point we created a dictionary of Au AEs (Fig.4.4) that contains all AE typical of different shape NPs. In particular, we calculated the SOAP atomic spectra of two ideal icosahedral Au NPs: Ih_{309}, simulated in Fig.4.1 and in Fig.4.2, and a larger one composed of 923 Au atoms (Fig.4.4**a**, right: in blue). We also calculated the SOAP atomic spectra of three decahedra composed of 348 (Dh_{348}, simulated in the next section), 1086 and 1734 Au atoms, and two truncated octahedra, composed of 309 (To_{309}, simulated in the next sections) and 807 Au atoms (To_{807}). In the AE dictionary, we also added an additional To_{976} (not included in the figure). Such a collection of NPs allowed us to maximize the number of sample AEs, obtaining a complete SOAP dictionary for *Ih*, *Dh*, and *To* Au NPs.

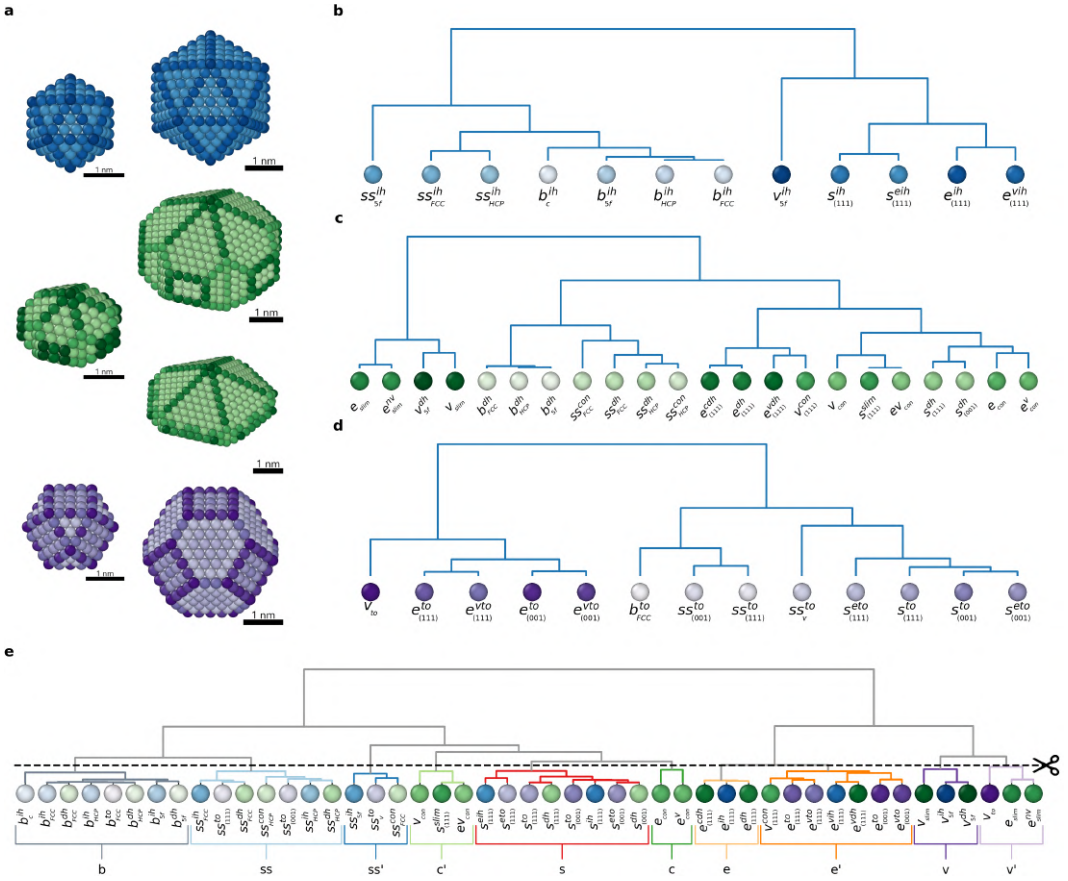


Fig. 4.4 **The dictionary of atomic SOAP NP environments.** (**a**) Icosahedral (blue), decahedral (green) and truncated-octahedral Au NPs used to generate the SOAP dictionary of AEs. Together with the NPs that we simulate herein (the three on the left), also larger size NPs are included in the dictionary, in order to guarantee that this contains all AEs typical of the NP families. (**b**) Dendrogram connecting the various SOAP AEs proper of icosahedral NPs (blue) connected based on their SOAP distance. (**c**) Dendrogram for the SOAP AEs native of ideal decahedral NPs (green). (**d**) Dendrogram for the SOAP AEs native of ideal decahedral NPs (purple). (**e**) Global dendrogram connecting all the AEs of the various NP types, hierarchically classified based on their SOAP similarity using the SOAP distance. Cutting the dendrogram at a certain SOAP distance provides a coarse-grained dictionary that groups together the AEs with SOAP distance $d_{SOAP} \leq 0.08$. *Reproduced with permission from Ref* [171].

The obtained AE dictionary contains a total of 47 different SOAP environments(see Methods subsection 4.4.3 for details about the creation of this dendrogram). Fig.4.4**b** shows the 12 ones typical of icosahedral Au NPs, organized hierarchically in a dendrogram based on their SOAP distance and similarity (see previous Chapter for more detail about this metrics). Fig.4.4**c** shows the dendrogram containing the 22

environments characteristic of the decahedral Au NP. Fig.4.4**d** shows the dendrogram of the 13 SOAP AEs typical of truncated-octahedral and cuboctahedral NPs. Finally, Fig.4.4**e** shows the complete SOAP dictionary, containing all AEs proper of *Ih*, *Dh* and *To* Au NPs, organized based on their SOAP distance and similarity. Such a dictionary of SOAP spectra can be then used to compare and classify the AEs that emerge along the MD simulations of a given NP, and to understand if on specific NP AEs emerge that are closer to those present in NPs of other shapes. While the dendrogram of Fig.4.5**e** contains the complete information, this also shows, *e.g.*, that most of the bulk environments across different-shape NPs are basically identical to each other, as expected. Thus, to ensure to capture of relevant variations in our analysis, we opted to "truncate" the dendrogram at the distance of $d_{SOAP} = 0.08$ (and considering as relevant only differences larger than this), see Fig.4.12 and Fig.4.11 for an example on how the choice of this parameter influences the environments and the analysis of the dynamics. The cut at $d_{SOAP} = 0.08$ reduces the 47 AEs to 10 AEs, improving the clarity and the statistical relevance of the subsequent analysis. Nonetheless, the resolution of such analysis can be in principle adapted, based on the relevance of the difference between the AEs. This "cut" provides a coarse-grained analysis, which regroups all AEs with $d_{SOAP} < 0.08$ in macro-clusters all SOAP environments

In particular, the truncated dendrogram of Fig.4.5**a** shows the final 10 AEs considered in the analysis. The *b* AE collects all NP bulk environments. The *ss* and *ss'* AEs collect all sub-surface AEs: *ss* identify the AEs under the FCC(111) and FCC(001) NP faces and those under the NP edges, *ss'* identify the "non-standard" subsurface AEs under the vertexes and the convex elements. The *c* and *c'* AEs enclose the concave environments. The *s*, *e* and *e'* enclose all surface AEs: *s* collects the AEs proper of FCC(111) and FCC(001) faces, *e* and *e'* collects edge AEs, while *v* and *v'* those proper of vertexes. We used the "coarse-grained" SOAP dictionary for analyzing our MD simulations and distinguishing between native and non-native AEs emerging in the simulated NPs.

4.2.3 A dynamic dance of native and non-native AEs shaping the surface identity of Au NPs

In Fig.4.5**b** we show the Ih_{309} NP in its ideal configuration (at 0 K: after minimization) and at various temperatures (top-to-bottom: 300 K, 400 K, 500 K). In particular, the Ih_{309} snapshots at 300 K, 400 K, 500 K correspond to the same MD frames of Fig.4.1**e**, and Fig.4.2**a** and e, but in Fig.4.5**b** the atoms are colored based on the similarity of their atomic environments and those contained in the SOAP dictionary of Fig.4.5**a** (*top–down* analysis). This analysis allows us to track in detail which ones of the AEs populating the Ih_{309} NP at the various temperatures belong to the family of the native ones, typical of icosahedral NPs, and which ones are non-native – namely, closer to those natives of different shape NPs, such as *e.g.*, decahedra or truncated octahedra. For each NP snapshot, a pie chart (bottom-right) shows the percentage of surface atoms belonging to native (in gray) and non-native AEs (in pink) in the Ih_{309}. The analysis shows how the percentage of emerging non-native environments increases with increasing temperature, essentially due to increased thermal fluctuations and surface reconstructions. The histograms of Fig.4.5**c** report the average number of atoms belonging to each AE in the last 1 μs of MD (equilibrated-phase MD trajectories). For each AE, we represent the count with four columns (Fig.4.5**c**): the first one refers to the AE populations in the ideal Ih_{309} NP, and the other three columns refer to the AE populations in the same NPs at the three simulated temperatures. The *v'* (purple), *s'* (light green), *c* (green) AEs are non-native AEs, in that these are not present in the ideal Ih_{309} (identified by arrows), and in icosahedral NPs in general, but emerge in Ih_{309} with temperature.

At each MD time-step we know the cluster each atom belongs to so that we can track where the atoms come from and where they go in terms of AEs all along the MD trajectories. This allows us to draw the chord diagrams of Fig.4.5**d**, showing the dynamic interconnections between the various AEs populating the NP.

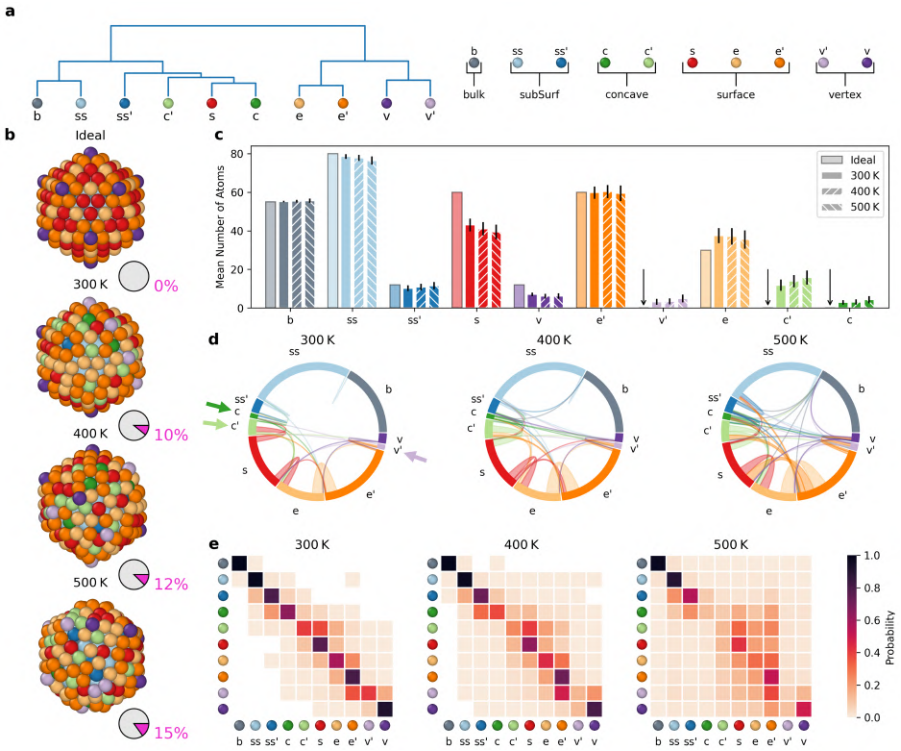


Fig. 4.5 ***Top-down* data-driven reconstruction of the innate dynamics and statistical identity of the Ih_{309} NP.** (**a**) Dictionary of AEs and associated dendrogram used for the *top-down* analyses, obtained *via* cutting the complete dendrogram at $d_{SOAP} = 0.08$ (as shown in Fig.4.4e). (**b**) The Ih_{309} NP before simulation start (top, 0 K), in steady state MD frames, taken from the MD simulations at temperature 300 K, 400 K, or 500 K (top-to-down). Coloring based on the dictionary AEs (**a**): the pie charts indicate in magenta the percentage of atoms on the NP surface that do not belong to environments native to ideal icosahedral NPs. (**c**) Histogram counting the average number of atoms in each cluster during the last 1 μs of MD at 300 K, 400 K, 500 K (second-left to right columns for each AE), compared to the AE populations in the ideal (0 K) Ih_{309} NP (leftmost column for each AE). Standard deviations as vertical black lines. An arrow in place of the first column highlights the absence of certain AEs in the ideal Ih_{309} NP – *i.e.*, these AEs are non-native of ideal icosahedral, and emerge with temperature (*e.g.*, in Ih_{309}: v', c' and c AEs in light-purple, light- and dark-green respectively). (**d**) Chord diagrams showing the interconnection between all AEs communicate with each other in Ih_{309} at various temperatures. Non-native AEs emerging in Ih_{309} are identified by colored arrows in the chord plot at 300 K (*cf.* main text for details). (**e**) Normalized transition matrices reporting the probabilities for atoms in the Ih_{309} NP at the various temperatures to remain in a given AE (p_{ii}) or to exchange into another one ($p_{i\to j}$) in the time interval of $dt = 1$ ns. *Reproduced with permission from Ref*[171].

Qualitatively, the width of the corona arcs represents the total number of transitions(or not) that happened to a given cluster during the simulation, and the chords between two clusters show the interconnections between the various AEs, the more

the base of the chord is extended the more the cluster has given atom to the one it is connected to. The color of the chords indicates the net flux (*e.g.*, the chord connecting the red (s) and violet (v') clusters is violet, meaning that more atoms are observed to undergo a transition in $v' \rightarrow s$ direction than *vice versa*). The results qualitatively show with what AEs the non-native *v'*, *s'*, and *c* ones are primarily connected, suggesting where these non-native AEs come from and where they go. In particular, *v'* is connected with *e* edge atoms (light orange). The non-native concave *c* and *c'* AEs are connected with edge (orange), surface (red), and vertex (violet) atoms. The chord diagrams show that the exchange between the AEs increases with temperature (increasing number of chords and of chords' widths moving left-to-right in Fig.4.5**d**). As seen also in Fig.4.1, this analysis confirms that at 300 K only surface AEs exchange dynamically, while *ss* and *b* clusters remain relatively static and separated (dynamic surface and static interior). Increasing the temperature, and in particular, at 500 K (Fig.4.5**d**: right), the interior of the NP starts to communicate with the surface (see cyan and gray chords going towards surface AEs).

To obtain more quantitative information on the complex atomic dynamics present in the NPs, we calculated the transition probabilities for atoms belonging to such AEs to remain in or undergo a transition into the different AEs in $\Delta t = 1$ ns (same analysis of Fig.4.1**h**, Fig.4.2**d** and Fig.4.2**h**, but with this new set of *top-down* detected AEs). The transition matrices of Fig.4.5**e** show that deep AEs (*b* and *ss*) have diagonal entries residence probabilities $p_{ii} \sim 1$ (dark colors). This confirms the rather static behavior of the interior of the NPs at all temperatures. At 500 K the blue *ss'* AE starts to communicate with the surface of the NP and, in particular, with the green *c* and *c'* concave AEs. The matrices of Fig.4.5**e** thus show that the formation of the rosettes on Ih_{309} comes from such deep states, as well as (in large part) from the surface (red, *s*) and edge (*e*,*e'*) AEs. From the transition matrices of Fig.4.5 it is possible to estimate, *e.g.*, the transition probabilities, rates, timescales, and lifetimes of all these *top-down* detected AEs as done from that of Fig.4.1h. For example, in the Ih_{309} NP at 300 K the non-native concave *c'* AE (light green) has a transition probability to the surface (red, *s*) AE of $p_{c' \rightarrow s} \sim 0.4$, indicating a transition rate of $k_{c' \rightarrow s} \sim 0.4$ ns^{-1} and characteristic transition timescale $\tau_{c' \rightarrow s} = 2.5$ ns [171]Given that in the *c'* row of the matrix the $c' \rightarrow s$ transition is by far the fastest one, in good approximation, this allows estimating the bottom limit of the lifetime of one atom in the *c'* AE in the range of $\tau_{c'} \sim 2.5$ ns. Similar estimations of the characteristic timescales for all transitions between the AEs in the NP can be easily performed

from all the $p_{i\to j}$ reported in the transition matrices of Fig.4.5**e**. Such analyses thus provide not only an estimation of the average composition of an NP but also, and perhaps even more interestingly for practical applications (*e.g.* reactivity), detailed information on the lifetime of all the native and non-native AEs populating it. In fact, the capability of an AE to activate a chemical reaction is directly related to its lifetime *vs.* by the characteristic time of the reaction itself on that AE. For example, while it is known that different atomic sites have, *e.g.*, different reactivity and efficiency in catalyzing chemical reactions,[161] obtaining a structural/dynamical map showing how long all AEs in the NPs live (τ_i) and how quickly they interconvert into other ones ($\tau_{i\to j}$) is key to understand their effective efficiency . In fact, from a statistical point of view, if one AE has a given average lifetime τ_i, but the characteristic timescale for the reaction to occur on that specific AE is $\tau_{react} > \tau_i$, the probability for effectively activating the reaction on that AE would be proportional to τ_i/τ_{react}. This would indicate of how many times, in principle, the reactant should get in touch with the same AE to effectively activate a given reaction. Of course, performing practical estimations in this sense would require focusing on a realistic case and also estimating the reactivity of all visited AEs in the NP. While this is not the main point of this chapter, this is certainly feasible, which underlines the potential of the approach. Moreover, we stress that such a purely probabilistic interpretation stands as far as the reactive species do not significantly alter the dynamics and features of the AEs present on the NP – *e.g.*, no or negligible chemisorption (if such a condition does not hold, a proper reactive parametrization and simulation of the system is needed, where new AEs may appear on the NP surface upon interaction with the reactants).[341]

Similar as to what was observed in the matrix of Fig.4.2**h**, the matrix of Fig.4.5**e** (right) shows that at 500 K the surface of Ih_{309} NP is basically pre-melted[340] (p_{ii} of surface AEs < 0.5, meaning that in $\Delta t = 1$ ns the atoms in those AEs have a higher probability to move to another AE, rather than to remain in the same one). This indicates that the entire Ih_{309} has atomic dynamics faster than the nanosecond scale (liquid-like dynamics). On the other hand, at the resolution of our analyses, the atomic dynamics on the NP surface appear as more “discrete” at 300 K and 400 K (solid-like dynamics).

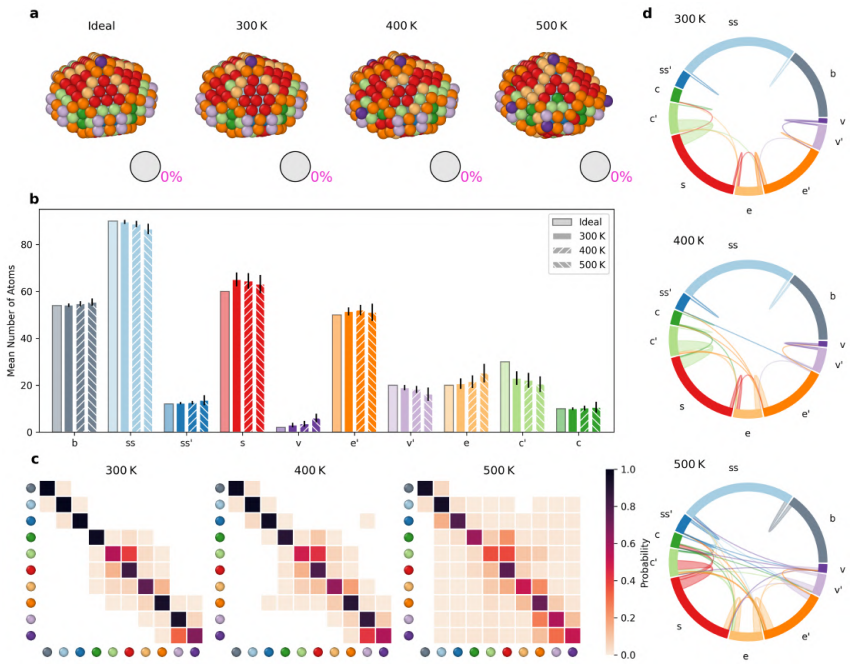


Fig. 4.6 ***Top-down* analysis of AEs in the Dh_{348} NP.** (**a**) Snapshots of the ideal Dh_{348} (0 K) and at 300 K, 400 K and 500 K. Atoms within the NPs are colored based on the AEs dictionary. (**b**) Histogram counting the average number of atoms in each AE during the last 1 μs of MD simulation at 300 K, 400 K, 500 K (second-left to right columns for each AE). Standard deviations as vertical black lines. (**c**) Normalized transition matrices reporting the probabilities for atoms in the Dh_{348} NP to remain in a given AE (p_{ii}) or to exchange into another one ($p_{i\rightarrow j}$) in the time interval of $dt = 1$ ns at the various temperatures. (**d**) The chord diagrams show the dynamic interconnections between all AEs detected in the NP at different temperatures. *Reproduced with permission from Ref* [171].

We repeated the same analysis for the Dh_{348} NP (Fig.4.10). The data show that this NP is more stable than Ih_{309} at all simulated temperatures. This NP has been chosen because its ideal conformation shows at least one atom per each cluster in the set that we identified with the cut. Consequently, this results in the pie charts of Fig.4.10a showing always 0 % pink. At 300 K and 400 K, the internal *b* and *ss* AEs are not in communication with the surface. At 500 K, some communication arises but also at such high temperatures in this case the atomic dynamics on the NP surface appear as "discrete" and closer to that of Ih_{309} at 300 K and 400 K (solid-like dynamics).

Among the investigated NPs, To_{309} (Fig.4.7) is a very interesting case. In fact, the To_{309} arrangement is known to be a non-favorable FCC arrangement. Indeed, To_{309} is more dynamic than Dh_{348} at all investigated temperatures. Interestingly,

at 300 K this NP is found less stable and more dynamic than Ih_{309}, with $\sim 24\%$ of its surface atoms in non-native AEs (pie charts in Fig.4.10**a** and histograms of Fig.4.10**b**). However, at 500 K, To_{309} is found more stable than Ih_{309} and its surface shows "discrete" atomic dynamics.

We compared the chord diagrams obtained with the *top-down* and *bottom-up* analysis (see Fig.4.9) to evaluate their consistency in terms of fluxes. The results show that the two analyses are consistent in this regard. However, comparing actual kinetics is not straightforward because the AEs identified by the two analyses are not directly related. The *top-down* analysis defines AEs based on a dictionary, while the *bottom-up* analysis derives them from the MD trajectories. Moreover, the consistency between the two analyses strongly depends on the cutoff for the dendrogram in the top-down analysis. Both analyses are therefore complementary and provide different insights into the atomic dynamics of Au NP's.

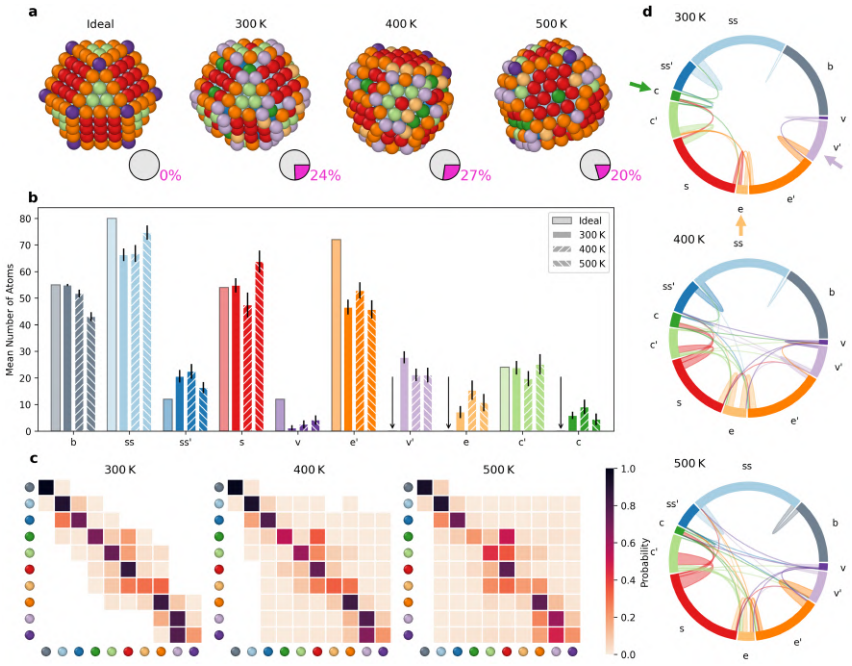


Fig. 4.7 ***Top-down* analysis of AEs in the To_{309} NP.** (**a**) Snapshots of the ideal To_{309} (0 K) and at 300 K, 400 K and 500 K. Atoms within the NPs are colored based on the AEs dictionary (the pie charts indicate in magenta the percentage of atoms on the NP surface that do not belong to environments native to ideal truncated-octahedral or cuboctahedral NPs. (**b**) Histogram counting the average number of atoms in each AE during the last 1 μs of MD simulation at 300 K, 400 K, 500 K (second-left to right columns for each AE). Standard deviations as vertical black lines. (**c**) Normalized transition matrices reporting the probabilities for atoms in the To_{309} NP to remain in a given AE (p_{ii}) or to exchange into another one ($p_{i\to j}$) in the time interval of $dt = 1$ ns at the various temperatures. (**d**) The chord diagrams show the dynamic interconnections between all AEs detected in the NP at different temperatures. *Reproduced with permission from Ref* [171].

Altogether, these results show that such analysis is transferable and flexible. In particular, this can be used *(i)* to obtain a thorough characterization of the complex atomic dynamics of the NPs that is difficult to attain with other approaches, and *ii* to compare and classify different types of metal NPs based on the AEs that emerge and populate their structure and on their complex dynamics.

4.2.4 Thermal effects on structural stability and dynamics of Ih_{309} Gold Nanoparticle

This section delves deeper into the structural stability and dynamic behavior of gold nanoparticles, specifically focusing on the Ih_{309} configuration, to enhance our understanding of the underlying mechanisms governing these systems, serving as confirmation and support for the analyses presented earlier.

As depicted in Fig.4.8, the histograms show the average number of atoms per cluster (AE) for the Ih_{309} nanoparticle at three distinct temperatures: 300 K, 400 K, and 500 K. This analysis, carried out using the *bottom-up* classification approach, reveals notable shifts in the distribution of atomic environments as the temperature increases. Specifically, at higher temperatures, we observe a significant increase in atomic mobility and reconfiguration within the nanoparticle. This indicates an enhanced dynamism, suggesting potential alterations in structural stability as the system is subjected to thermal agitation.

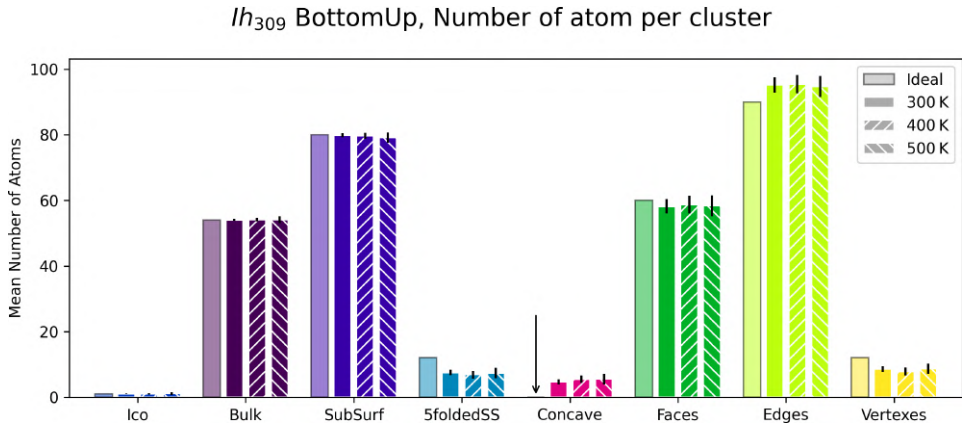


Fig. 4.8 **Histograms of AEs for Ih_{309}.** Histograms showing the average number of atoms per cluster, calculated for the simulations Ih_{309} at 300 K, 400 K, and 500 K analyzed with the *bottom-up* classification. *Reproduced with permission from Ref* [171].

To further complement these findings, 4.1 presents detailed energy calculations for various 309-atom gold clusters, including the Ih_{309} and To_{309} configurations. The Best309 configuration, which approximates a decahedral structure, exhibits the lowest energy, underscoring its higher stability relative to other configurations. Interestingly, the Ih_{309} configuration with three rosettes, though slightly higher in energy, demonstrates significant stability, implying its energetic favorability under certain conditions. This nuanced energy landscape suggests that while Best309 is the

most stable, Ih_{309} configurations can also maintain stability depending on specific environmental factors.

cluster	Energy	ΔE from Lowest
Ih_{309}	-1125.87 eV	+3.16 eV
Ih_{309} With3Rosette	-1126.86 eV	+2.17 eV
To_{309}	-1125.3 eV	+3.73 eV
Best309($\sim Dh$)	-1129.03 eV	0

Table 4.1 Energies of the ideal 309 Au clusters, for an Au_{309} near the global minimum configuration, and an Ih_{309} with 3 rosettes making an anti-Mackay face.

Delving deeper into the differences in net fluxes and the influence of varying cutoff distances on the clustering of AEs, Fig.4.9 compares the net fluxes of atomic environments in the Ih_{309} nanoparticle across different temperatures using both *bottom-up* and *top-down* analyses. The discrepancies observed between these two methodologies highlight the inherent complexity of atomic dynamics. This underscores the necessity of employing multiple classification approaches to garner a comprehensive understanding of nanoparticle behavior, as each method offers unique insights into the atomic interactions and transitions occurring within the system.

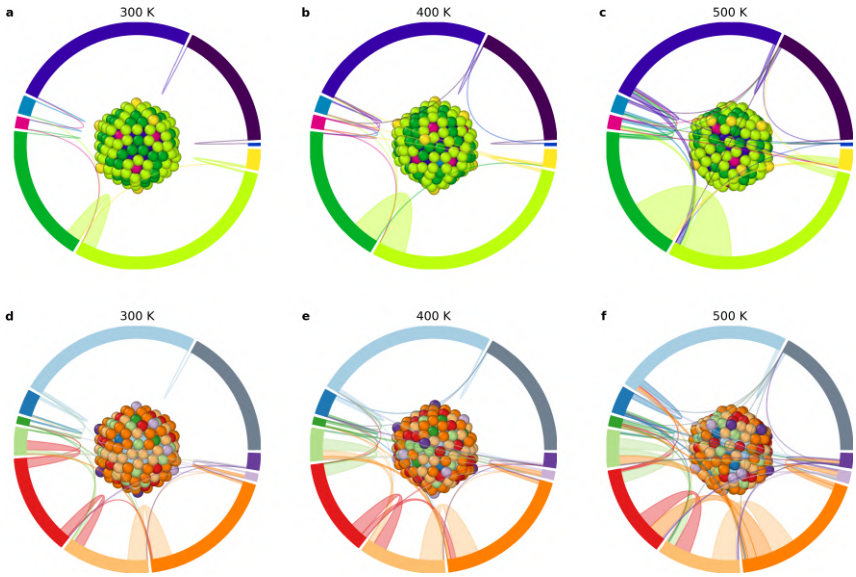


Fig. 4.9 **Chord diagrams for Ih_{309} at various temperatures.** A parallel of the net fluxes in the Ih_{309} at various temperatures, for the BU and the TD analyses. The two analyses are not completely overlapping due to a missing one-to-one relationship between the AEs found in the BU analysis and the AE selected in the TD analysis. *Reproduced with permission from Ref* [171].

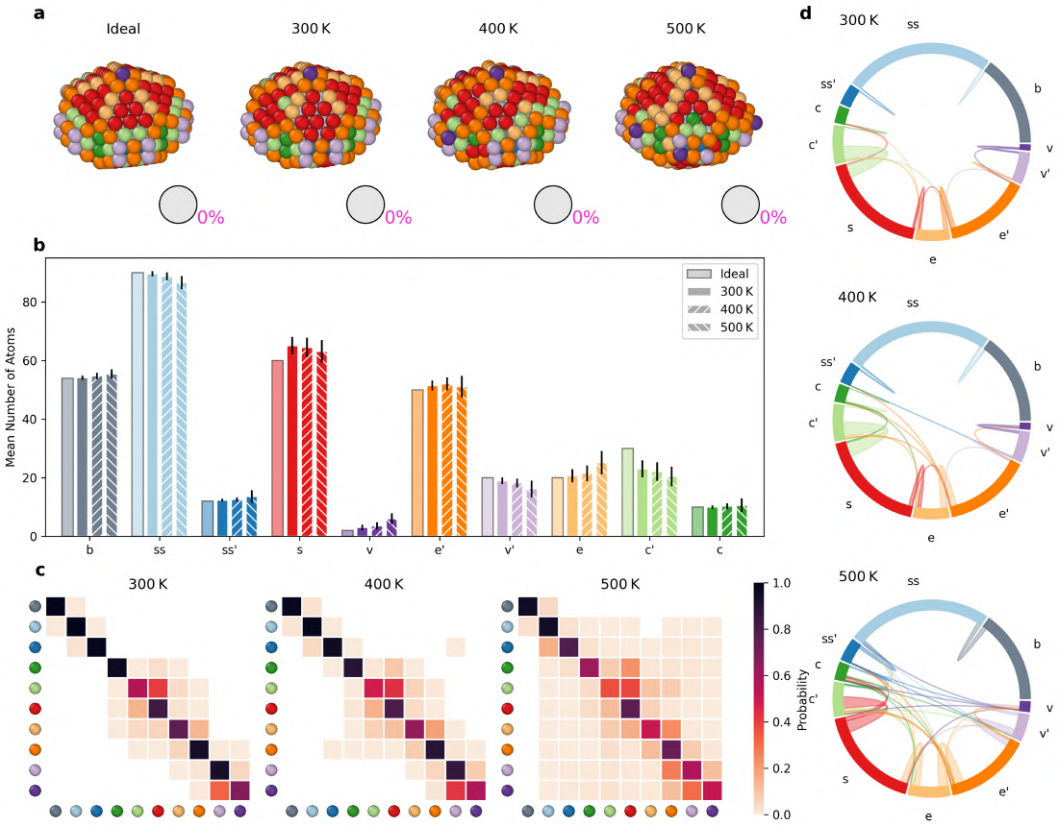


Fig. 4.10 ***Top-down* analysis of AEs in the Dh_{348} NP.** (**a**) Snapshots of the ideal Dh_{348} (0 K) and at 300 K, 400 K and 500 K. Atoms within the NPs are colored based on the AEs dictionary. (**b**) Histogram counting the average number of atoms in each AE during the last 1 μs of MD simulation at 300 K, 400 K, 500 K (second-left to right columns for each AE). Standard deviations as vertical black lines. (**c**) Normalized transition matrices reporting the probabilities for atoms in the Dh_{348} NP to remain in a given AE (p_{ii}) or to exchange into another one ($p_{i\rightarrow j}$) in the time interval of $dt = 1$ ns at the various temperatures. (**d**) The chord diagrams show the dynamic interconnections between all AEs detected in the NP at different temperatures. *Reproduced with permission from Ref* [171].

Building on this, Fig.4.11 explores how different cutoff distances affect the transition probabilities and net fluxes for the Ih_{309} nanoparticle at various temperatures. The transition matrices, which display the grouped AEs from the 0.08 cutoff used in the main analysis, illustrate how varying cutoff values can significantly impact the interpretation of atomic transitions and dynamic behavior. This analysis highlights the critical role of cutoff selection in effectively capturing the nuances of atomic movements and interactions, thereby providing a more accurate depiction of the system's dynamic behavior. These detailed analyses clearly highlight how the different atomic environments within the Ih_{309} nanoparticles significantly influence their structural

stability and dynamic behavior as a function of temperature. The energy data show that, although the Best309 configuration is the most stable, the Ih_{309} configuration with three rosettes also maintains significant stability under certain conditions. The differences in net fluxes between the *bottom-up* and *top-down* analyses reveal the intrinsic complexity of atomic dynamics, underscoring the necessity of multiple approaches for a comprehensive understanding.

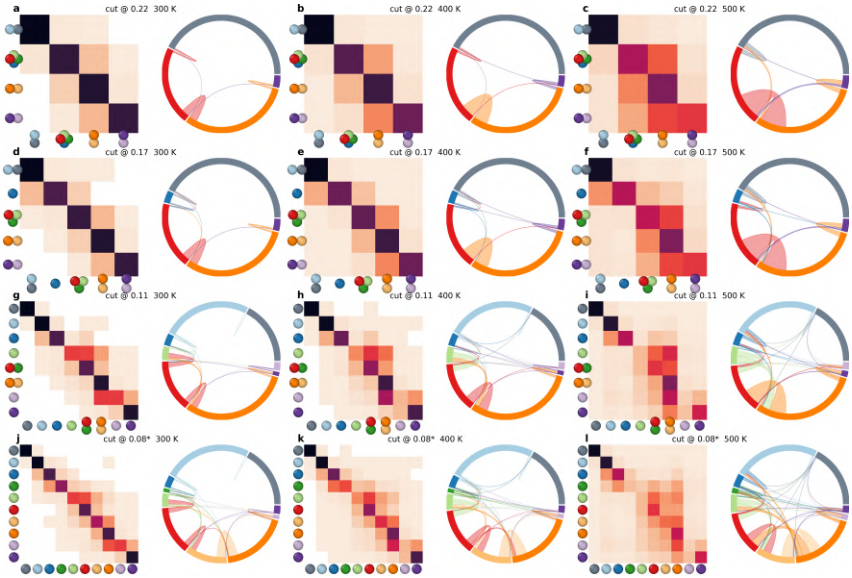


Fig. 4.11 The influence of the number of AEs due to the chosen cut (0.22, 0.17, 0.11 and the used 0.08) on the transition probability and on the net flux for the Ih_{309} at various temperatures. In the transition matrices, we show the grouped AEs from the cut=0.08 that we used in the main analysis. *Reproduced with permission from Ref* [171].

Furthermore, the impact of varying cutoff distances on the clustering of atomic environments emphasizes an important trade-off between detail and clarity in the analysis. Accurate selection of the cutoff is crucial for capturing the atomic transitions and interactions within the nanoparticle, allowing for a more precise representation of the system's dynamic behavior. This level of detail in atomic transitions and fluxes at different temperatures provides valuable insights into the mechanisms governing stability and structural transformations in gold nanoparticles.

4.3 Conclusions

Understanding the intricate atomic dynamics on the surface of metal NPs in relevant regimes is paramount for unveiling the physical chemistry and diverse properties of the nanoparticles. This characterization, while fundamental, typically poses significant challenges, both experimentally and computationally. In this work, maintaing the same approach described in Chapter 3, we achieve such a detailed characterization *via* the use of a concerted ML approach that includes a combination of a *bottom-up* and *top-down* data-driven analyses of steady-state MD trajectories of various types of Au NPs. In the first step, the *bottom-up* ML analysis detects in a purely data-driven way the main atomic environments that populate an NP in MD steady state based on the levels of order/disorder and the structural similarity between them as captured by high-dimensional SOAP data extracted from the MD trajectories, and further dimensionality reduction and unsupervised clustering (Figures 4.1 and 4.2). The choice of the SOAP descriptor, having a defined distance metric, allows us to perform advanced analysis techniques, *i.e.* density-based clustering, to better understand and classify the atomic environments at the surface. Tracking the individual atoms along the MD and classifying them based on the detected SOAP environments allows resolving the complex atomic dynamic that is present on the NP surface at different temperatures (see transition matrices in Fig.4.1 and in Fig.4.2). In particular, this allowed us to identify in a data-driven unbiased way all the AEs that populate the NPs in the MD local equilibrium and to understand where these come from and where they go in terms of interconversion into other AEs. On a second step, a *top-down* data-driven classification based on the similarity/distance between/from the SOAP spectra of each atom at each sampled step of the steady state MD trajectories and the SOAP spectra characteristic of a variety of different Au NPs contained in a SOAP spectra dictionary. Such a SOAP dictionary contains the SOAP spectra of all AEs proper of ideal, *e.g.*, icosahedral, decahedral, and truncated-octahedral Au NPs, which allows us to learn which ones of the AEs appearing on the surface of a given simulated NP in MD steady state are "native" of that type of NP, and which ones are "non-native" (*i.e.*, typical of ideal NPs of a different shape). We repeat such analysis on three types of different-shape Au NPs (Figures 4.5, 4.7 and 4.10: *i.e.*, Ih_{309}, Dh_{348} and To_{309}). Notably, such analysis allows us not only to estimate the "*statistical identity*"[172] of the various NPs – namely, what do the NPs look-like in relevant dynamic regimes (histograms of Figures 4.5, 4.7

and 4.10) –, but also obtain relevant dynamic information as on the average lifetime and interconnection rates of all visited atomic (native and non-native) environments present on their surface (transition matrices in Figures 4.5, 4.7 and 4.10). While the dictionary of AEs presented herein is based on the analyzed set of Au NPs, it can be easily expanded to include more particle geometries. Nevertheless, the provided analysis already shows the generality of such an approach, *i.e.*, it can be applied to any metal NP. Moreover our analysis can be easily applied without further tuning to metallic systems in which vacancies arise, *i.e.*, on that case the user would just need to add some new environments to the AE dictionary.

In conclusion, this study highlights a paradigm shift from static structural analyses to dynamic evaluations, a transition that is even more crucial for NPs than for surfaces due to the dominance of surface dynamics, even at very low temperatures. By focusing on these dynamics, we can determine energy barriers, reconstruct the free energy landscape, and establish probability distributions and lifetimes of various atomic states. In contrast, approaches that rely solely on the internal thermodynamics of the system are unable to resolve these dynamic aspects and thus cannot provide the same detailed insights as our methods based on reconstructing the system's behavior from its dynamics.[172, 342, 234, 205]

This knowledge can optimize the performance of Au NPs in applications that depend on surface properties and features, such as catalysis, sensor devices, and biomedical applications. We anticipate that the methodologies presented will find broad applications in other metallic NP systems, opening new avenues for discovering structural-dynamic-property relationships across a variety of similar metal NP systems. This approach enhances our understanding of metallic NPs' behaviors and properties, revealing complexities that traditional static analyses fail to capture.

4.4 Methods

4.4.1 Atomistic models and MD simulations of the NPs

The atomistic models for the Ih_{309}, Dh_{348}, and To_{309} NPs were built with the tool "clusterCreator".[343] Preliminary basin hopping calculations showed that, at these sizes, Au favors the formation of decahedral NPs, followed by the icosahedron and the cuboctahedron. To simulate the NPs, we used the SMATB [227, 271, 339]

potential available in LAMMPS [344] .[334] (see dedicated section on FF of Chapter 2 for insights about this potential) The NP models were initially minimized using the built-in command in LAMMPS (set up with etol = 10^{-6} ftol = 10^{-8}, maxiter = 1000 and maxeval=10000), then we performed a small thermalization of 20000 MD steps with the timestep set to 1 fs on the NP with the velocities initialized to the desired temperature and with the thermostat with the same settings of the main simulation. We then simulated different Au NPs at temperatures of 300 K, 400 K, and 500 K. All MD simulations were conducted in the canonical ensemble using the LAMMPS's Langevin thermostat, using a timestep of 5 fs, and a damping parameter for the Langevin thermostat set to 100 ps. We simulated each NP system for a total of 2 μs of MD. During the simulations, all NP systems reached a steady state in the MD regime (equilibrium)[171]. All our analyses were thus conducted on 1000 frames taken every 1 ns along the last 1 μs of each MD simulation.

4.4.2 SOAP analysis

The SOAP spectra of each atom in the NPs (Fig.4.1**a**) were calculated at each of the 1000 MD snapshots taken from the last 1 μs of the simulations (every 1 ns). We thus come out with SOAP datasets containing a total of 309'000, 348'000, or 309'000 SOAP spectra for Ih_{309}, Dh_{348}, and To_{309} simulated systems respectively, at each temperature. We used *dscribe* [345] to generate the SOAP vectors with the following parameters: $r_{cut} = \sim 4.48$ Å (corresponding to 110 % of the Au FCC lattice parameter, which includes in the calculation up to the first two neighbors in FCC, and up to the third in the HCP case even in case of some small local fluctuations. This specific value was determined through various tests and trials, and comparisons with other descriptors, representing the best compromise between computational cost and retained information, ensuring a comprehensive and precise representation in our analysis. We set up the l_{max} parameters for the spherical harmonics to 8, and the n_{max} parameter to set up the number of radial basis functions to use to 8. With these parameters, the SOAP spectrum for each atom is a vector of 576 components (of which 324 are unique).

4.4.3 *Top–down* analysis: similarity, distance and dictionary

To apply the SOAP classification we built a dictionary, with the same approach described in the previous chapter: we attempted to create the most complete dictionary for icosahedral, decahedral, and octahedral NPs' AEs by choosing the most different environments from various minimized Au NPs.

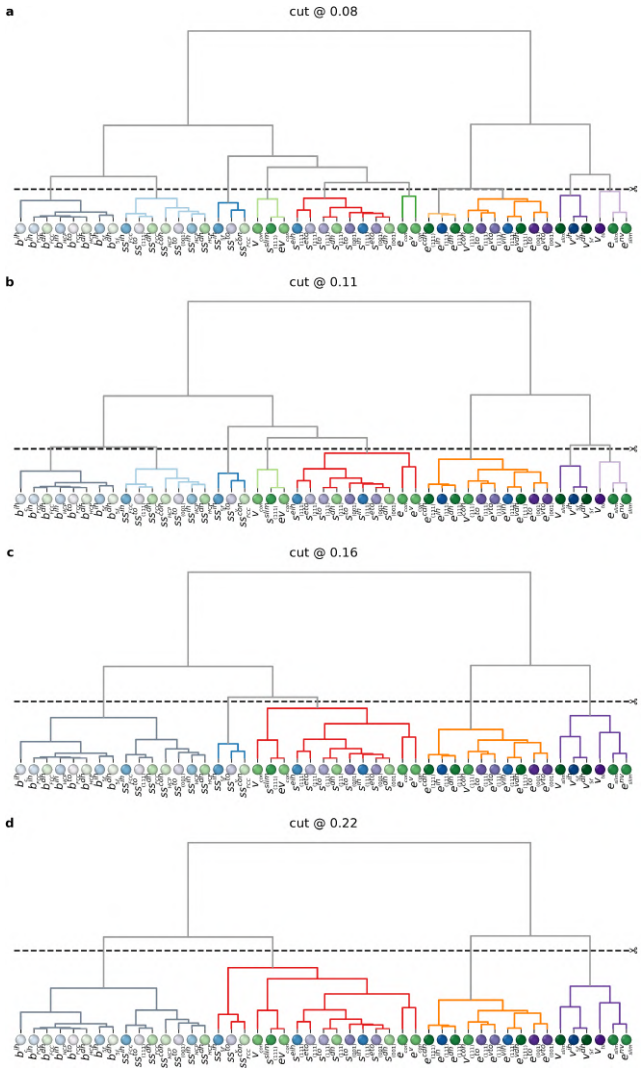


Fig. 4.12 The influence of the chosen cut (0.22, 0.17, 0.11 and the used 0.08) on the dendrogram of Fig.4.4: by increasing the cut distance we reduce the number of clusters, and so we lose details about the geometry of the AEs. *Reproduced with permission from Ref* [171].

To enrich our dictionary, together with the NPs that we effectively simulated in this work, we also included larger size NPs possessing a higher variety of AEs

in their ideal state. We obtained a dictionary of 47 elements. To simplify its usage, we hierarchically classified its elements using the hierarchical clustering algorithms implemented in scipy[346]. First of all, we used the d_{soap} (see previous chapter for the mathematical derivation) for calculating the distance between each of the environments belonging to the dictionary. Then we created a binary tree that represents this classification by using the "*complete*" algorithm for hierarchical clustering, which at each step couples the closest elements in the set and assigns to the newly formed couple the largest distance from each remaining element of the set, and uses the new distance in the next steps until it has completed the classification. We represent this tree in the dendrogram in Fig.4.4**e**, where we show that we have chosen to apply a cut at the distance of 0.08 [*dSOAP*].

Fig.4.12 elucidates the impact of different cutoff distances (0.22, 0.17, 0.11, and 0.08) on the dendrogram clustering of AEs . As the cutoff distance increases, there is a notable reduction in the number of clusters, which consequently leads to a loss of detailed geometric information about the AEs. This figure accentuates the trade-off between detail and clarity in AE analysis, emphasizing the importance of selecting appropriate cutoffs to balance comprehensive insight with manageable complexity. The careful selection of cutoff distances is thus crucial for accurately capturing the intricacies of atomic arrangements and transitions within the nanoparticle. This cut at 0.08 [*dSOAP*] leads to the creation of 10 different groups of dictionary entries (that can be seen more clearly in Fig.4.5**a**) with similar geometrical characteristics, from the original 47 AEs. During the MD simulations analysis, we assigned an environment to one of these 10 clusters in two steps. The first step is to classify it as one of the 47 elements of the original environment dictionary: we do this simply by assigning it to the closest element of the dictionary in terms of the SOAP distance. The second step is to classify our analyzed environment by assigning it to the cluster to which its closest reference belongs.

Chapter 5

Sampling real-time atomic dynamics in metal nanoparticles by combining experiments, simulations, and machine learning

This work has been carried out in collaboration with the experimental groups of Prof. Sara Bals and Prof. Sandra Van Aert at the University of Antwerp (Belgium). Our study emphasizes the computational results while also briefly summarizing the experimental methods, reported in subsection 5.4.2 and 5.4.3 of the *Methods* sections. For more comprehensive details on the experimental techniques, please refer to the paper titled "*3D Atomic Structure of Supported Metallic Nanoparticles Estimated from 2D ADF STEM Images: A Combination of Atom-Counting and a Local Minima Search Algorithm* "[331].

The work presented in this section marks a transition from theoretical investigations to a more complete understanding that integrates both experimental and computational approaches. A critical aspect of computational simulations is the challenge of sampling and the risk of becoming trapped in local energy minima. MD simulations, while powerful for revealing atomic-level dynamics, typically start from predefined structural models of the systems under investigation. These initial structures are often idealized and might not fully embody the complexity and heterogeneity of NPs synthesized under experimentally relevant conditions, which can feature a range of defects, dislocations, and surface irregularities, as we described

in previous chapters. This discrepancy between the starting configurations used in simulations and the actual structures of NPs in experimental settings may introduce inaccuracies, specifically in the simulations' ability to accurately reproduce NPs' equilibrium dynamics. The main issue lies in the sampling methodology inherent in classical MD simulations, which explore the potential energy landscape of the system to predict its behavior over time. However, if the initial structure is an idealized version, the simulation may only explore a limited portion of the potential energy landscape, thereby increasing the risk of the system becoming trapped in a local energy minimum, especially at relatively low temperatures that correspond to the experimental conditions (300-700 K). These local minima can significantly differ from the true equilibrium state (the global minimum) of the NP under experimentally relevant conditions. Consequently, this leads to results that are heavily dependent on the initial structure, providing only an indicative, rather than accurate, representation of the NP's properties and behavior. For this reason, to ensure reliable predictions, it is crucial to use realistic initial configurations that embody the actual physical state of the NPs. This approach helps the simulations capture true dynamics and interactions, making the insights more applicable to real-world conditions. Indeed, in this context, computational simulations also offer several significant advantages. They allow for direct tracking of atomic movements over time and provide continuous and detailed insights into microscopic dynamic processes, capturing fast dynamics that experimental methods might miss. Additionally, they can explore a wide range of conditions and variables, including different temperatures, pressures, and chemical environments.[171, 54, 347–351]

In contrast, state-of-the-art experimental approaches can capture the actual physical state of NPs under real-world conditions, reflecting true environmental conditions, including defects, dislocations, and surface irregularities[331]. This real-world applicability ensures that the results are relevant and accurate under practical conditions. Furthermore, experimental data provide a fundamental check for computational models, helping to refine and validate simulations, thus ensuring that theoretical predictions align with observed behaviors. Recent advancements, such as annular dark-field scanning transmission electron microscopy (ADF-STEM), have enabled the accurate reconstruction of the 3D atomistic structures of NPs from microscopy images taken at temperatures relevant to various applications. However, experimental approaches also have limitations. A major issue is the discrepancy between the frequency of image acquisition and the timescales of NP dynamics. Typical

ADF-STEM setups capture snapshots of the nanoparticle structure at intervals of approximately 0.1 to 1 second[331]. However, the real atomic dynamics often occur on much shorter timescales, typically in the *ps* to *ns* range. This limited time resolution hinders the ability to track individual atomic movements accurately, which is crucial for our analysis and essential for studying the dynamics of the system with the high resolution needed to resolve the specific behaviors we aim to understand.

Integrating experimental data into computational analyses is essential to overcome the limitations of each method when used independently[331, 352–355]. By grounding simulations in experimentally derived structures, we can enhance their accuracy. This strategy enables a more precise exploration of the potential energy landscape, ensuring that simulations reflect the true dynamics of NPs as they exist in real-world applications.

Recognizing these challenges, this chapter introduces an innovative, integrated experimental-computational strategy designed to resolve the atomistic dynamics of metal NPs under conditions that represent real-world scenarios. This approach enables a detailed exploration of the NPs' dynamic properties over timescales and conditions directly relevant to their applications. By combining the strengths of both methods, we can achieve a more comprehensive and realistic understanding of NP dynamics and properties, ultimately leading to more accurate and applicable insights.

Full bibliographic reference: **Cioni, M.**, Delle Piane, M., Polino, D., Rapetti, D., Crippa, M., Irmak, E. A., Van Aert, S., Bals, S., & Pavan, G. M.(2024). "Sampling Real-Time Atomic Dynamics in Metal Nanoparticles by Combining Experiments, Simulations, and Machine Learning." *Advanced Science*, 2307261. Wiley Online Library. [1]

5.1 Introduction

As described in Chapter 1 and shown in the previous Chapter 4, the high surface mobility of Au NPs is a characteristic feature that plays a critical role in determining their unique properties [327, 171, 356]. This is particularly important in small NPs, where the high surface-to-volume ratio results in a significant proportion of atoms

[1] My contribution to this article, as the first author, involved collaborating with the experimental team, performing the simulations and analyses, contributing to the interpretation of the results, and contributing to the writing of the manuscript.

residing on the NP surface, displaying greater mobility compared to those in the bulk even at fairly low temperature.[330, 357, 320, 358] The dynamic atomic rearrangements occurring on the NP surface significantly influence the optical, electronic, and catalytic properties.[359, 321, 327, 360, 361] However, the high atomic mobility of NPs introduces substantial challenges for both experimental and theoretical investigations into NP structures. Capturing both the static structure and real-time atomic dynamics of NPs is crucial to fully grasp their behavior and effectively control their properties for different applications. Recent experimental advancements, as described in Chapter 1, such as annular dark-field scanning transmission electron microscopy (ADF-STEM) have allowed reconstruction of the atomistic structure of NPs from microscopy images taken at relevant temperatures.[330, 322, 331] However, the time resolution of these techniques is insufficient to capture the rapid atomic dynamics, which typically unfold at much shorter timescales (ps to ns).[171] Techniques like ultrafast electron diffraction (UED) and high-resolution transmission electron microscopy (HRTEM) provide enhanced temporal resolution, but HAADF-STEM stands out by providing indispensable three-dimensional structural information.[331, 362] Combining HAADF-STEM with MD simulations provides detailed spatial and 3D structural insights along with higher temporal resolution, offering a comprehensive approach to understand NP dynamics.[171, 247, 351] Advanced structural and dynamical descriptors, such as smooth overlap of atomic positions (SOAP)[202], local environments and neighbors shuffling (LENS)[203], and TimeSOAP[363], enhanced by ML, allow reconstruction of atomic environments and dynamics within NPs. It should be noted, however, that such analyses are obtained from MD trajectories acquired from simulations that start from ideal nanoparticle structures, which may differ from the structure typical of the same NPs under experimentally relevant conditions. This makes it difficult to guarantee that the extracted data provide a reliable reconstruction of the equilibrium dynamics of these NPs since one of the main limitations of classical MD simulations lies in the sampling and in the risk of entrapment in local energy minima. In this work, we demonstrate the potential of combining state-of-the-art experimental and computational approaches to overcome these limitations. By integrating high-resolution ADF-STEM images with MD simulations, we capture the real-time dynamics of Au NPs. This method provides a comprehensive analysis of NP dynamics, closely reflecting real-world conditions. This approach begins with the acquisition of high-resolution ADF-STEM images, enabling precise 3D reconstructions of the NP.[331] These reconstructed

structures serve as starting points for MD simulations, allowing us to study the dynamics of Au NPs at different temperatures. Our analysis shows that the dynamics reconstructed from MD simulations represent an equilibrium ensemble. This combined experimental/computational approach provides deep insights into the real atomic-scale dynamics of metal NPs and links these to their macroscopic properties.

5.2 Results

5.2.1 A combined experimental-computational approach

The research group of our collaborators S.Bals and S. Van Aert recently developed a novel approach that combines atom-counting and iterative local minima search algorithms to reconstruct the 3D structure of supported NPs from experimental single-view 2D ADF-STEM images (Figure 5.1**a**).[331] This technique has been applied for example in the demonstrative case study reported herein to Au NPs supported on CeO_2 at a temperature of 673 K. The methodology begins with ADF STEM imaging, essential for attaining atomic-level details of NPs. This step is enhanced by advanced atom-counting techniques employing statistical parameter estimation, which precisely determines the composition and density of the NP. The data obtained from these techniques lay the groundwork for the next phase, involving the integration of these findings into MD simulations to create a preliminary 3D model of the NP[331].

Further refinement of this initial model is achieved through a local minima search algorithm which adjusts atomic positions within the model and thoroughly evaluates the system's energy landscape. These careful adjustments are key in approximating the NP's real structure.

An innovative element of this approach is the incorporation of molecular dynamics structural relaxation at an experimental temperature of 673 K. This step is vital to ensure that the reconstructed models are not only theoretically accurate but also realistically represent the NPs' behavior under specific, experimentally relevant temperature conditions [364–368]; indeed, aligning the models with real-world scenarios is crucial for their practical applicability.

Validation of the reconstructed 3D structures is conducted through an extensive comparison with experimental data. This validation focuses on the precision of atom

positions and the overall morphology of the NPs, ensuring that the reconstructed models are not only theoretically reliable but also congruent with observed experimental behaviors. This step is crucial in confirming the practical viability and accuracy of the reconstructed models.

For more details on the 3D reconstruction process, see the Methods section (subsections 5.4.3 and 5.4.2). Additionally, for a more complete overview and detailed methodological insights, reference paper [331] offers extensive information.

In particular, this approach was used as a first step to generate a series of atom-counting maps from 10 snapshots taken every 0.6 *s* along a 6 *s* of ADF-STEM sampling(Figure 5.1**a**). This represents the intrinsic limitations of this experimental approach: a 0.6*s* is too long to track the motion of atoms between two consecutive frames. In Figure 5.1**b** we present snapshots of the obtained NPs, color-coded based on a scheme that corresponds to the number of atoms in each atomic column. As previously mentioned, the reconstructed 3D structures of the NPs correspond to 10 frames, captured over a total observation time of 6 *s* using ADF-STEM[331] (Figure 5.1**c**). Although the time intervals between the reconstructed NP structures are relatively long (0.6 seconds), they significantly contribute to our comprehension of atomic surface dynamics in environments similar to practical applications[365, 364, 366, 368]. This provides a profound insight into the dynamic activities on NP surfaces, reflecting conditions encountered in real-world experimental settings.

To extract information on the atomic environments emerging on the NP during the experimental data acquisition, we employed the same approach[171] based on SOAP power spectra[202] (Figure 5.1**d**), to analyze and better interpret what happens into the atomic structure of these real NP experimental snapshots. For this reason, in this chapter, the section concerning the choice of SOAP parameters will not be explained as it is the same as in the previous chapter.

The extracted SOAP spectra (2×10^6 for each of the 10 frames) of the Au atoms have been then classified based on a general AE dictionary[171] of SOAP AEs, the same as the previous Chapter 4 (see subsection 4.2.2). Such comprehensive AE dictionary, not explained in detail in this chapter, includes 47 SOAP spectra of all the AEs present in ideal Au NPs (at 0 K) of various sizes and morphologies at 0 K (Figure 5.1**d**). This provides us with an essential tool for monitoring the AEs that are present in the real NP under experimental conditions and classifying them based on their similarity to the AEs contained in the dictionary. Figure 5.1**e** displays

the 47 AEs defined in our SOAP environments dictionary, presented in a circular dendrogram arranged according to their similarity based on SOAP fingerprints. The dotted inner circle indicates that truncating the dendrogram, we consider only SOAP distance larger than 0.8 as relevant [171] (an excessive resolution increases the noise and would emphasize irrelevant differences). This choice simplifies the dictionary from 47 to 11 without sacrificing significant variations[171] (Figure 5.1**f**). In the adopted color scheme of the dendrogram of Figure 5.1**f**, the colors belonging to the purple palette refer to native AEs typical of ideal truncated octahedral NPs, while all the other colors identify AEs that are more similar to AEs proper of other ideal NPs' morphologies (e.g. icosahedral, decahedral, etc.)

Considering such "coarse-grained" dictionary, we colored the atoms in the ten experimentally reconstructed NP structures based on the similarity of their SOAP atomic environments to those within our SOAP dictionary (Figure 5.1**g**). This first analysis reveals the dynamic nature of the atomic environments within the NPs in experimental conditions over the 6-second data acquisition, during which the atoms of this octahedral Au NP move, and non-native AEs (colored in yellow, green, and red) emerge on the NP surface. Far from being static, these NP structures show remarkable variability driven by thermal effects. The positions of vertices and edges within the NP displayed substantial shifts, underscoring the ongoing structural transformations. Our method thus provides a unique view of the atomic-level dynamics of NP, using Machine Learning (ML) analysis to illuminate structural evolution from experimental static frames.

Such atomic-level ML analysis underlines that understanding the atomic dynamics present in these NPs is key to understand their properties in experimentally-relevant conditions. While these experimental snapshots of the NP structures provide initial important evidence, they are spaced by time intervals of 0.6 *s* (Figure 5.2**a**). This is a substantial time gap, especially considering that atomic dynamics typically unfold on much faster timescales, such as pico- and nano-seconds. This temporal mismatch makes it impossible to reconstruct the atomic dynamics directly from such ADF-STEM reconstructed snapshots, as, e.g., it is not possible to attribute an identity to the individual atoms nor to monitor their movements from one snapshot to the subsequent one.

To address this issue, we utilized the high spatiotemporal sampling resolution provided by atomistic MD simulations and the advantage that MD can keep track of the atomic ID over time allowing for more detailed tracking of atomic movements.

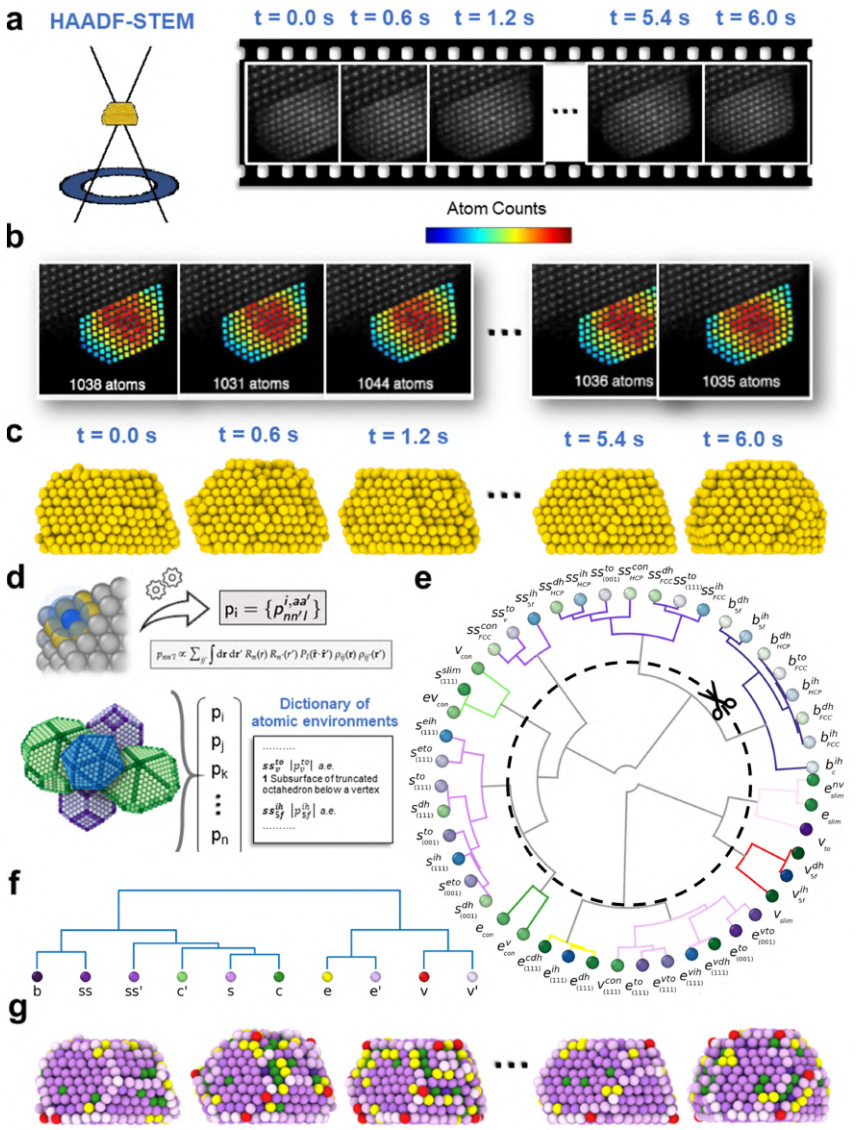


Fig. 5.1 **Stepwise process of applying the SOAP analysis to the experimental structures of Au NPs.** (**a**) Left: Schematic representation of the High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Right: Ten consecutive ADF-STEM frames of observed Au NPs at 673 K. (**b**) Atom-counting maps corresponding to the ten frames from (a), with the color indicating the atom count per column. (**c**) Final reconstructed 3D structures of the observed NPs. (**d**) Top: Each atom in an Au NP (depicted in blue) is assigned a SOAP vector, with the cutoff radius shown as a transparent sphere ($r_{cut} =\sim$ 4.48 Å, corresponding to 110 % of the Au FCC lattice parameter). Bottom: Construction of a SOAP dictionary of atomic environments (AEs) using icosahedral (blue), decahedral (green), and truncated-octahedral (purple) Au NPs, which also includes larger NPs for comprehensive AE representation. (**e**) A global dendrogram connecting AEs from different NPs. This dendrogram visualizes hierarchical clustering of AEs based on SOAP distances, with branches connecting the different clusters. By cutting at a SOAP distance threshold of 0.08 we form a coarse-grained dictionary, reported in (**f**). (**g**) Final 3D structures, with atoms colored according to the SOAP classifications from the AEs dictionary. *Reproduced with permission from Ref* [342]

5.2.2 Combining atomistic-scale MD simulations and experimental level sampling

We started 10 independent MD simulations using the 10 experimentally reconstructed configurations depicted in fig. 5.2**a**, with each starting structure containing a varying number of atoms ranging from 1031 to 1044. These simulations were conducted at temperatures of 300 K and 673 K (consistent with experimental conditions[365, 364, 366, 368]), enabling us to obtain trajectories from which it is possible to track the movements of the individual atoms in the NP and between the AEs that emerge within them; in these simulations, we employed the SMATB potential [227, 271, 339], which has been demonstrated to accurately describe the dynamics of gold NPs[171] as shown in previous Chapter. It is important to note that we did not include the substrate in our simulations. For more detailed information on our simulation methodology, please refer to subsection 5.4.1 of the Methods section. Each MD simulation lasts for 2 μs, and we collected 1000 frames (every 1 ns in MD) from the final 1 μs of the trajectory. This allows for a sampling frequency of 1 ns^{-1}, which is a sufficient sampling frequency to track the atomic motions of this system in 300-673K conditions. During this period, we computed SOAP spectra for all the atoms in the NP. It's important to mention that the chosen time window for our analysis ensures that any observed communication or exchange among the AEs pertains to processes occurring on the nanosecond timescale or slower. This effectively reduces the likelihood that the AE exchanges are influenced by thermal vibrations. This led to $\sim 10^6$ SOAP spectra for each MD, comprehensive for 10^7 SOAP spectra for all the 10 MD simulations. As demonstrated in the previous Chapter (see subsection 4.4.2 of Chapter 4), we utilized a cutoff radius of 4.48 Å, corresponding to 110% of the lattice pair distance of gold, which includes in the calculation up to the first two neighbors[171]. The selection of the cutoff is a critical decision[369], and for this reason, this specific value was chosen to achieve a balance between computational efficiency and the fidelity of information retained, ensuring a thorough and accurate representation in our analysis[171]. Differently from the previous analysis on the static experimental frames, these MD simulations permit us to track the SOAP AEs to which each atom of the NP belongs to over time. This allows us, e.g. to quantify the propensity of each atom to remain in a certain AE or to undergo a transition to a different one at every δt ($1ns$). By delving deeper into the dynamics of these atomic environments, we were able to quantify the NP's

stability and analyze the dynamics of exchange between its various constitutive AEs. We generated histograms from the final 1 μs of the MD simulations, which provide an average atom count associated with the population of each AE (Figure 5.2**c**). Comparing the histograms obtained, e.g., at 673 K vs those of the starting NP configurations, provides information on the NP stability and it is interesting to note how the AEs histograms calculated from the MD at 673 K do not deviate much from those obtained from the experimental structures, despite the considerable dynamics observed along the MD.

Figure 5.2**b**-**c** show that during the various MD runs, dynamic atomic rearrangements can be observed; vertical arrows in the histograms indicate AEs that are not present in the starting frame, but that may emerge with temperature: e.g. red "v" AE. Furthermore, Figure 5.2**b**-**c** (left to right) show that, while some variability between the MD runs can be expected, the histograms do not change much in the various systems configurations. This implies that these MD simulations offer a detailed view of the atomic dynamics within these NPs. By observing them at an atomistic resolution over microsecond-long time windows, we can track their equilibrium trajectories under conditions that are relevant to real-world experiments. This allows concatenating the various 1μs-long analyzed MD trajectories, obtaining 10 μs of sampling of the equilibrium MD of the Au NP.

We repeated the same analysis, as a control case, by running the 10 MD simulations at a lower temperature (300 K). In this case, the analysis shows, as expected, less dynamic activity in the NPs compared to 673 K. However, the overall conclusions remain the same [342]

The next step involves a statistical analysis of the SOAP data, which allows us to quantify the dynamics of the NP. Figure 5.3**a** illustrates the initial structures taken at 0, 2.4, and 6 s, and the corresponding structures obtained after 2μs of MD at 673 K (color code according to the dictionary of Figure 5.1) The NP surface maintains its structural integrity throughout the simulation, suggesting that the truncated octahedral structure remains relatively stable overall, despite the high atomic mobility. To characterize the intricate atomic dynamics present in the NPs, we calculated the transition probabilities for atoms between these AEs.

The transition matrices in Figure 5.3**b** indicate the probability of an atom, with a specific AE at time t_i, remaining in the same AE (diagonal entries) or transitioning into a different AE (off-diagonal entries) within the analysis time interval (δt).

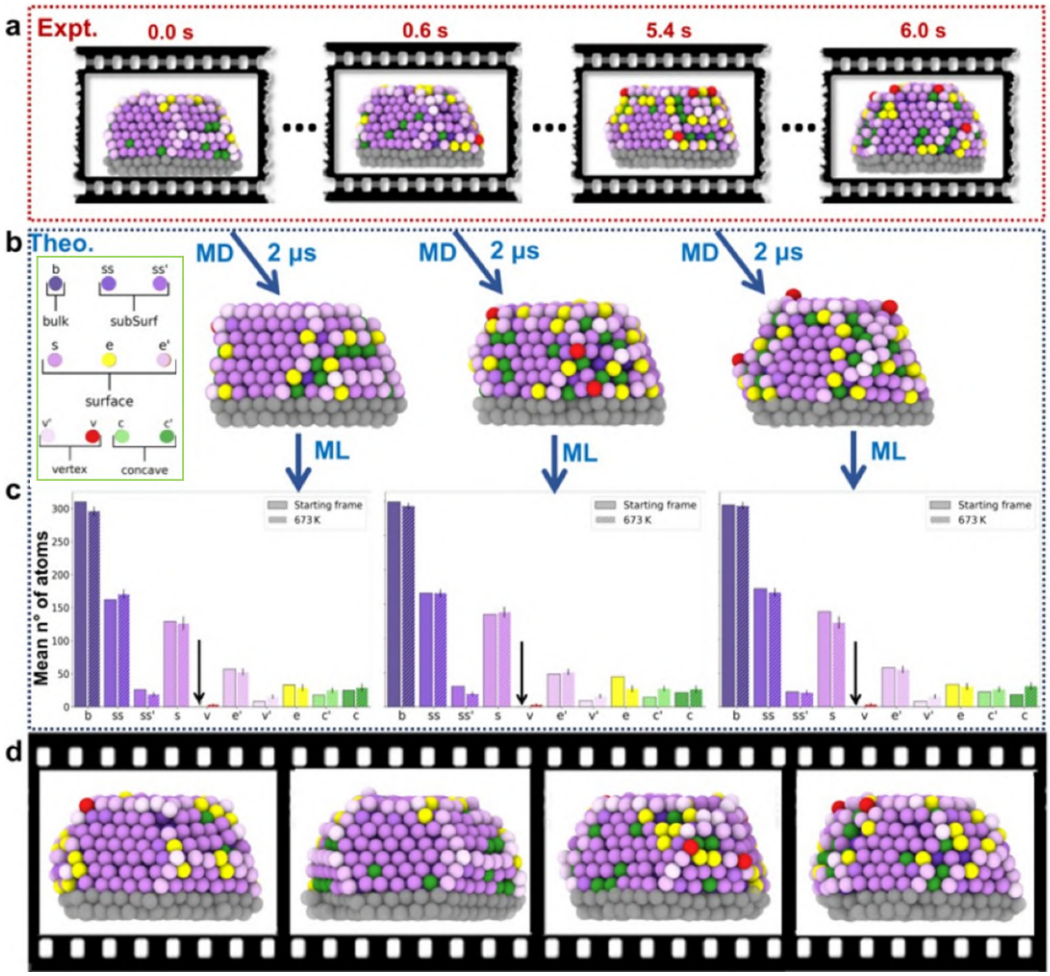


Fig. 5.2 **Schematic representation of the combined approach for capturing atomic-scale dynamics in Au NPs.** (**a**) Experimental frames showing the time resolution of 0.6 s, highlighting the temporal gaps between the snapshots. (**b**) Final frames after 2 μs MD simulations, with atoms colored according to the dictionary developed from the SOAP spectra analysis, effectively filling the temporal gaps between experimental snapshots. (**c**) Histograms displaying the average count of atoms associated with each atomic environment, providing a measure of system stability. Standard deviations as vertical black lines. (**d**) Conceptual representation of the full 6-second movie reconstruction, with gaps filled by MD simulations. *Reproduced with permission from Ref* [342].

At 673 K, all cells in the transition matrix are colored, indicating non-zero transition probabilities across all AEs within the NP. This observation indicates a marked increase in atomic mobility and interchange between various NP regions at higher temperatures [171, 172, 248, 204].

A visual representation of these dynamic exchanges is well rendered by the use of chord diagrams of Figure 5.3**c**, where the size of the arc sections is proportional to the population of the various AE, while the width of the chords corresponds to the intensity of atomic exchanges between them. At 673K, these analyses show high communication and dynamic exchange in the NP. This does not pertain to surface AEs only but this atomic exchange can also be observed between the innermost NP bulk (*b*) and the least coordinated surface AEs, e.g. vertexes AEs (*v* and *v*′), or between sub-surface and surface AEs (*ss*→*v* and *ss*→*e'*). On the other hand, at 300K, dynamic exchanges are predominantly observed among AEs with similar coordination numbers, such as *v*→*v'*, *b*→*ss*, indicating limited atomic mobility and constrained transitions. Despite the striking internal atomic dynamics observed at 673 K, in such conditions, the NP still preserves its truncated octahedral structure (Figure 5.3); it is interesting to note that these analyses performed on the MD trajectories starting from the experimental configurations captured at 0 *s*, 0.6 *s*,...,6.0 *s* provide very similar results (Figure 5.3*b-c* left-to-right). This demonstrates that these MD provide reliable pictures of the internal atomic dynamics present in these NPs in equilibrium regimes and in experimental relevant conditions. Furthermore, this also allows us to concatenate all data in a unique dataset, useful to improve the statistical confidence of our analysis.

5.2.3 Reconstructing the realistic atomic dynamics in Au NPs in experimentally-relevant conditions

Merging together the results obtained from the 10 (independent) MD trajectories, we could obtain an average and comprehensive picture of the dynamics of Ceria-supported Au NPs. Figure 5.4 averages the data derived from our MD trajectories providing meaningful insight into the atomic ensemble's dynamics and transitions present in these NPs over the entire experimental sampling timescale. Figure 5.4**a** and Figure 5.4**c** provide an equilibrium representation of the AEs dynamics at 300K and 673K respectively. The average histograms, similar to those in Figure 5.3, confirm our earlier observations of the AE distribution, showing that the NP's dynamics at thermal equilibrium maintain similar characteristics over the full experimental time-scale. The chord diagrams and transition matrices, displayed in Figure 5.4**a** (middle and right) and Figure 5.4**c** (middle and right), capture at atomistic resolution the average AEs transition probabilities that characterize such NP in experimental

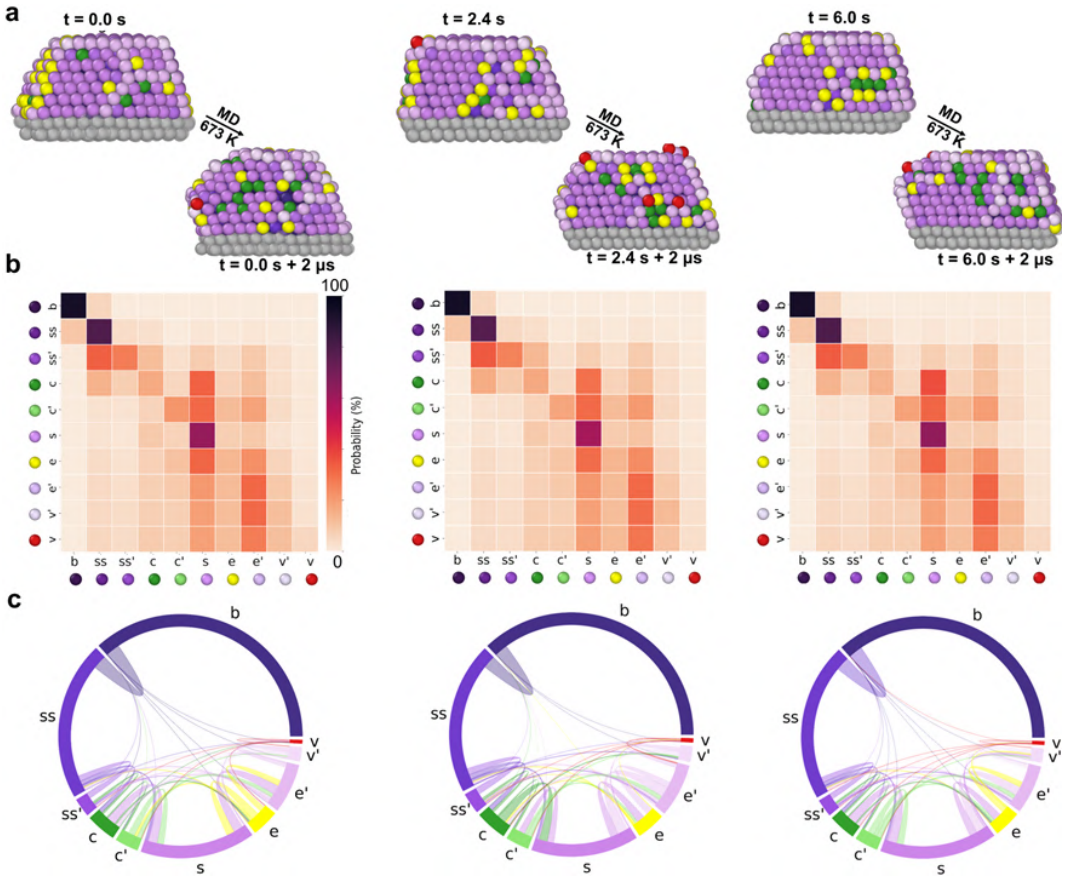


Fig. 5.3 **Quantitative analysis of atomic transitions and stability in reconstructed Au NPs at different temperatures.** (**a**) Initial structures at time points 0.0 s, 2.4 s, and 6.0 s, and the corresponding final structures post-MD simulations at 673K, color-coded according to the SOAP dictionary. (**b**) Histograms representing the average number of atoms per AE during the final microsecond of MD simulations. The leftmost column for each AE indicates the initial population, with standard deviations represented by vertical black lines. The absence of certain AEs in the initial structures is denoted by arrows. (**c**) Normalized transition matrices demonstrating the probability of atoms remaining in a given AEi (p_{ii}) or transitioning to a different AEj ($p_{i\to j}$) within a time interval of $\delta t = 1$ ns. *Reproduced with permission from Ref* [342].

conditions and in relevant observation timescales, obtained from the 10 individual $1\mu s$ MD windows. These average diagrams reveal similar patterns to the ones observed in (Figure 5.3), taken along $6s$ of global experimental samplings, providing a statistically robust perspective of the equilibrium atomic structure and dynamics of these NPs. The details of panels Figure 5.4**b** and Figure 5.4**d**, show how, selecting one specific AE (e.g. s), from the matrices one can obtain specific transition probability of the dynamic interconnections with the other ones in the NP.

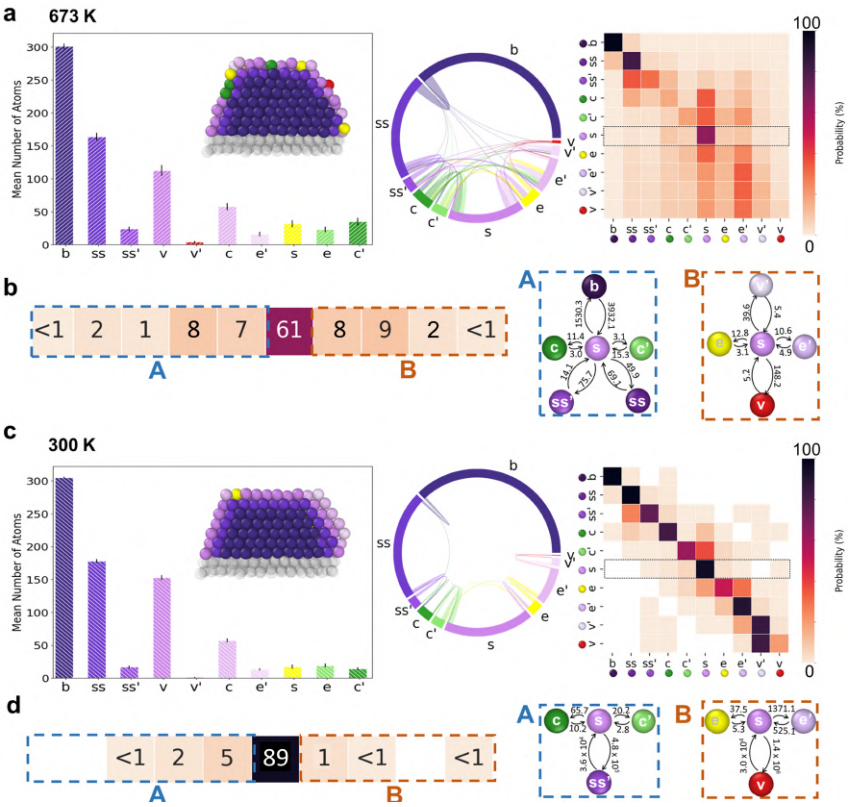


Fig. 5.4 **Detailed analysis of atomic transitions and mobility in Au NPs across the experimental time-scale.** (**a**) Equilibrium representation of atomic environments at 673K, featuring histograms similar to Figure 5.3, confirming AE distribution. The inset shows the equilibrium 3D structure of the NP, cut to show the interior. In the middle and right panels, chord diagrams and transition matrices encapsulate the aggregated fluxes and transition probabilities for the combined trajectories, representing atomic behaviors over the full experimental time. (**b**) Focused examination of specific transition probabilities from the surface (s) AEs at 673K, demonstrating the likelihood of atoms on the NP's flat faces transitioning to other AEs. The rightmost panels display the characteristic times (in ns) of transitions related to surface AEs. (**c**) and (**d**) report the same results, but at 300K. *Reproduced with permission from Ref* [342].

From transition probabilities, between AEi and AEj ($p_{i\to j}$) one can obtain information on the average lifetimes of different AEs within the NP and on the average transition rates for the atomic exchange between them. The off-diagonal elements of the transition matrix, denoted as $p_{i\to j}$, provide data on the probability for transitioning from AEi to AEj over the sampling time interval (δt), which we preset at 1 ns for our investigation. This choice represents the best compromise between capturing significant atomic AE exchanges and managing computational costs. This period effectively balances resolution and computational efficiency, ensuring that

observed AE exchanges reflect sustained, dynamic processes within the NPs. This timeframe minimizes the influence of transient thermal fluctuations, focusing on the more substantial changes we aim to observe, and aligns with our goal of accurately depicting meaningful dynamics in nanoparticle behavior. We observe a high mobility of atoms within this AE at 673K, where transitions are observed to/from all other AEs, including the innermost bulk AE, behaving similarly to a viscous liquid. Specifically, at 673K, surface atoms have a probability of approximately 61% of remaining in AE *s* during the 1-ns sampling interval, behaving like a viscous liquid.

This probability increases to $\sim$ 91% at 300K. However, even at this lower temperature, transitions from the surface (s) to the concave AEs (c, c') and edges (e) can still be observed with probabilities significantly above $\sim$ 1% (in $dt = 1ns$).

By dividing $p_{i\to j}$ by dt, one can estimate the transition rate ($k_{i\to j}$), and consequently calculate the characteristic timescale ($\tau_{i\to j}$) expected for each transition (the reciprocal of $k_{i\to j}$). The right parts of Figure 5.4**b** and Figure 5.4**d** show the characteristic timescale of transitions involving surface AEs (s). Comparing the data extracted at 673 and 300K, the transition times diminish substantially at higher temperatures, exemplified by the characteristic timescales for $s\to c$ transition shrinking from $\sim$ 66 ns at 300K to $\sim$11 ns at 673K. Such different dynamics as a function of temperature, are even more pronounced for transitions between the inner AEs; the interior of the NP is almost static at 300K e.g., the $s\to ss'$ (*i.e.*, surface to subsurface) transition time drops from $\sim 10^4$ ns at 300K to $\sim$14 ns at 673K, a diminution of three orders of magnitude. These analyses demonstrate how such a combined experimental/computational approach allows achieving a resolution of the *ps* scale, to reconstruct the atomic dynamics present over a real 6-second experimental time window. Such level of detail provides crucial insights, such as, e.g., how long an AE exists in realistic conditions, which is key to understanding the surface properties of these NPs.

5.2.4 Rare local transitions on the NP surface in experimental conditions

While the equilibrium and average dynamic picture discussed thus far are useful, they could mask significant local fluctuations, key for the NP properties. Such analyses based on pattern recognition of the statistically dominant AEs may lose information, in particular, on fluctuations/transitions that may sparsely occur along the MD trajectories and that have negligible statistical weight. To address this issue, we have completed our previous analysis using a different abstract descriptor called "Local Environments and Neighbors Shuffling" (LENS) [203]. (See Chapter 2 for more details) LENS allows detecting and tracking rare local fluctuations, by monitoring how much every Au atom in the NP changes neighbor individual atoms identities (IDs) every 1 ns along the MD trajectories, which are typically overlooked in pattern recognition structural (e.g. SOAP) based analyses. Figure 5.5**a** illustrates the time-series of LENS signals, denoted as $\delta_i(t)$, obtained from one of the MD trajectories at 300K. These signals were computed starting from the experimentally reconstructed structure obtained at 2.4 s. Additionally, the figure includes the Kernel Density Estimate (KDE) of the LENS distribution and the interconnection dendrogram, covering the final 1μs of our AuNP MD simulation at 300K. Atoms exhibiting elevated $\delta_i(t)$ values are suggestive of pronounced dynamism within their atomic environments. These elevated $\delta_i(t)$ values are indicative of considerable variations in the number and identities of neighboring atoms, signifying notable changes in the local neighborhood of each atom. Consequently, atoms characterized by persistently elevated $\delta_i(t)$ values are systematically classified into the more dynamically active regions. This criterion underpins our methodological approach in differentiating between various dynamical states of the atoms, with the most dynamically active clusters comprising those atoms exhibiting the highest values of $\delta_i(t)$.

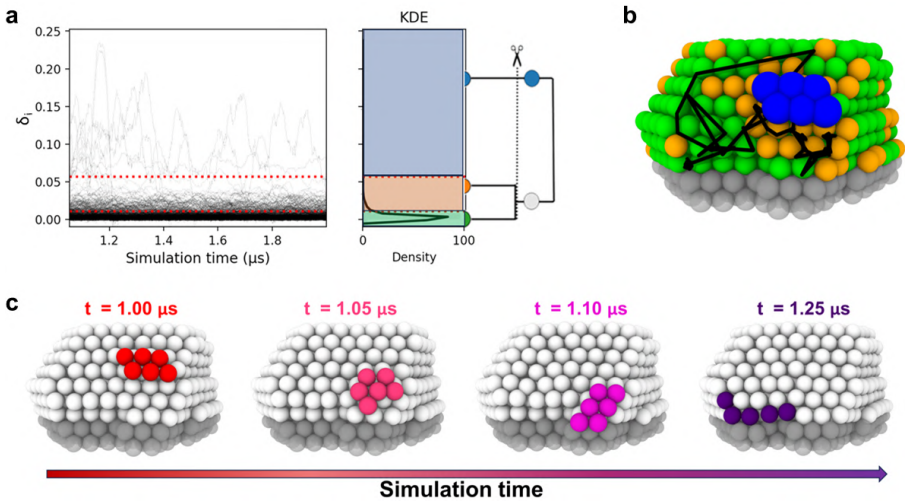


Fig. 5.5 **Exploring local fluctuations and dynamic domains within reconstructed Au NPs.** (**a**) Left: Time-series of LENS signals ($\delta_i(t)$), Kernel Density Estimate (KDE) of the LENS distribution [203], and interconnection dendrogram for a 1-microsecond window of our Au NP MD at 300K. The KDE provides an overview of the LENS distribution, while the dendrogram illustrates the interconnections between dynamic domains. (**b**) An example of a local fluctuation, where a small group of Au atoms exhibits rapid diffusion on the NP surface. Atoms are colored according to the LENS cluster represented in the KDE plot. The corresponding movements during the MD simulation on the underlying NP surface are highlighted by black trajectory lines. This dynamic behavior of a localized group of atoms adds complexity to the overall NP behavior, with potential implications for system properties and reactivity (**c**) Four snapshots at different times (indicated above) illustrate the movement of the aforementioned group of dynamic atoms. The color scheme used is relative to the simulation time, as represented by the arrow below. *Reproduced with permission from Ref* [342].

More precisely, by applying K-means clustering[370] to the LENS data, we identify 3 distinct environments, represented in the right panel of Figure 5.5**a**, characterized by different LENS signals (local dynamics), enabling the construction of an associated dendrogram illustrating their adjacency.

This detailed dynamic analysis complements our equilibrium SOAP analysis, highlighting localized dynamic areas amidst a backdrop of relative stability, and underscoring the complexity of NP behavior. Indeed, in line with the equilibrium SOAP analysis, the majority of NP atoms remain within a relatively "static" region, visualized by the green domains. However, a sparse subset of atoms participates in local transitions and diffusion, leading to the emergence of localized dynamic regions shown in blue. Figure 5.5**b** provides an illustrative example of these local fluctuations, where a few Au atoms demonstrate rapid movements on the NP surface.

Such infrequent behaviors, which may not be as evident in other analyses like those in Figure 5.4, are effectively detected by LENS, providing a deeper insight into the microscopic and macroscopic features characterizing the atomic dynamics of these metal NPs under experimentally relevant conditions.

5.3 Conclusions

In this work, we presented a combined experimental/computational approach that allows us to characterize the (complex) atomic dynamics present in Au NPs in realistic relevant conditions. Thanks to the capability of high-resolution ADF-STEM microscopy, we demonstrate the potential of this approach by obtaining ten images of a real Au NP every 0.6 seconds along a total of 6 seconds of experimental data acquisition. These images are then used to reconstruct as many atomic precise 3D structures of the Au NP, which are then used as a starting point for ten microseconds-long independent MD simulations. By integrating advanced descriptors of atomic environments with machine learning, we can track the atomic-scale rearrangements of the individual atoms on the NP over time, combining in this way the high spatiotemporal resolution of fully atomistic MD and the advantage of starting from multiple (time-decorrelated – experimental dt=0.6 s) NP structures. This allows us to minimize the typical sampling limitations of MD simulations starting, e.g., from initially perfect (ideal) NP structures.

MD simulations starting from NP structures taken every 0.6 *s* along a multi-second experimental data acquisition show consistent and conserved atomic dynamics within the NP. This is true both at 673 K (same temperature as during the data acquisition) and at 300 K (see Figure 5.2, Figure 5.3 and Figure 5.4). Although the simulations at 673 K show pronounced atomic dynamics, the nanoparticle maintains its octahedral structure. This stability, observed under experimental conditions, illustrates that the dynamics in our simulations align with the equilibrium behavior expected in practical environments where NPs are actively employed.[365, 364, 366, 368] This means that such an approach allows us to reconstruct in a reliable – notably, with a spatial resolution of the Å, and the time resolution of the picosecond – way the structural dynamics that such metal NPs have when observed along an experimentally-relevant seconds time-window.[331]

The ML approach used herein to analyze the MD trajectories allows us to identify and classify all the AEs that emerge within the Au NP in realistic conditions, and to quantify, e.g., their lifetimes, transitions, etc. The dynamics reconstructed via SOAP and LENS data identify dominant average dynamic behaviors as well as sparse concerted movements, such as the "terrace sliding" on the NP face shown in Figure 5.5 that, being rare events, are typically difficult to capture with conventional pattern recognition approaches. On the one hand, such insights provide a new, exquisitely dynamic (qualitative) view of these atomic NPs. At the same time, the analyses of these MD simulations provide a useful approach to interpret and rationalize, in the future, the properties of such NPs in experimentally relevant conditions (e.g., their reactivity, stability, etc.).

We believe that the comprehensive understanding of the internal atomic dynamics that it is possible to attain for metal NPs with such a combined experimental/computational approach will offer a fundamental tool for rational control of NP properties. On a first simpler step, for example, it provides a direct way to understand the effect of the environmental conditions on the behavior expected from these NPs. Increasing with the complexity, it will also pave the way toward a better understanding of the properties of such metal systems, e.g., under the exposure to various stimuli or revealing, for example, the effect of the interaction with different entities (e.g., reactants, molecules, etc.) on the structural dynamics of the NP and, vice versa, the effect of the dynamics of the AEs that populate the NPs on the NPs' properties (e.g., interactions, reactivity, etc).

In summary, our work not only advances the current state of the art in studying metal NPs but also highlights the remarkable stability of these systems under realistic conditions. By integrating experimental data and computational simulations, we bridge the gap between theory and practice, offering valuable insights into the dynamic behavior of metal NPs in real-world scenarios. We are confident that our approach will serve as a foundation for further advancements in NP research, enabling precise control and optimization of NP properties for various applications.

5.4 Methods

5.4.1 Atomistic MD details

To simulate the NPs, we used the SMATB[227, 271, 339] potential, available in LAMMPS [344], acknowledged for its specific application to Au NP. We also validate our approach through extensive comparison with experimental data, emphasizing atom position precision and nanoparticle morphology to ensure theoretical and practical congruence of our models. The NP models were initially minimized using the built-in command in LAMMPS (set up with etol = 10^{-6} ftol = 10^{-8}, maxiter = 1000 and maxeval=10000), then we performed a small thermalization of 20000 MD steps with the timestep set to 1 fs on the NP with the velocities initialized to the desired temperature and with the thermostat with the same settings of the main simulation. We then simulated 10 reconstructed Au NPs[331] at temperatures of 300 and 673 K. The number of atoms in each configuration, starting from the first structure to the tenth, are as follows: 1038, 1031, 1044, 1044, 1047, 1037, 1042, 1030, 1036, and 1035 All MD simulations were conducted in the canonical ensemble using the LAMMPS's Langevin thermostat, using a timestep of 5 fs, and a damping parameter for the Langevin thermostat set to 100 ps. We simulated each NP system for a total of 2 μs of MD. During the simulations, all NP systems reached a steady state in the MD regime (equilibrium). All our analyses were thus conducted on 1000 frames taken every 1 ns along the last 1 μs of each MD simulation, during which the populations of all detected AEs are plateaued[342]. It is important to note that in our simulations, we did not explicitly include the ceria (CeO_2) substrate on which the Au NPsarticles are often supported. [371, 367, 372, 373] Instead, to mimic the effect of the substrate on the NPs, we applied a potential to the last layers of the NPs. This approach was chosen as our primary interest was in understanding the intrinsic dynamics of the Au NPss themselves, rather than the interactions between the NPs and the ceria substrate. By focusing on the Au NPs and employing the SMATB potential[227, 271, 339], which has been shown to accurately describe the dynamic behavior of these systems [171], we were able to more effectively investigate the properties and behavior of the NPs under the conditions of interest. A key observation from our study is that the structure obtained after MD simulations closely resembles the structure reconstructed experimentally. This similarity indicates that the potential we applied successfully reproduces the impact of the substrate on the structure

of the NP. This decision to omit the explicit simulation of the substrate allowed us to concentrate our computational resources and analysis on the aspects most critical to our aim. Indeed, including the substrate would have made our analysis system dependent, and so we opted for keeping our computational framework more affordable and gaining in terms of generality and applicability.

5.4.2 Atom-counting for NPs at high temperature

The atom counting is based on so-called scattering cross-sections (SCSs), representing the total intensity of electrons scattered toward the ADF detector for every atomic column. These SCSs can be quantified using statistical parameter estimation theory [374]. To achieve this, images are modeled as a superposition of Gaussian functions using the StatSTEM software. From the estimated model parameters, which encompass the positions, heights, and widths of all atomic columns, the SCS values are determined. In a subsequent analysis, the distribution of the SCSs of all atomic columns is decomposed into overlapping normal distributions, i.e., a Gaussian mixture model. This allows us to count the number of atoms in a particular atomic column with single-atom sensitivity. However, to ensure the reliability of our results, especially when prior information is lacking, we validate our findings by comparing them to reference SCS values obtained through accurate multislice simulations. These simulations, conducted using MULTEM [375], account for the unique characteristics of the detector, including the non-uniformity of the real detector surface [376]. Furthermore, to achieve the highest level of quantitative accuracy, we incorporate the temperature-dependent Debye-Waller factors into our image simulations. This accounts for changes in the root mean square deviation of Au atoms concerning temperature variations. An appropriate parameterization [377] is used for this purpose.

5.4.3 Reconstruction based on atom-counting results

To obtain the 3D atomic structure of the Au NP from the estimated number of atoms in each atomic column, our collaborator applied the proposed method to a simulated system as a validation step. They used the atom-counts procedure to generate an initial 3D model of the Au NP by arranging the atoms symmetrically around the central plane, based on the known specimen orientation ([110]) and the crystal

structure. Then they fixed the distance between adjacent Au atoms, along the beam direction, according to the lattice parameter[331].

After this, they employed an iterative local minima search algorithm to construct the final 3D structure using the starting input model. This process is designed to comprehensively navigate the energy landscape and prevent confinement to nearby local minima. In each iteration, a random atomic column was displaced upwards or downwards within the interval of [-a, a], where "a" represents the lattice parameter of the FCC Au structure. Based on the resulting change in the system's energy ΔE, calculated as the difference between the energy of the new configuration and the previous one, they computed the Boltzmann probability factor P[347, 378], utilizing Boltzmann's constant k_B and a selected temperature of 673 K:

$$P = e^{\left(\frac{-\Delta E}{k_B T}\right)} \tag{5.1}$$

If P for the candidate structure exceeded a specified threshold, it was accepted and used for the next iteration; otherwise, the previous configuration continued. This process was repeated for 2000 iterations, maintaining the structural integrity of the nanoparticle until significant deviations occurred due to the displaced atomic columns. A threshold of 0.9 was selected for efficient exploration within the energy range near the input model, accounting for computational advantages.

Each candidate structure linked to a local minimum underwent MD relaxation in a canonical ensemble at 673 K for 5 ns using a Nose-Hoover thermostat. Unlike standard energy minimization, this temperature-specific MD relaxation enables studying structures observable at elevated temperatures. Indeed, at elevated temperatures, the anisotropy of surface energy diminishes. This phenomenon leads to the emergence of rounded features in the equilibrium shape, along with the presence of kinks and steps on the surface. These surface irregularities serve as sites for atom sources and growth, facilitating the diffusion of adatoms [379]. During the iterative search and MD simulations, the EAM potential described Au atom interaction, and the interaction between CeO_2 support and the particle was considered using LJ interaction[331].

To select the most plausible 3D structure, a fitness function (f) was defined.

$$f = \frac{E}{atom} + \alpha\chi \tag{5.2}$$

It incorporated the average potential energy per atom ($\frac{E}{atom}$) and a quantitative goodness-of-fit measure (χ) of candidate structures with the reference observation. The fitness function balanced$\frac{E}{atom}$ and χ using an empirically chosen weighting parameter (α). This function's design is taken from Yu *et al.*[380], utilizing atom counts and projected atomic column displacements for the discrepancy definition.

The minimum in the fitness graph yielded the final 3D structure aligning best with the reference ADF STEM image in terms of atom count and projected atomic column position[331]. The retrieved structure primarily comprises 100 and 111 facets separated by edges and corners. A quantitative comparison with the exact 3D model of the reference image verified the proposed methodology[331]. Discrepancies in the number of atoms between reconstructed and original input model atomic columns were minimal, attributed to methodological limitations and atom movement during MD relaxations. Surface structure analysis highlighted an accuracy of over 95% in identifying the reconstructed Au NP's surface structure[331]. The iterative local minima search algorithm was then applied to reconstruct the 3D structure of the experimentally investigated NPs of Figure 5.1**g**

It is important to note that in our case, images are recorded sequentially. The recording dwell time per pixel in this ADF STEM experiment is 0.6 μs, and the total recording time required to capture one atomic column is therefore in the range of several microseconds. During this time, the atomic structure is likely to be averaged out experimentally. Each atomic column is revisited after 0.6 seconds. This sequential recording approach introduces inherent temporal averaging, which can contribute to mitigating the effects of atomic vibrations and enhancing the overall signal-to-noise ratio. It effectively accounts for the dynamic nature of atomic motion during imaging, particularly at elevated temperatures. By revisiting each atomic column within a relatively short time frame, the experiment captures a series of atomic snapshots, contributing to a more accurate representation of the dynamic behavior of the NP's atomic structure.

In conclusion, the integration of atom-counting with an iterative local minima search algorithm, incorporating temperature effects and particle-support interaction, facilitated the accurate and precise reconstruction of the 3D structures for both simulated and experimentally observed supported NPs. Unlike approaches solely relying on energy minimization, our method outperforms previous techniques that combined 3D atom counting with MC or MD simulations. This approach overcomes inherent limitations by effectively navigating the local energy landscape to pinpoint

the local minimum corresponding to the imaged NP structure. This capability enables the successful estimation of target structures, encompassing atomic column positions and surface atomic configuration, as observed in ADF STEM images. Thus, our methodology offers a robust means to retrieve comprehensive 3D atomic-scale insights into stable and metastable structures in experimental relevant conditions.

5.4.4 Temporal analysis

To investigate the temporal behavior, we calculated transition matrices based on the cluster information of each atom throughout the simulation [248, 204]. Transition matrices are constructed by accumulating a table where the elements represent the number of transitions from state i to state j, or from state i to state i, observed at each time step. We further obtain the probabilities for an atom to transition to a specific atomic environment (or to remain in the same environment) after each time step (with a time interval of $\delta t = 1$ ns in our analyses) by normalizing each row to 1. In the figures presenting the transition matrices, unobserved transitions are denoted by blank squares. By employing these methodologies, we were able to perform a detailed analysis of the atomic environments and their temporal behavior in the simulated systems. We note that the 1 ns timeframe chosen for our analysis is crucial in distinguishing processes occurring on the nanosecond scale or slower. This consideration significantly reduces the likelihood that the exchanges between atomic environments are merely attributable to thermal vibrations. Additionally, the SOAP spectra for each atom in the NPs were calculated based on 1000 molecular dynamics snapshots, taken every 1 ns from the last microsecond of the simulations. This approach, which uses unprocessed atomic positions, is critical for an accurate representation of the dynamics within the NPs.

Chapter 6

Local Dynamic Descriptors in the study of metallic systems

The macroscopic behavior of complex systems is significantly influenced by local fluctuations, which are critical for understanding the dynamics of these systems but that are typically difficult to detect and control[203, 363]. These rare events are crucial in various phenomena, including nucleation, defect emergence and propagation, phase transitions, and triggered atomic motions in metal systems[203, 363, 171, 342]. Understanding and managing these fluctuations are key to advancing our knowledge of these processes. As highlighted in Chapter 2, to enhance the analysis and interpretation of these complex processes, conventional human-based descriptors are progressively being replaced by abstract descriptors[202, 381–384, 204, 385], often combined with supervised and unsupervised ML methods[171, 172, 386–389, 248, 390, 261, 391].

While human-based descriptors can provide an accurate understanding of intricate physical-chemical mechanisms, they also heavily rely on prior knowledge of the system. This limits their transferability among different systems. The design and the use of abstract and high-dimensional descriptors allowing for more general representations and providing a broader perspective on system behavior would guarantee a more general applicability to various systems. Consequently, typical approaches based on dimensionality reduction principles (e.g., linear PCA, kernel-PCA[392], t-SNE[393]) are frequently employed to extract information from such datasets. The reduced datasets are then classified using diverse clustering meth-

ods (e.g., KMeans[394], Gaussian Mixture Models (GMM)[395], DBSCAN[396], HDBSCAN[297]) to facilitate interpretation.

However, when relying on structure-based descriptors, these approaches may have limitations. While they effectively detect dominant structural environments in the system, they may fail to capture local time-dependent events that are sparsely observed within the trajectory. These transitions, although statistically infrequent, play a crucial role in the overall behavior of the system[203, 363]. The absence of an adaptive resolution capable of capturing non-dominant events presents two significant challenges: firstly, it results in a loss of information by failing to detect fluctuations within the system; secondly, these fluctuations may be inaccurately classified within the dominant clusters, thereby contaminating them[234].

Thus far, in the works presented in this thesis, we have primarily focused on the descriptions provided by SOAP as high-dimensional descriptors for characterizing the most prevalent and persistent atomic environments in our metallic systems[172, 171, 342]. However, in the last section of the previous chapter, through the introduction of LENS-based analysis[203], we highlighted the advantages of using a dynamic descriptor to capture these rare (less statistically significant) but crucial phenomena that get typically lost in SOAP-based pattern-recondition analysis. (see Figure 5.5).

Starting from these considerations, the first section of this chapter presents the results of applying LENS and SOAP to the Cu(211) surface, already studied in Chapter 3. The results reported here are part of the work: Crippa, M., Cardellini, A., **Cioni, M.**, Csányi, G., & Pavan, G. M. (2023). "Machine learning of microscopic structure-dynamics relationships in complex molecular systems." *Machine Learning: Science and Technology*, 4(4), 045044. DOI: 10.1088/26322153/acf1e2[203].[1]

The second part will discuss the application of LENS to analyze the deformation of a copper system, as examined in our ongoing project (preprinted not yet available) *"From Local Environments to Overarching Behavior via Tracking and Relating Structural and Dynamical Events"*.[2] We will demonstrate how LENS can capture collective phenomena that are essential for understanding the physics of the system.

[1] My contribution to this manuscript involved analysing and interpreting the results for the Cu FCC-211, which are reported in this chapter.

[2] My contribution to this work involved performing simulation and analysis for the Cu sample reported in this chapter.

6.1 Introduction

The primary goal of this chapter is to illustrate the limitations of relying solely on structural descriptors for understanding dynamically evolving metallic systems. To achieve this, we employed machine learning in the study of microscopic structure-dynamics relationships in metallic systems, demonstrating the advantages of combining different descriptors to enhance our understanding.

In the first part of this chapter, we will show how to improve the accuracy of structural and dynamic classifications by integrating dynamic descriptors such as LENS[203] with structural descriptors like SOAP[202]. This approach aims to reduce noise and degeneracy issues, providing deeper insights into microscopic structure-dynamic relationships. Combining these descriptors offers a comprehensive characterization of complex systems, revealing the origins of dynamic fluctuations. In the second section, we will present an example demonstrating LENS's ability to detect dynamic events that SOAP cannot capture, specifically focusing on the dislocation of a copper sample. This example will highlight LENS's potential in identifying crucial dynamic phenomena essential for understanding the physical properties of the system. The results from this project will soon be submitted as a preprint.

6.2 Machine Learning of microscopic structure-dynamics relationships in complex molecular systems

6.2.1 Results

After the introduction of the new SOAP&LENS descriptor and detailed methodologies in Chapter 2, we now explore their practical application to a Cu(211) FCC metal surface [172]. This DPMD simulation at T=600 K reveals that this slab displays a structurally diverse surface with intricate internal atomic dynamics (Figure 6.1a, right). Understanding the underlying mechanisms behind such dynamics is essential for comprehending the properties of such metal systems [244–246].

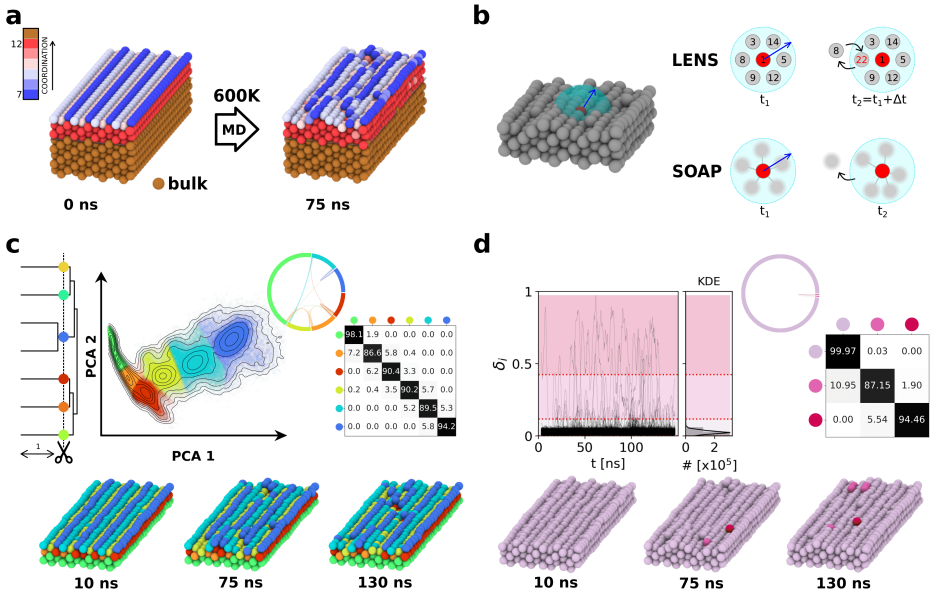


Fig. 6.1 **Flow of the analysis.** (a) FCC 211 copper slab snapshots colored by atom coordination (excluding bulk) at t=0 ns and after 75 ns of an MD trajectory at T=600 K. (b) LENS and SOAP: given the local neighborhood (cyan sphere) of each atom (red atom) in the system, LENS tracks the *identity* of the neighbor atoms within it (no information on their geometrical organization is retained), while SOAP captures their structural arrangement (without tracking their identity: it is a permutationally-invariant description). (c) SOAP-based analysis of Cu(211) system. Left: HC-based dendrogram and dendrogram cutting, defining the merged macro-clusters. Middle: PCA of the SOAP dataset (first two principal components), colored based on the detected macro-clusters. Right: chord diagram (fluxes) and transition probability matrix for the dynamical transitions between the macro-clusters (SOAP environments). Bottom: surface MD snapshots where atoms are colored based on the classification: bulk atoms in green, sub-surface in orange and red, surface "valleys" in yellow, faces in cyan, and edges in blue. (d) LENS analysis of Cu(211) at 600 K. Left: LENS time-series and classification.[203] Right: chord diagram (fluxes) and transition probability matrix. Bottom: MD snapshots with atoms colored based on the LENS clusters: more/less dynamic atoms in brighter/lighter colors. *Adapted with permission from Ref* [234].

We have already observed that a structural-descriptor-based analysis, such as using SOAP combined with dimensionality reduction and density-based clustering, captures the most prevalent and dominant conformation domains within the system.[172, 203, 363]. In contrast, a pure LENS analysis, based on the reshuffling of the neighborhood over time, captures the dynamical features of the system (see Figure 6.1b).

To investigate both the structural and dynamical properties of the Cu(211) system, we adopted a similar *bottom-up* protocol as described in previous chapters.

This strategy includes, as a first step, a representation of the system *via* the SOAP descriptor. One SOAP spectrum is extracted for each of the 900 atoms (three top-most layers, although the SOAP spectra also consider the presence of the 1500 bottom-side atoms as neighbors, they are not included in the analysis because we are interested in the dynamics of the surface[172]) in 482 snapshots taken every Δt=0.3 ns along 145 ns of MD simulation, for a total of 482×900 spectra. A PCA is then used to reduce the dimensionality of the SOAP spectra dataset, considering the first n-PCA components to retain at least 99.5% of the variance. An unsupervised clustering algorithm (HDBSCAN*[297], see Methods sections) is finally adopted to rationalize the data and identify the dominant AEs on the surface[234]. From the atoms' transition between the clusters over time, we compute a transition probability matrix, as already described in Chapters 3,4 and 5. From the transition probability matrix, we construct a Hierarchical Clustering (HC) based dendrogram merging the clusters with high dynamic interconnection. The dendrogram is cut to retrieve only meaningful clusters, colored accordingly in the PCA plot of Figure 6.1c, where only the first two PCA components are reported.

The results demonstrate how SOAP can successfully distinguish diverse structural environments within this system, including the bulk (green), subsurface (orange and red), surface valleys (yellow), faces, and edges (cyan and blue), identified in different colors in Figure 6.1c. The dynamic interconnections between the various clusters on the surface are also represented by the chord diagrams in Figure 6.1c on the right. In these chord diagrams, the dimensions of the arcs stand for the population of the various clusters, while the dimensions of the strings connecting them give visual information on how pronounced the atomic flow is in Δt, and thus on their dynamic interconnection. Moreover, we also obtained the transition probabilities matrix (% to undergo transition in Δt=0.3 ns) between the HC-merged clusters (Figure 6.1c right).

Separately, we also performed a LENS analysis (see Methods for detail) on the same 482 snapshots extracted by the same MD trajectory. This analysis of the system reveals intriguing surface events that are not captured (or highlighted) by the static SOAP-based analysis of structure as described above. Specifically, a few Cu atoms are seen to detach from the crystalline structure of the surface and diffuse on it very fast. On the one hand, since these diffusing atoms are characterized by a high rate of reshuffling of their neighbors, they are clearly identified by LENS as a separate environment in the dataset (Figure 6.1d). On the other hand, a comparison of Figure 6.1c and Figure 6.1d shows how, since these points are sparse and have negligible

statistical weight in the dataset, these are overlooked in a pattern recognition approach such as that, for example, of Figure 6.1c. In particular, in the SOAP analysis of Figure 6.1c, it is possible to note that the diffusing atoms (magenta in Figure 6.1d), are merged into the SOAP cluster identifying the edges of the Cu(211) surface. This merging causes two problems: it distorts the representation of the SOAP cluster and results in the loss of important information about dynamic events.

We developed a combined approach based on the basic assumption that a structural environment at a certain time might influence the dynamical events within the subsequent time interval. Starting for example at time t_1, a SOAP spectrum $\mathbf{p}_i^{t_1}$ is computed for each particle i in the system. We also calculate its LENS value for the immediately subsequent time interval $\delta_i^{t_2-t_1}$. By including the LENS term as an extra component into each SOAP power spectrum, we thus obtain a new vector $\chi i^{t_1} = (\mathbf{p}i^{t_1}, \delta_i^{t_2-t_1})$ containing information on the structural properties in the neighbor environment surrounding atom i at time t_1 and its evolution in the subsequent time interval $t_2 - t_1$. The SOAP spectrum and LENS scalar component are opportunely normalized to have the same weight in the dataset (see Chapter 2 for details). Iterating this procedure for the whole trajectory, we thus obtain a new dataset (SOAP&LENS dataset) comprising $N = N_{particle} \times N_{frames}$ vectors, each one of dimension $n+1$, where n is the SOAP spectrum dimension (structural information) and 1 the LENS (dynamical) component. Such an updated dataset effectively contains information on the instantaneous environments surrounding each particle i and how they are prone to change over time at the resolution Δt (0.3 ns) of our analysis. This ensures that each vector has $n+1$ components containing information about the structural environments of each atom's neighborhood at time t_i, plus information about the fluctuations emerging from these environments in the subsequent time interval Δt. This method allows us to delineate a new concept for classification, as reported in Figure 6.2. On the left side, Figure 6.2a shows the PCA of the SOAP dataset projected onto the first two PCA components. On the right side, Figure 6.2a shows the same projection for the new SOAP&LENS combined dataset. Notably, while the majority of the data has an almost identical distribution on the two PCAs, a distinct cloud of points appears as evidently separated from the rest in the combined dataset (top-right: highlighted by the red circle).This identifies a new environment that, in the combined SOAP+LENS analysis, appears separate, whereas in the SOAP-only analysis, it was merged into other environments.

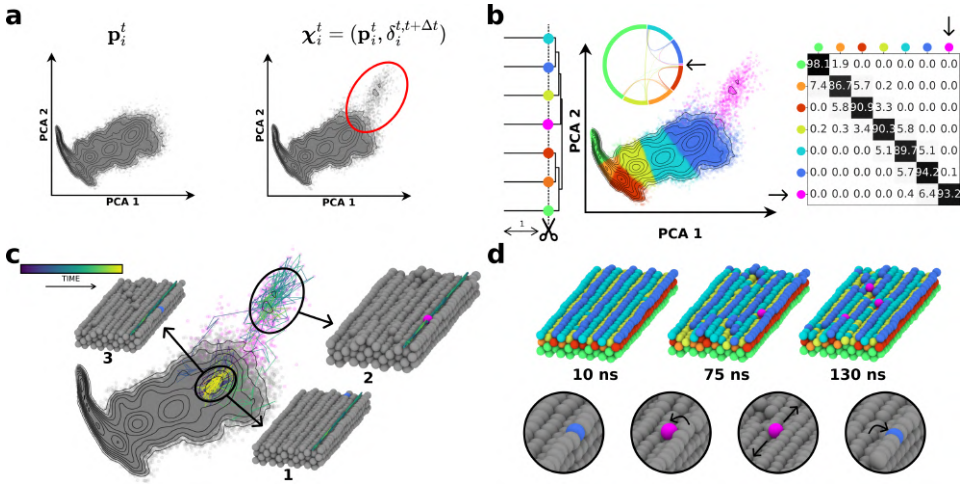


Fig. 6.2 **Combined SOAP&LENS analysis of a Cu(211) surface at 600 K.** (a) Left: First two PCA components of the SOAP power spectra of the Cu(211) system. Right: First two PCA components of the SOAP&LENS combined descriptor: the new cloud of points emerging in the PCA projection of the χ vector is highlighted by the red circle. (b) SOAP&LENS based analysis of Cu(211) system. Left: HC-based dendrogram and dendrogram cutting, defining the merged macro-clusters (accordingly to clusters in Figure 6.1a, except for a new pink cluster). Middle: PCA of the SOAP&LENS dataset (first two principal components), colored based on the detected macro-clusters and chord diagram (fluxes). Right: transition probability matrix for the dynamical transition between the macro-clusters, highlighting the new cluster in pink. (c) Trajectory of an atom detaching from an edge, running on the surface, and re-entering into the edge. The trajectory is shown both on the PCA plot and on the snapshots, colored from blue to yellow according to the time evolution. (d) Three surface MD snapshots colored based on the classification: bulk atoms in green, sub-surface in orange and red, surface "valleys" in yellow, faces in cyan, edges in blue, and pink atoms detaching from the edges and running on the surface (an example of this process is reported in the zoom below). *Reproduced with permission from Ref* [234].

As shown in Figure 6.2b, unsupervised HDBSCAN* clustering combined with HC-based merging reveals that such a separated domain on the SOAP&LENS PCA identifies a distinct, specific local environment. The clustering parameters used for the analyses of Figures 6.1c and 6.2b are exactly the same. This comparison shows how the classification of Figure 6.1c (SOAP only) is enriched via the detection of a new LENS environment identified by the pink color (highlighted by the arrows in the transition matrix and chord diagram of Figure 6.2b). As done for both the SOAP and LENS independent analyses, we can reconstruct the evolution of the detected environments by following the AE belonging to all atoms at every time step (see the chord diagram and transition probability matrix in Figure 6.2b, right).

This combined SOAP&LENS analysis offers distinct advantages over the purely SOAP-based approach. The decoupling of this additional pink LENS environment not only provides a more complete description of what happens on the Cu(211) surface at 600 K but also improves the statistical precision in the classification of the SOAP environments. Differentiating the structural from the dynamical environments reduces errors in detecting the SOAP AEs in the SOAP&LENS dataset. The PCA area identified by the red oval in Figure 6.2b, which corresponds to a well-defined LENS AE, merges into the SOAP AEs in the PCA of Figure 6.1c, creating errors and increased uncertainty. Thus, when combined, two distinct descriptors such as SOAP and LENS complement and improve each other. Furthermore, this approach allows tracking the origin of local dynamical (LENS) fluctuations occurring on the surface, outlining microscopic structure-dynamics relationships.

The off-diagonal entry in the matrix of Figure 6.2b representing the transition of atoms from the edge AE (in blue) to the pink (LENS) environment (about 0.1% probability) reveals that those atoms diffusing with high-speed on the metal surface come from the surface edges. After their creation and diffusion, these diffusing pink atoms are then reabsorbed into the surface edges (about 6.4% probability). The large imbalance between the probabilities for the creation and annihilation of these LENS diffusing atoms (about 0.1% vs. about 6.4%) indicates that the emergence of such fast atoms is a rare event. Yet, detecting such diffusing atoms is key to understanding the behavior of the system. Figure 6.2c provides an example of the structural variation of an atom undergoing such a transition, following its trajectory both on the PCA plot and along the MD. The atom's trajectory is color-coded based on the MD simulation time, from dark blue to yellow, showing atoms that, after residing within the surface edges (dark blue lines, example snapshot 1), detach and diffuse on the surface becoming part of this pink LENS environment (green lines, example snapshot 2), and then are reabsorbed into the edges (yellow lines, example snapshot 3). Figure 6.2d shows a complete representation of the Cu(211) surface colored based on corresponding SOAP&LENS environments. Unlike the snapshots of Figures 6.1c and 6.1d, this comprehensive approach captures all the key SOAP as well as LENS environments, providing a more complete characterization of this system. By combining these two descriptors, it becomes evident that the motion of atoms diffusing on the surface (pink AE) originates from fluctuations within the SOAP environment, which defines the edges of the surface (blue AE).

6.2.2 Conclusions

In conclusion, the findings presented in this section emphasize the limitations of relying exclusively on structural descriptors to comprehend the physics of dynamically evolving copper surface. By combining the dynamic descriptor LENS with the structural descriptor SOAP, we realized multiple benefits. This integration improves the precision of both structural and dynamic classifications by minimizing noise and addressing the degeneracy issues that are common in individual analyses. Moreover, it enables the identification of specific structural environments that generate dynamic behaviors, providing deeper insights into microscopic-scale structure-dynamic relationships. This approach is particularly useful in capturing rare but crucial fluctuations on the Cu(211) surface at 600 K, which are often overlooked by traditional pattern recognition methods. The combination of SOAP and LENS not only provides a more complete characterization of the system but also allows for a better understanding of the origins of dynamic fluctuations, thereby offering a comprehensive tool for analyzing complex systems across various domains.

6.2.3 Methods

DPMD simulation

The atomistic model of Cu(211) surface is composed of 2400 atoms. The MD simulation is conducted at T=600 K *via* LAMMPS software[344] using a Neural Network potential built using the DeepMD platform,[180] as described in detail in Chapter3. The sampled trajectories are 150 ns long. A total of 502 frames are extracted every Δt=0.3 ns along the MD trajectory and used for the analysis.

SOAP analaysis

To describe the structural environment surrounding each particle within the simulations, we use the SOAP descriptor. We compute the SOAP spectrum $\mathbf{p}_i^t$ representing the local structural environment of each particle i at every timestep t within a cut-off radius r_{cut} = 6 Å The SOAP vectors are generated using *dscribe*[345], and both l_{max} and n_{max} parameters for spherical harmonics, and number of radial basis functions are set to 8, the same values of Chapter 3. Then, we applied the Principal Component Analysis algorithm to each dataset (as implemented in the SciPy python package

[346]), reducing the dimensionality of the representation to the first n-components, in order to reach a certain cumulative variance within each system. To analyze the reduced data of the we applied the HDBSCAN* [296] clustering algorithm set up with `min_cluster_size=80`. We used soft-clustering to assign the point classified as noise to their closer cluster. From the cluster transition probability matrix we found the relations within the environments *via* Hierarchical Clustering algorithm. Then, merging the ones closer than 1 in terms of the chosen metrics (`correlation`) and linkage (`average`), we obtained 6 macro-clusters.

LENS analysis

We compute the $\delta_i(t)$ signals for Cu(211) following a similar procedure reported in Crippa *et al.*[203], reducing the noise by using a Savitzky–Golay [397] filter (as implemented in the SciPy python package [346]). Each δ_i(t) signal is smoothed using a common polynomial order parameter of $p = 2$ and a time-window of 20 frames.

6.3 LENS tracking of collective phenomena in metals: the case of dislocation

6.3.1 Results

As widely known, brittle fractures in materials are controlled by the formation and amplification of microscopic defects. Herein, we focus on an initial test case, namely a Cu bulk subjected to a constant strain rate at $T = 300$ K [398], thus mimicking phenomena occurring during a tensile fracture. Figure 6.3A displays simulation results featuring a stress-strain curve alongside the potential energy profile. This visualization clearly shows the transition from elastic to plastic deformation, thereby validating the simulation's capacity to capture these critical phases accurately.

Building on this foundation, we use the LENS framework to gain insights into a selected distinct plastic event, as highlighted in Figure 6.3A. In Figure 6.3B, the temporal evolution of the LENS descriptor is traced, revealing a pronounced peak. Specifically, SOAP is not capable of detecting this event because it does not correspond to any significant structural rearrangement. Instead, the event involves

the reshuffling of neighboring atoms without altering their relative positions, which SOAP fails to capture. Indeed, SOAP is designed to be invariant to the permutation of atoms within the local environment, providing a consistent description of the structural arrangement of atoms regardless of their order. This invariance is useful for identifying common structural motifs but limits SOAP's ability to detect dynamic events where the identity of neighboring atoms changes without significant structural rearrangement. Conversely, LENS is not invariant to the permutation of atoms. LENS specifically tracks how the identities of neighboring atoms change over time, capturing the dynamics of reshuffling even when the relative positions of the atoms remain constant. This non-invariance allows LENS to detect rare and significant dynamic events that are missed by purely structural descriptors like SOAP.

This observation aligns with the traditional view that plastic deformations in metals are predominantly facilitated through dislocation planes. This assertion is supported by the findings in Figure 6.3C and D, where temporal and spatial correlation analyses are reported. These analyses demonstrate that LENS spikes are both closely timed (within 5 ps) and localized in space (the distance between atoms experiencing LENS fluctuations corresponds to the inter-atomic spacing, as shown in g(r) of Panel D). It is important to note that these events not only do not involve significant structural rearrangements, but they are also statistically rare, occurring within 5 ps in a comprehensive 150 ns simulation. This brevity is highlighted by presenting only a short segment of the LENS signals in the graph. Although these fleeting events are critical for understanding local dynamics, they represent a minor fraction of the total simulation time and may appear statistically insignificant in the broader temporal context. Consequently, pattern recognition-based analyses, such as SOAP, fail to detect these peaks and instead average them out with other signals.

Lastly, Figure 6.3F provides a three-dimensional representation of these regions, distinctly marking coherent areas of plastic deformation and identifying rows of high-LENS atoms, corresponding to dislocation lines. Elucidating the microscopic mechanisms of plastic deformation, such high-resolution analysis highlights the critical role of the reshuffling within the local atomic neighborhood, as captured by the LENS component, in understanding the deformation processes under applied stress. Shifting the focus from structural re-arrangements to dynamical reshuffling proves the accuracy of our descriptor in detecting such collective phenomena.

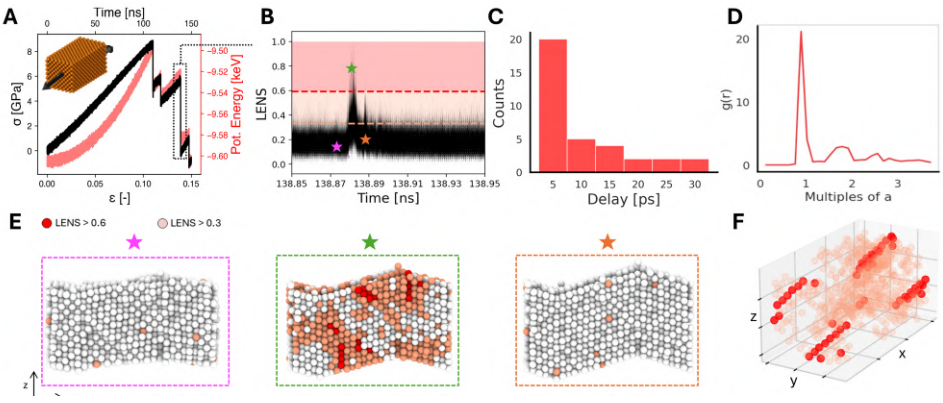


Fig. 6.3 **LENS analysis of Cu bulk metal during constant strain rate at T = 300 K[398].** (A) Stress-strain curve (black) over time with corresponding potential energy profile (red), indicating the regions of elastic and plastic deformation. The inset shows the simulated bulk copper structure, with arrows indicating the direction of strain.(B) Time-series of the LENS signal within the same timeframe, color-coded by intensity thresholds (LENS $>$ 0.3 in pink, LENS $>$ 0.6 in red); stars mark snapshots detailed in (E). This panel reveals how changes in LENS values dominate the LEAP signal over time. (C) Temporal correlation analysis of high-intensity LENS signals, showing that most LENS spikes occur within 5 ps. (D) Radial distribution function of atoms with high LENS signals, illustrating their spatial proximity. (E) Snapshots corresponding to the marked events in (B), color-coded to show the intensity of their LENS signals, highlighting both spatial and temporal correlations. (F) Three-dimensional representation of atomic positions with high LENS signals, emphasizing regions of coherent plastic deformation, identified as dislocation lines.

6.3.2 Conclusions

In this last section, we have demonstrated how LENS can elucidate the complex behavior of a metallic system by focusing on local neighbor fluctuations. By monitoring the identity changes of individual atoms over time, LENS effectively detects dynamic reshuffling events that are crucial for understanding the system's physics.

In particular, we examined the deformation mechanism of a copper system under stress. LENS reveals a distinct pattern where dynamic events, rather than static structural rearrangements, dominate the process. Such events are not detectable by SOAP because they do not involve significant structural changes, and their rare occurrence means they have minimal statistical weight in a SOAP-based analysis. This ability to capture rapid, localized fluctuations provides a clearer picture of the time-space correlation of metal deformation activities. LENS uniquely identifies these subtle but critical phenomena, offering deeper insights into the microscopic

mechanisms driving macroscopic behaviors in metals. This underscores the importance of integrating dynamic information for a comprehensive analysis of complex molecular systems.

Chapter 7

The effects of dynamics on amide reactivity in the cavity of a coordination supramolecular cage

The investigation of chemical reactivity is inherently complex, and elucidating the physicochemical mechanisms that govern these processes poses significant challenges, particularly from an experimental perspective. This complexity has driven a substantial interest in the use of computational simulations as a tool to explore and understand these intricate processes in greater detail.

However, traditional simulation approaches often rely on static models[164–170, 399], which inherently limit their ability to represent the dynamic nature of chemical reactions accurately. This simplification overlooks the crucial role of molecular motion and the dynamic interactions that occur during a reaction[159, 160, 235–239, 244, 245, 185, 201, 400]

In this thesis, we aim to incorporate the concept of dynamics into the study of reactivity, emphasizing the crucial role of dynamic behavior in computational approaches across various systems[172, 171, 342]. Our research on Cu surfaces and Au NPs has shown that under realistic conditions, microscopic dynamics significantly influence the structure of these systems. The continuous interconversion of atomic environments affects the availability of reactive sites and specific atomic configurations. This dynamic behavior is fundamental, as it directly impacts the properties and performance of these materials, which are intricately governed by the reactive sites and structural configurations available over time.[161, 185, 247, 250, 251]. Another

compelling example of the necessity of considering dynamic processes, though applied to a different type of system, is presented in our work "*Reactivity in Dynamic Host-Guest Systems at Atomistic Resolution: Amide Hydrolysis under Confinement in the Cavity of a Coordination Cage*"[401]. This study investigates the hydrolysis of amides within coordination cages, demonstrating how spatial confinement and dynamic host-guest interactions can accelerate the reaction. The encapsulation of amides within the cage cavity not only enhances the reaction rate but also stabilizes reactive conformers, illustrating the profound impact of dynamic factors on chemical reactivity.

Building on the foundational work of Takezawa *et al.*[402], which provided detailed structural elucidation of host-guest complexes using X-ray crystallography and NMR spectroscopy, this study goes further. While these techniques offer a precise but static depiction of the encapsulated system, showing atomic positions and molecular conformations at a particular time, they do not capture the dynamic processes of the molecules, such as conformational changes, binding/unbinding events, and transient states that can significantly influence chemical reactivity.

Capturing these dynamic processes provides a more accurate and predictive model of material behavior, whether involving metallic surfaces or molecular systems within coordination cages. The necessity of a dynamic approach becomes clear as it enables the observation of intermediate states and transient conformers, which are critical for understanding reaction mechanisms.

Full bibliographic reference: Delle Piane, M., Pesce, L., **Cioni, M.**, & Pavan, G. M. (2022). Reconstructing reactivity in dynamic host–guest systems at atomistic resolution: Amide hydrolysis under confinement in the cavity of a coordination cage. *Chemical Science*, 13(37), 11232-11245.[1] This work, my first manuscript during my PhD, significantly influenced the direction of my subsequent research activities. The insights gained here prompted me to extend my focus on the importance of integrating dynamic analyses in both molecular and metallic systems.

[1] In this first work, my contribution included performing Well Tempered MetaD and Infrequent Metad simulations for the isomerization of amide, reparametrization of the dihedral (see Methods) and contributed to the interpretation of the results.

7.1 Introduction

Chemists have developed synthetic cavities in the attempt to mimic natural catalytic systems, allowing for high selectivity in reactant encapsulation[403, 402, 404]. Supramolecular coordination cages, in particular, have been designed to host catalytic reactions within their internal cavities[402, 405–408]. These systems can manipulate reactivity, regio-selectivity, and enantio-selectivity by altering the structural and electronic properties of the cages or the guests[403, 409].Notably, mimicking proteolytic enzymes,[410] recent work has shown that encapsulating amide guests in coordination cages significantly accelerates amide hydrolysis by promoting reactive (*cis* isomers) configurations[402].

Designing these synthetic systems requires understanding the molecular and chemical-physical factors controlling their reactivity and mastering these to create new catalytic systems. However, the dynamic equilibrium of guest exchange in/out of the cages complicates this understanding[409].

Several factors control reactivity in these systems, including guest binding/unbinding dynamics, solubility in the outer solution, reaction timescales, entrapment in metastable states, and molecular concentrations. Non-covalent cavity–guest interactions play a major role since they control the binding affinity and, *de facto*, the probability of observing the reactant in the reactive site.[409] Despite significant efforts,[403, 408, 411, 412], it remains challenging to experimentally understand the molecular factors controlling reactivity in systems where reactant and product are in continuous molecular exchange.

Comprehensive descriptions that explicitly take into account the role of the internal dynamics of such complex dynamic systems are scarce. As a case study, we focused on a case study by Fujita's group,[402] where encapsulating amide guests within a coordination cage significantly enhanced amide hydrolysis, providing a test case for studying spontaneous reactions hindered by high kinetic barriers.[413] X-ray crystallography and NMR characterized the encapsulation of electron-rich diaryl amides into different octahedral coordination cages. The acceleration of amide hydrolysis (14x) when amide **2** is co-encapsulated with co-guest cage **3** inside cage **1** has been linked to the stabilization of reactive (*cis*-twisted) amide conformations within the cage host.[402]

To gain submolecular-resolution insights into this dynamic system, we combined all-atom MD and MetaD simulations, reconstructing the structural and dynamical

features of the host-guest system. *Ab initio* MetaD simulations[414] enabled us to study the chemical reactivity of the amide guest in its favored conformations within the cage. Coupling this with thermodynamic/kinetic studies of the amide and co-guest exchange, we developed a general scheme revealing key factors controlling the reaction enhancement.

7.2 Results

7.2.1 Atomistic modeling of the host-guest system

As a representative example of a supramolecular host, here we focus on the coordination cage **1** reported in Figure 7.1a recently employed by Takezawa et al. to host the hydrolysis of encapsulated amide guests.[402] This is an octahedral coordination cage composed of four self-assembled electron-deficient panel ligands (2,4,6-tris(4pyridyl)-1,3,5-triazine) and six metal-based corners (*cis*-endcapped Pd(ii) complexes).[415] As a first step, starting from the X-ray crystal structure reported in the literature,[402] we built an AA (All-Atom) model for cage **1** which was then preliminary minimized and equilibrated in explicit water solvent and in standard (room) conditions of temperature and pressure *via* a classical MD simulation. As the main guest, here we focus on guest **2**, an N-(2,4-Dimethoxyphenyl)thiophene-2-carboxamide, an electron-rich diaryl amide (Figure 7.1a) that was experimentally shown to produce the considerable reaction acceleration following to encapsulation in cage **1** (Figure 7.1b).[402] The central bond in amide guest **2** can undergo hydrolysis according to the reaction schematized in Figure 7.1b (top). Guest **2** is mildly apolar (logP=2.77)[401]. Experimentally-obtained crystal structures show that cage **1** can encapsulate at the same time up to two **2** molecules in its internal cavity[402]. An AA model was developed for guest **2**, paying particular attention to the accuracy in the force field parametrization of the central amide bond dihedral, which defines the *trans* and *cis* conformers of the amide, how much one is energetically favored respect to the other, and the related transition barrier. To achieve this, we optimized the original amide ω dihedral force field parameters of GAFF, to obtain a *trans*-to-*cis* isomerization free energy profile consistent with the experimental data available for N-metyl-acetamide (see Methods)[416].

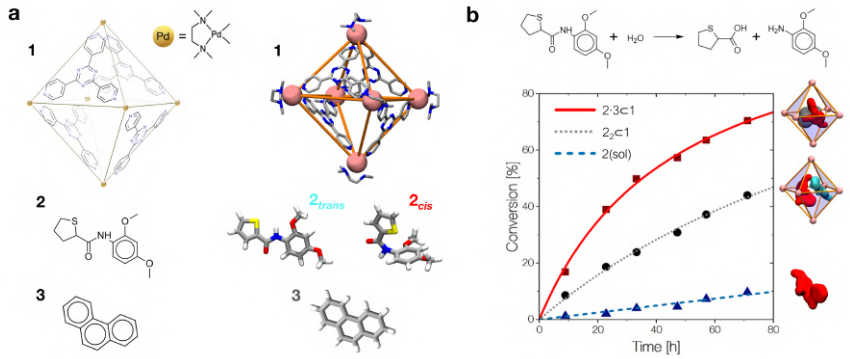


Fig. 7.1 **The host-guest systems studied in this work**. (**a**, top) The octahedral coordination cage **1** used as a reference in this work is composed of four panel ligands (2,4,6-tris(4-pyridyl)-1,3,5-triazine) and six metal corners (*cis*-endcapped Pd(II) complexes): chemical structure on the left and AA model on the right. (**a**, bottom) Chemical structure of amide guest **2** and of pirine co-guest **3**. AA models are reported on the right of each guest structure (*cis* and *trans* isomers are shown for **2**). (**b**, top) Hydrolysis reaction of the amide bond of **2**. (**b**, bottom) Experimentally observed percentage of hydrolyzed **2** over time. Conversion data are reported for the hydrolysis of **2** free in solution (in blue), in the case when two **2** guests are co-encapsulated in cage **1** (black), and for **2** co-encapsulated with co-guest **3** in cage **1** (in red: $\sim 14\times$ hydrolysis acceleration compared to the blue curve). [402] *Reproduced with permission from Ref* [401].

We built inclusion complexes where guest **2** is encapsulated within cage **1** in different stoichiometries: *i.e.*, $\mathbf{2} \subset \mathbf{1}$ and $\mathbf{2}_2 \subset \mathbf{1}$, where respectively one or two **2** amide guests are encapsulated inside cage **1**. In particular, the AA model for the $\mathbf{2}_2 \subset \mathbf{1}$ complex was built starting from the available experimental crystal structure for this complex[402], while that for the $\mathbf{2} \subset \mathbf{1}$ was obtained by deletion of one of the **2** guests. Aromatic amides, like **2**, exist mainly in a *trans*-planar conformation in solution. However, experimental evidence demonstrated that **2** may adopt a *cis*-twisted conformation within the T_d symmetric cavity (forming a pseudo-S_4 symmetric dimer) in cage **1**.[402] In particular, X-ray and NMR measurements showed signals corresponding to a *cis*:*trans* 1:1 **2** dimer in the $\mathbf{2}_2 \subset \mathbf{1}$ complex, and to a *cis*-twisted conformation in the $\mathbf{2}\cdot\mathbf{3} \subset \mathbf{1}$ complex. To the purpose of our investigation, here we decided to model all possible combinations of conformers, mainly aiming to explore the presence of any correlation between confinement, crowding and the rotation of the amide bond. We also parametrized an AA model for co-guest **3** (Figure 7.1a). Starting also in this case from a corresponding experimentally available crystal structure,[402] we use this to build an additional AA model for the $\mathbf{2}\cdot\mathbf{3} \subset \mathbf{1}$ complex, a ternary inclusion complex where one **2** guest and one **3** co-guest are simultaneously

encapsulated in cage **1**. Particularly interesting for this study, the $\mathbf{2}\cdot\mathbf{3}\subset\mathbf{1}$ system was observed experimentally to produce the larger hydrolysis acceleration among all explored cases (Figure 7.1b: $\sim$ 14x acceleration compared to the free amide).[402] When in the cage, **2** tends to stay shifted from the geometrical center of the cavity to maximize interactions with the walls, with the $\mathbf{2}_{trans}$ conformer showing in general a larger shift compared to $\mathbf{2}_{cis}$ (Figure 7.2b). This is imputable to the thiophene ring, which tends to partially stick out of the cavity preferring interaction with the solvent (scheme of Figure 7.2a, top: in cyan). Co-encapsulation with **3** induces a larger decentralization of the amide guests. Noteworthy, the augmented growth in such a case is observed to cause a significant reduction in terms of $\mathbf{2}_{cis}$ mobility in the cage cavity, as evidenced by the narrower distribution in Figure 7.2b (solid red curve). This behavior correlates with an augmented number of contacts between **2** and the cage **1** (Figure 7.2c). In general, the $\mathbf{2}_{cis}$ guest shows increased contacts than the $\mathbf{2}_{trans}$ one, while this trend is even increased when co-guest **3** is also present within **1** (higher crowding). We will discuss in the next section how this affects what conformer ($\mathbf{2}_{trans}$ *vs.* $\mathbf{2}_{cis}$) is more favored within the cage in the different complexes.

The Solvent Accessible Surface Area (SASA) of the guests shows that **2** is less exposed to the solvent when this is encapsulated within **1**, and even less when this is co-encapsulated together with **3** (Figure 7.2d). In general, we observe that the $\mathbf{2}_{cis}$ conformer has a smaller SASA, and is more compact than $\mathbf{2}_{trans}$ in all cases. These SASA and contacts data provide information on the different packing of the $\mathbf{2}_{cis}$ and $\mathbf{2}_{trans}$ conformers within the cage in increasing crowding conditions. These are consistent with the free-energy surfaces (FESs) reconstructed from the histograms extracted from the MD of Figure 7.2a, where it is evident how $\mathbf{2}_{cis}$ allows for a tighter packing within the cage (bottom). At the same time, the $\mathbf{2}_{cis}$ conformer is found less mobile within **1**, as it is demonstrated by a narrower FES dark minimum (minimum energy configuration) compared to that of the $\mathbf{2}_{trans}$ conformer (top).

As a first and generally rate-limiting step, the amide hydrolysis reaction requires the nucleophilic attack of water (or of OH^-, if the reaction occurs as catalyzed in basic conditions) to the carbonyl carbon.[413, 417] While the cavity of cage **1** is markedly apolar, and the encapsulation of the guests in the cavity is essentially driven by hydrophobic effects, some accessibility by the solvent is therefore still required for the reaction to take place.

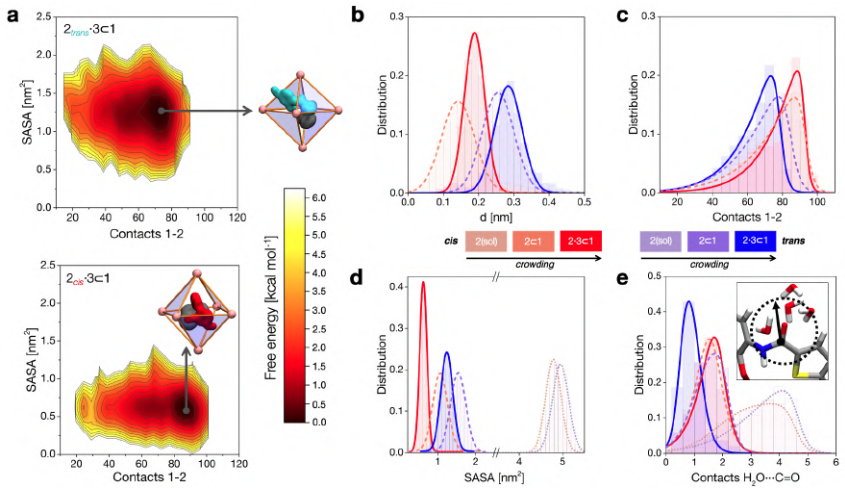


Fig. 7.2 **Into the host-guest complexes at atomistic resolution**. (**a**) Free-energy surfaces (FESs) computed from the equilibrium MD trajectories, showing the most favorable complex configurations as a function of the contacts between **1** and **2** (x axis) and of the Solvent Accessible Surface Area (SASA) of **2** within the cage (y axis). FESs are reported respectively for the $\mathbf{2}_{trans}$ (top) and $\mathbf{2}_{cis}$ (bottom) co-encapsulated with co-guest **3** in the cage. For each case, a representative scheme is shown of the encapsulated structures. (**b**) Histogram calculated from the equilibrium MD trajectories of the distance between the geometric center of cage **1** and the center of **2**. (**c**) Histogram of the contacts between **1** and **2**. (**d**) Histogram of the SASA of **2** in different complexation conditions (dotted distributions at large SASA values for the free guests in solution are reported for comparison). (**e**) Number contacts between the Oxygen atom in the amide of **2** and the water molecules in the system. A color scale is used in panels (**b-e**) to show increasing crowding conditions for **2** within the cage. *Reproduced with permission from Ref* [401].

We calculated the number of contacts between the solvent molecules and the carbonyl group of the **2** guest in the various cases. In general, we can observe that the amide of $\mathbf{2}_{trans}$ is slightly more accessible to the solvent compared to that of $\mathbf{2}_{cis}$ when the guests are free in solution (dotted distributions), and a similar depletion is observed upon mono-guest confinement (dashed). However, the situation is surprisingly switched in the tightly packed $\mathbf{2}\cdot\mathbf{3}\subset\mathbf{1}$ systems (Figure 7.2e). In highly confined conditions, the amide of $\mathbf{2}_{cis}$ is found more exposed to the solvent within cage **1** compared to that of $\mathbf{2}_{trans}$ (red *vs.* blue solid curves). This suggests that, while on the one hand the *cis* conformer of **2** is more tightly packed within the cage, the equilibrium configuration of $\mathbf{2}_{cis}$ in the $\mathbf{2}\cdot\mathbf{3}\subset\mathbf{1}$ complex may allow, at the same time, for an increased propensity to react.

7.2.2 Relationship between amide conformations and reactivity

We explore at a deeper level the difference in terms of reactivity of the *trans vs. cis* conformers of **2**. We do this by decomposing the study into two phases. First, using AA MetaD simulations, we compute the free energy landscape for all conformers accessible by the guests in and out of the cage, obtaining information on the probability of effectively visiting and observing reactive guest conformers in the cage cavity *vs.* in solution. Then, using *ab initio* MetaD simulations, we estimate the relative hydrolysis reactivity of the different accessible conformations that the guest can assume (*i.e.*, characterized by different ω values) in or out of the cage cavity.

In realistic conditions, amide bonds are usually found in *trans* configurations with a torsion angle ω close to π, with a sparse population in *cis* conformation ($\omega \sim 0$). The degree of steric conflict of the two residues flanking the amide bond is typically larger in *cis* amides, resulting, for example, in only $\sim 5-6\%$ occurrence of *cis* peptide bonds in protein structures[418, 419]. To estimate the relative probabilities for finding different conformers in different conditions, we used WT-MetaD simulations, providing free energy profiles of the ω isomerization.

The *trans*-to-*cis* isomerization of ω angles consists of a local conformation change that is often compensated by local variations of the backbone angles ϕ and ψ of the residues flanking the amide.[418] To assess how the free energy profile for the isomerization of ω is affected by the torsion of ψ and ϕ, we selected these 3 dihedral angles as our CVs and ran WT-MetaD simulations activating/biasing the *trans*-to-*cis* transition of **2** (*i.e.*, the torsion around the amide bond) in different conditions (Figure 7.3a-b).

Preliminary WT-MetaD simulations showed that the free energy profile of the ω isomerization is not particularly influenced by the ϕ and ψ torsions. Well-converged WT-MetaD runs allowed us to reconstruct the differences in free energies between the conformers (ΔG) and estimate the free energy barriers (to this end we used infrequent WT-MetaD simulations, as recrossing WT-MetaD simulations may underestimate the barrier heights, as explained in Chapter 2). The results in Figure 7.3b compare four cases where: (i) **2** is free in solution (Figure 7.3d: dotted curve, *cis* conformer in pink), (ii) **2** is encapsulated in cage **1** (dashed curve, *cis* in dark pink), (iii) the isomerizing **2** is co-encapsulated in cage **1** with another *trans* **2** guest (encapsulated **2** dimer: dot-dashed curve, *cis* in light red), and (iv) **2** is co-encapsulated in cage **1** with co-guest **3** (dot-dashed curve, *cis* in light red). The results show that the

Process	$\Delta G_{trans\rightarrow cis}$ [kcal mol^{-1}]	K_{conf}	$\Delta G^{\ddagger}_{trans\rightarrow cis}$ [kcal mol^{-1}]	$t_{trans\rightarrow cis}$ [s]	$t_{cis\rightarrow trans}$ [s]
$\mathbf{2}_{trans} \rightleftarrows \mathbf{2}_{cis}$	6.5	1.8×10^{-5}	22.3	$3.5\times10^{+3}$	1.4×10^{-1}
$\mathbf{2}_{trans} \subset \mathbf{1} \rightleftarrows \mathbf{2}_{cis} \subset \mathbf{1}$	5.0	2.3×10^{-4}	22.0	$2.3\times10^{+3}$	1.6×10^{0}
$\mathbf{2}_{trans}\cdot\mathbf{2}_{trans} \subset \mathbf{1} \rightleftarrows \mathbf{2}_{cis}\cdot\mathbf{2}_{trans} \subset \mathbf{1}^{a}$	4.0	1.2×10^{-3}	25.9	$1.5\times10^{+6}$	$4.7\times10^{+2}$
$\mathbf{2}_{trans}\cdot\mathbf{3} \subset \mathbf{1} \rightleftarrows \mathbf{2}_{cis}\cdot\mathbf{3} \subset \mathbf{1}$	2.7	1.1×10^{-2}	19.8	$5.4\times10^{+1}$	1.4×10^{0}

[a] The second $\mathbf{2}_{trans}$ guest was kept in *trans* conformation during the simulation.

Table 7.1 **Thermodynamic and kinetic data for 2 isomerization in all simulated complexes**. Free energy differences ($\Delta G_{trans\rightarrow cis}$) at 300K related to the *trans*-to-*cis* isomerization, equilibrium constants for the conformational change (K_{conf}), the height of the free-energy barriers ($\Delta G^{\ddagger}_{trans\rightarrow cis}$) from the *trans* state, characteristic timescales ($t_{trans\rightarrow cis}$ and $t_{cis\rightarrow trans}$) are reported.

stability of the conformers of **2** is significantly affected by confinement. The free energy differences between the *cis* and *trans* conformers (Figure 7.3b) indicate that, while *trans* is always the most stable configuration of the guest, the *cis* conformer is more and more stabilized as the crowding in the cage cavity increases. The transition barrier also decreases while increasing the crowding. This is captured by the ΔG and the K_{conf} values, as well as by the relative probability profiles P_{ω}^{conf} of Figure 7.3b. In particular, the P_{ω}^{conf} plots the relative probability for different conformers (ω) of **2** to the *trans* conformer in all simulated complexes. We move from a *cis*:*trans* ratio of $\sim 10^{-7}$:1 for one **2** free in solution to $\sim 10^{-5}$:1 in the mono-encapsulated system ($\mathbf{2} \subset \mathbf{1}$), $\sim 10^{-4}$:1 when two **2** guests are co-encapsulated in the cage ($\mathbf{2}_2 \subset \mathbf{1}$), to $\sim 10^{-2}$:1 in the $\mathbf{2}\cdot\mathbf{3} \subset \mathbf{1}$ system. The **2** dimer encapsulation ($\mathbf{2}_2 \subset \mathbf{1}$) system, where one of the two guests is kept fixed in a *trans* configuration in accordance with experiments,[402] falls in between the $\mathbf{2} \subset \mathbf{1}$ and $\mathbf{2}\cdot\mathbf{3} \subset \mathbf{1}$ cases. The strongly twisted at $\omega = \pi/4$ remains extremely unlikely in all systems, despite a similar thousand-fold stabilization *via* confinement. The most crowded case, $\mathbf{2}\cdot\mathbf{3} \subset \mathbf{1}$, shows a $\sim 10000\times$ increase in the probability for finding the more reactive *cis* conformer with respect to the case where **2** is free in solution. This is remarkable, considering that experimentally this case is the one showing the strongest acceleration in the hydrolysis reaction.[402] These simulations have allowed us to identify the confinement in the cage cavity as a key factor determining the relative probabilities of the different conformers of guest **2**. As a next step, we investigated the propensity of the accessible guest conformers to react.

The encapsulation of amide **2** within cage **1** was shown to enhance amide hydrolysis considerably in mildly basic conditions (experimental results of Figure 7.1b have been obtained at [NaOH] = 100 mM[402]).

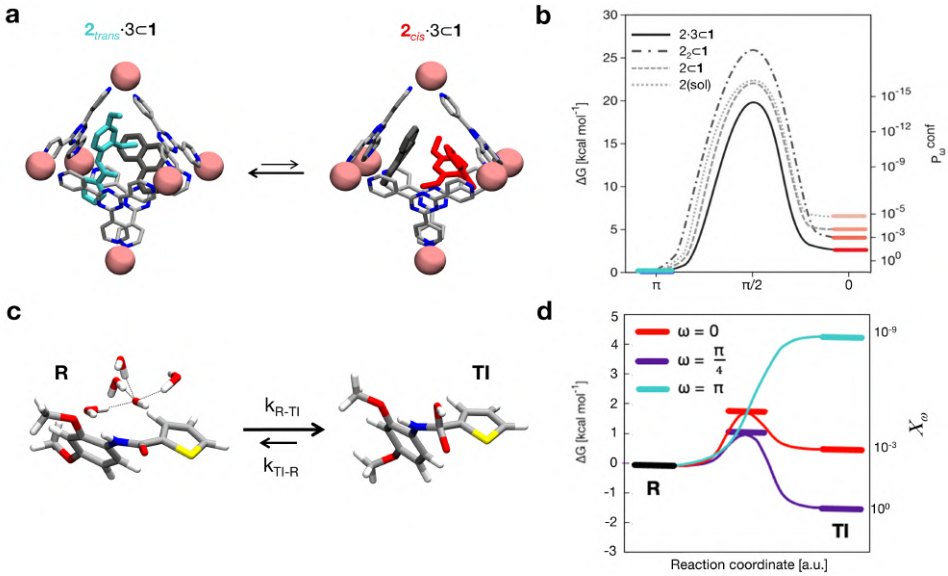


Fig. 7.3 **Conformers and reactivity of the amide 2**. (**a**) Isomerization of **2** in the cage, co-encapsulated with **3** (left: *trans*-**2**, right: *cis*-**2**). (**b**) Free energy profiles (shown as smoothed fits between the computed critical points) for the isomerization of **2** (i) when this is free in solution (dotted curve, *cis* in pink), (ii) when **2** is encapsulated in **1** (dashed curve, *cis* in dark pink), (iii) when this is co-encapsulated with another (*trans*) **2** guest in cage **1** (dot-dashed curve, *cis* in light red), and (iv) when **2** is encapsulated in **1** together with the co-guest **3** (solid curve, *cis* state in light red). The data show that increasing crowding stabilizes more and more the reactive **2** conformations (*e.g.*, *cis*) in the cage cavity. Right secondary *y* axis: relative probabilities (P_{ω}^{conf}) for the different conformations (ω) of **2** in the various host-guest systems calculated based on the ΔG values extracted from WT-MetaD simulations. (**c**) Reaction scheme modeling the first step of the hydrolysis, where the hydroxide anion (OH^-) approaches the amide group (state **R**) attacking the carbonyl group (formation of the transition intermediate **TI**). (**d**) Schemes of free energy profiles (shown as smoothed fits between the computed critical points) for hydroxide attack along the reaction coordinate as a function of the ω dihedral of **2** different values. The cyan profile refers to **2** in *trans* conformation, the red one refers to the reaction when **2** is *cis*, while the violet profile refers to the free energy profile of a *cis*-distorted configuration of **2** with $\omega = \pi/4$. A relative reactivity score for each amide conformer (χ_{ω}), normalized based on the maximum measured value (*i.e.*, that for $\omega=\pi/4$, set to 1), is associated with the simulated conformers of **2** (right secondary *y* axis). *Reproduced with permission from Ref* [401]

To simplify the investigation of the hydrolysis process, we focused on modeling the reaction of the amide only in free solution, without considering the confinement effects of the cage.In the absence of other catalysts, the hydrolysis of amides occurs via hydroxide attack, leading to the formation of a tetrahedral intermediate (TI), followed by the cleavage of the C–N bond [413].

The first step of the reaction is typically considered the rate-determining step of the process[417]. For this reason, here we focused our investigation only on the formation of the TI *via* nucleophilic attack by a solvated hydroxide (Figure 7.3c), when the amide in free in solution. Previously used, *e.g.*, to investigate the hydrolysis of formamide[417], we relied on *ab initio* well-tempered metadynamics (WT-MetaD) simulations[420] to study the reactivity of **2** as free in solution. Considering the computational cost of these simulations and the complexity of our systems, we employed a semi-empirical density-functional tight-binding (DF-TB) method,[421, 422] in its self-consistent charge corrected variant SCC-DFTB.[423] Recently shown to provide comparable accuracy to DFT with large basis sets in terms of prediction of barrier heights and reaction energies for organic molecules,[424–426] SCC-DFTB has indeed been already successfully used to investigate hydrolysis reactions in biological systems.[427, 428]

We simplified our models by studying a system with an OH^- near the amide (Figure 7.3a), and constraining the C–N–C(O)–C dihedral (ω) of amide **2** to representative values, to simulate an attack to different conformers. We compared **2** conformers with ω equal to 0 (*cis* conformer), π (*trans* conformer), and $\pi/4$ (a twisted *cis* conformer).

We employed replica infrequent *ab initio* WT-MetaD simulations[414] to obtain information on the reaction coefficients (rate of hydroxide attack and TI formation) for the various conformers of **2**. The reaction barrier of $\omega = \pi/2$ has been also tested, but this conformer was found too unstable to compute meaningful kinetic data. From multiple infrequent WT-MetaD runs activating/biasing the transition (R→TI, and the TI→R processes), we reconstructed the unbiased kinetics for the transition events and could estimate the characteristic transition times, τ_{off} and τ_{on}. The kinetic constant for hydroxide release can be calculated as $k_{off} = 1/\tau_{off}$. The kinetic constant for hydroxide attack (k_{on}) can be obtained in a similar way from the τ_{on} and accounting for the OH^- concentration in the system. In our simplified model, the OH^- ion is positioned near the amide at solvation distance and is constrained from diffusing into the solution. This setup mimics a solvation box where OH^- ions replace water molecules, effectively removing ion diffusion from the reaction site. This approximation ensures that ion availability is consistent with a pH of 14, focusing solely on the reactivity without considering diffusion effects.

This simulation approach allowed us to obtain a thermodynamic and dynamic characterization of the reaction as a function of the amide conformation (for some

relevant discrete ω values). The results of Figure 7.3d show a strong dependency of reactivity on the ω dihedral. As expected, the lowest reactivity against hydroxide attack was observed for the *trans* conformer, while the *cis* and the twisted *cis* amide conformers were found as more reactive. In particular, the latter is the only conformer with a $K_{reac} = k_{on}/k_{off} > 1$ (Figure 7.3d: **R** is higher in free energy than **TI**).

The higher reactivity of the *cis* conformer compared to the *trans* can be attributed to the fact that, in the *cis* form, the spatial arrangement of the substituents around the amide bond leads to reduced steric hindrance, allowing for easier access of the hydroxide ion to the carbonyl group. In contrast, the *trans* conformer, being more extended, results in greater steric constraints, making the nucleophilic attack more difficult. Furthermore, the twisted *cis* conformer exhibits even higher reactivity because its partial structural distortion increases the electrophilicity of the carbonyl carbon, facilitating nucleophilic attack by the hydroxide ion. This conformational twist likely destabilizes the ground state, lowering the activation barrier for the formation of the tetrahedral intermediate (TI).

In particular, from the K data, we can obtain relative reactivity scores (χ_ω) useful to compare the reactivity between the different amide conformers. The χ_ω scores of Figure 7.3d (right secondary *y* axis) clearly show how, compared to the twisted *cis* amide conformer ($\omega = \pi/4$), the *cis* amide ($\omega = 0$) is ~ 1000 times less reactive, while the reactivity of the *trans* amide conformer ($\omega = \pi$) is basically negligible ($\sim 10^9$ times less reactive than the $\pi/4$ conformer).

Altogether, the results of Figure 7.3 provide a new perspective to obtain a reactivity ranking of the conformers in the different systems, which accounts for the configurations of the amide guest which are more/less probable in or out the cage cavity and their relative reactivity. (i) The most probable conformer in all states, $\mathbf{2}_{trans}$, is also the least reactive. (ii) The most reactive twisted conformers ($\omega = \pi/4$ or, *e.g.*, $\pi/2$) are, at the same time, highly improbable, even at increased molecular crowding. (iii) The $\mathbf{2}_{cis}$ conformer, moderately reactive (but sensibly more reactive than the $\mathbf{2}_{trans}$ one), is unfavored in solution against $\mathbf{2}_{trans}$, but it becomes more and more relatively favored as the crowding increases upon confinement, emerging as the prominent reactive species in the cage. Our results then converge on the idea that the reaction acceleration by confinement is strongly governed by the other dynamic processes of the system, particularly the stabilizing effect of confinement on the more reactive conformers.

The available experimental X-ray structures for these complexes show a $\mathbf{2}_{cis}$–$\mathbf{2}_{trans}$ dimer in the $\mathbf{2}_2 \subset \mathbf{1}$ case, and a $\mathbf{2}_{cis}$-twisted conformation in the $\mathbf{2} \cdot \mathbf{3} \subset \mathbf{1}$ complex. This seems to indicate that in these complexes a $\mathbf{2}_{cis}$ conformer is more favored than $\mathbf{2}_{trans}$. While this may seem to contradict the simulation results discussed above, it is worth noting that all the results collected up to this phase are valid only under the assumption that the encapsulated guests remain always within the cage cavity. Nonetheless, these are host-guest systems, in which the probability of finding the guests within the cage obeys a well-defined supramolecular equilibrium. Estimating the effective probability of finding the guests within the cage requires also studying the dynamics of guest encapsulation/exchange in-and-out the cavity. As it will be demonstrated in the next section, accounting also for the intrinsic supramolecular dynamics of these host-guest systems provides results that are globally in very good agreement with all available experimental evidence.

7.2.3 Dynamics of amide encapsulation/expulsion in-and-out the cage cavity

Up to this point, we have compared the reactivity of the amide guest in solution versus when encapsulated in cage **1**. To complete the picture, we must consider the effective probability of reactive guest conformers in the cage cavity. Host-guest systems follow a supramolecular equilibrium where the affinity between host and guest determines the likelihood of encapsulation or expulsion. Characterizing the thermodynamics and kinetics of guest encapsulation/expulsion is crucial.

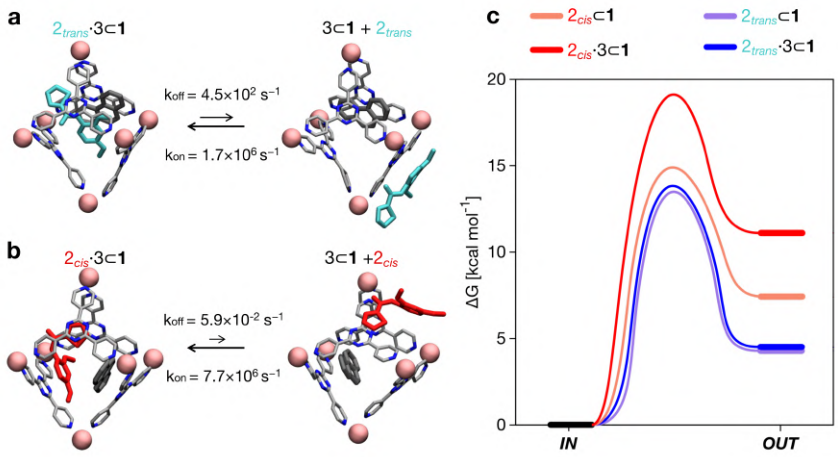


Fig. 7.4 **Equilibrium and kinetics of amide guests encapsulation/expulsion in/out the cage cavity**. (**a**) Equilibrium and kinetics for the encapsulation/expulsion of $\mathbf{2}_{trans}$ in/out the cage when **1** is hosting also guest **3**. Above and below the arrows of the equilibrium reaction are reported respectively the kinetic constants k_{off} and k_{on} estimated from the WT-MetaD simulations. The k_{on} values are also reported as considering the concentration present in the system (in brackets). (**b**) Equilibrium and kinetics for the encapsulation/expulsion of $\mathbf{2}_{cis}$ in/out the cage cavity when **1** is hosting also guest **3**. (**c**) Free energy differences and barriers (ΔG and $\Delta G^{\ddagger}$) associated to the encapsulation/expulsion in the cavity of **2** when **1** is also hosting **3** (solid curves) or when **1** does not contain any other guest (dashed curves). Free energy profiles are shown as smoothed fits between the computed critical points. *Reproduced with permission from Ref* [401].

Encapsulation/expulsion of guests like **2** or **3** in/out of cage **1** involves high free energy barriers, making them rare events within classical MD simulation timescales. Using a WT-MetaD simulation protocol, we reconstructed the thermodynamics and kinetics of amide **2** encapsulation/expulsion in cage **1**; data are summarized in Table 7.2 and Figure 7.4. Comparing the encapsulation of **2** isomers in cage **1**, we observe that $\mathbf{2}_{cis}$ encapsulation is generally more favored than $\mathbf{2}_{trans}$. The dynamics of $\mathbf{2}_{trans}$ encapsulation/expulsion are minimally affected by other guests in the cage (k_{off} and k_{on} remain similar). However, co-guests significantly impact the dynamics and stability of $\mathbf{2}_{cis}$, stabilizing its encapsulation by 2-4 orders of magnitude in the presence of **3** (lower k_{off}: slower **2** expulsion). Table 7.2 shows k_{off} for $\mathbf{2}_{cis}$ expulsion from cage **1** drops from $\sim 7.1 \times 10^{1}$ s^{-1} (with only $\mathbf{2}_{cis}$) to ~ 1 s^{-1} or $\sim 5.9 \times 10^{-2}$ s^{-1} with co-encapsulated $\mathbf{2}_{trans}$ or **3**. These data suggest that host-guest equilibrium differences are controlled by affinity between the guest (**2**) and the host cavity. From ΔG, we estimate host-guest affinity constants (K_{enc}) (Table 7.2). K_{enc} values for $\mathbf{2}_{cis}$ are generally higher than for $\mathbf{2}_{trans}$, especially in

Process	ΔG [kcal mol^{-1}]	K_{enc}	$\Delta G^{\ddagger}_{off}$ [kcal mol^{-1}]	t_{off} [s]	k_{off} [s^{-1}]	k_{on} [s^{-1}]
$\mathbf{2}_{trans} + \mathbf{1} \rightleftarrows \mathbf{2}_{trans} \subset \mathbf{1}$	−4.3	1.5×10^{3}	13.5	1.2×10^{-3}	8.3×10^{2}	1.3×10^{6}
$\mathbf{2}_{cis} + \mathbf{1} \rightleftarrows \mathbf{2}_{cis} \subset \mathbf{1}$	−7.4	2.1×10^{5}	14.9	1.4×10^{-2}	7.1×10^{1}	1.8×10^{7}
$\mathbf{2}_{trans} + \mathbf{2}_{trans} \subset \mathbf{1} \rightleftarrows \mathbf{2}_{trans} \cdot \mathbf{2}_{trans} \subset \mathbf{1}$	−4.7	2.7×10^{3}	14.5	7.3×10^{-3}	1.4×10^{2}	3.7×10^{5}
$\mathbf{2}_{cis} + \mathbf{2}_{trans} \subset \mathbf{1} \rightleftarrows \mathbf{2}_{trans} \cdot \mathbf{2}_{cis} \subset \mathbf{1}$	−9.3	6.9×10^{6}	16.7	9.6×10^{-1}	1.0×10^{0}	7.1×10^{6}
$\mathbf{2}_{trans} + \mathbf{3} \subset \mathbf{1} \rightleftarrows \mathbf{2}_{trans} \cdot \mathbf{3} \subset \mathbf{1}$	−4.5	2.2×10^{3}	13.8	2.2×10^{-3}	4.5×10^{2}	1.0×10^{6}
$\mathbf{2}_{cis} + \mathbf{3} \subset \mathbf{1} \rightleftarrows \mathbf{2}_{cis} \cdot \mathbf{3} \subset \mathbf{1}$	−11.1	1.3×10^{8}	19.1	1.7×10^{1}	5.9×10^{-2}	7.7×10^{6}

Table 7.2 **Equilibrium and kinetics of the amide encapsulation/expulsion in/out cavity**. For each simulated host-guest complex, encapsulation free energies (ΔG), equilibrium constants K_{enc}, expulsion free energy barriers ($\Delta G^{\ddagger}_{off}$), characteristic in-cavity residence times (t_{off}), and the associated transition rates (k_{off} and k_{on}) estimated from the WT-MetaD simulations are reported.

2·**3**⊂**1** complexes. The probability of $\mathbf{2}_{cis}$ co-encapsulation with **3** is $\sim 100,000\times$ higher than $\mathbf{2}_{trans}$. This explains why X-ray structures always show encapsulated $\mathbf{2}_{cis}$ guests. High K_{enc} values indicate the amide guest is generally encapsulated in the cage cavity. From thermodynamic data, we extrapolate partition probability $P^{in} = K_{enc}/(1+K_{enc})$, showing a high probability of guest encapsulation ($P^{in} \sim 1$ in all cases).

7.2.4 Molecular determinants of reactivity in dynamic host-guest systems

We have thus far investigated, separately, the three main processes that concur to define the global dynamic picture of amide hydrolysis under confinement (Figure 7.5a): (i) encapsulation/expulsion of the guests in-and-out the cage cavity; (ii) isomerization around the ω dihedral in solution and in confinement; (iii) the reaction itself. For each of them, we have devised a probability metric and observed how it varies across the different investigated cases. Our last step is thus to put these processes (and probabilities) together, in order to capture the full dynamic complexity of the system.

To this end, we can define a reaction acceleration index, a, as the ratio between the observed reactivity with or without the presence of the cage in the system – *i.e.*, when the reactant, guest **2**, is encapsulated within the cage cavity ($K(\subset \mathbf{1})$) or when it is free in solution ($K(sol)$):

$$a = K(\subset \mathbf{1})/K(sol) \tag{7.1}$$

In the real system, amide hydrolysis can in principle take place both when **2** is encapsulated in the cage cavity or when this is out of the cage (with the observed reaction coefficients determined by the probabilities for finding the different reactive conformers – planar or twisted – in the two environments).

In general, the reaction acceleration a will thus depend on the likelihood that the hydrolysis of **2** occurs in *vs.* out of the cage. From our simulations, we have seen that the conformational free energy landscape of the amide guest may change upon encapsulation (changing the relative free energy difference between *cis* and *trans* conformers). As a consequence, the probability for crossing the rotational barrier around the amide bond also changes. In particular, we could observe that the more reactive $\mathbf{2}_{cis}$ conformer is more and more stabilized as the crowding increases in the cavity of cage **1** (Figure 7.3d). The simulations also show that the encapsulation of $\mathbf{2}_{cis}$ within the cage cavity is considerably more stable than that of $\mathbf{2}_{trans}$, showing a higher affinity and retention time (Figure 7.4 and Table 7.2). Altogether, this indicates that it is more likely to observe $\mathbf{2}_{cis}$ rather than $\mathbf{2}_{trans}$ encapsulated within the cavity of the cage, which is consistent with the fact that the $\mathbf{2}_{cis}$ conformer is present in the crystal structures obtained experimentally.[402]

The reactivity in the system depends on the propensity of the visited **2** conformers to react, their relative population in the different complexes, their probability of encapsulation (*i.e.*, the relative population ratio between having **2** in the cage *vs.* in solution at the equilibrium), and the solvent molecules accessibility to the amide (*i.e.*, the solvent is another key reactants) upon encapsulation. Noteworthy, are all these parameters that can be extracted from our simulations. In general, we can define a global reaction constant for the case when hydrolysis takes place within the cage cavity, $K(\subset \mathbf{1})$, as the sum of the reaction constants ($K_\omega(\subset \mathbf{1})$) for all amide conformers (ω) visited by the guest reactant **2** in the cage cavity:

$$K(\subset \mathbf{1}) = \sum_{\omega=0}^{\pi} K_\omega(\subset \mathbf{1}) = \sum_{\omega=0}^{\pi} (\chi_\omega \cdot P_\omega^{conf}(\subset \mathbf{1}) \cdot P_\omega^{in}(\subset \mathbf{1}) \cdot NW_\omega(\subset \mathbf{1})) \qquad (7.2)$$

where χ_ω is the hydrolysis reaction constant associated to the possible amide conformers ω (see Figure 7.3b), $P_\omega^{conf}(\subset \mathbf{1})$ is the relative statistical weight for all different conformers ω in the cage cavity (Figure 7.3d), $P_\omega^{in}(\subset \mathbf{1})$ is the probability for effectively having each specific conformer ω in the reactive environment – in this case, inside the cage cavity (see Table 7.2) –, and $NW_\omega(\subset \mathbf{1})$ is the average number of contacts between solvent molecules (water or OH^-: key reactants for amide

hydrolysis) and the amide's carbonyl (estimated as in Figure 7.2e). Accordingly, the global reaction constant in the absence of the cage in the system (**2** alone in solution), $K(sol)$, can be defined as:

$$K(sol) = \sum_{\omega=0}^{\pi} K_{\omega}(sol) = \sum_{\omega=0}^{\pi} (\chi_{\omega} \cdot P_{\omega}^{conf}(sol) \cdot P_{\omega}^{in}(sol) \cdot NW_{\omega}(sol)) \tag{7.3}$$

where in this case $P_{\omega}^{conf}(sol)$ and $NW_{\omega}(sol)$ refer respectively to the relative probabilities for **2**, when alone in the solvent, to assume the different conformers ω, and the corresponding number of amide carbonyl-solvent molecules contacts. In this case, in the absence of the cage in the system, guest **2** is by definition always out of the cage, and $P_{\omega}^{in}(sol)$ = 1. Thus, Equation7.3 simplifies into:

$$K(sol) = \sum_{\omega=0}^{\pi} K_{\omega}(sol) = \sum_{\omega=0}^{\pi} (\chi_{\omega} \cdot P_{\omega}^{conf}(sol) \cdot NW_{\omega}(sol)) \tag{7.4}$$

transforming the formula for the acceleration to:

$$a = \frac{\sum_{\omega=0}^{\pi}(\chi_{\omega} \cdot P_{\omega}^{conf}(\subset \mathbf{1}) \cdot P_{\omega}^{in}(\subset \mathbf{1}) \cdot NW_{\omega}(\subset \mathbf{1}))}{\sum_{\omega=0}^{\pi}(\chi_{\omega} \cdot P_{\omega}^{conf}(sol) \cdot NW_{\omega}(sol))} \tag{7.5}$$

Moreover, it is worth noting that, given the high ΔG values in Figure 7.4c and Table 7.2, when the cage is present in the system, the guests can be also considered as always encapsulated within the cage cavity, so that in Equation 7.2 the $P^{in}(\subset \mathbf{1})$ term tends to $\sim$1 (*vide supra*):

$$K(\subset \mathbf{1}) = \sum_{\omega=0}^{\pi} K_{\omega}(\subset \mathbf{1}) = \sum_{\omega=0}^{\pi} (\chi_{\omega} \cdot P_{\omega}^{conf}(\subset \mathbf{1}) \cdot NW_{\omega}(\subset \mathbf{1})) \tag{7.6}$$

reducing the formula for the acceleration to:

$$a = \frac{\sum_{\omega=0}^{\pi}(\chi_{\omega} \cdot P_{\omega}^{conf}(\subset \mathbf{1}) \cdot NW_{\omega}(\subset \mathbf{1}))}{\sum_{\omega=0}^{\pi}(\chi_{\omega} \cdot P_{\omega}^{conf}(sol) \cdot NW_{\omega}(sol))} \tag{7.7}$$

This means that, given the high propensity to guest encapsulation ($P^{in}(\subset \mathbf{1})$ $\sim$1), in this specific case the reaction acceleration in the system is found to be little dependent on the guest encapsulation/expulsion equilibrium. On the other hand, the reactivity turns out to be rather controlled by the fact the guest is more favored to

assume reactive conformations inside the cage cavity (compared to the case when this is free in solution). This is in full agreement with the available experimental evidence on these systems[402]. Finally, it is worth noting that while the summations in Equation 7.6 and Equation 7.4 run in principle over all possible values of ω (which have different reactivities, as seen in Figure 7.3b), the data in Figure 7.3d clearly show that, due to the intrinsically high isomerization barrier, the relative probability of observing twisted (and extremely reactive) **2** conformers (*e.g.*, $\pi/2$, $\pi/4$, etc.) is very low. These are distorted, very unstable conformers with a survival lifetime tending to zero, for which the product $\chi_\omega \cdot P^{conf}\omega \approx 0$. The unique conformers with a non-zero survival life and $P^{conf}\omega \neq 0$ are $\mathbf{2}_{cis}$ ($\omega = 0$) and $\mathbf{2}_{trans}$ ($\omega = \pi$). The latter, however, is substantially non-reactive (Figure 7.3b shows $\chi_\pi \approx 0$), so that also in this case $\chi_\pi \cdot P^{conf}\pi \approx 0$. Based on these observations, in our case, the reactivity of the system seems to be largely related to (i) how much the reactive **2***cis* conformer is over-stabilized and (ii) how accessible the amide is to the co-reactant solvent molecules in such a conformation within the cage cavity *vs.* in solution. Combining these data, we estimate the reaction acceleration a for the various host-guest complexes reported in Figure 7.5b). We observe that, in this case, the reactivity increases with the crowding in the system. While a $\sim$26-fold acceleration is computed for the mono-encapsulated case ($\mathbf{2} \subset \mathbf{1}$), a double-encapsulation gives a $\sim$64-fold increase for the $\mathbf{2}_2 \subset \mathbf{1}$ system. A dramatic $a \sim 150$ is obtained for the $\mathbf{2} \cdot \mathbf{3} \subset \mathbf{1}$ complex. While such estimated a values may differ quantitatively from those obtained from the experiments (this can be expected, given the deviations of such ideal models from realistic systems/conditions), the trends can be still safely compared. Figure 7.5c shows a remarkable trend between our calculated acceleration data and the experimental ones. This validates our simulation approach. It is worth noting that the mono-encapsulation case ($\mathbf{2} \subset \mathbf{1}$) does not have an experimental counterpart, due to the tendency of **2** to dimerize within the cage. Nonetheless, this extra-case (where crowding is lower than in, *e.g.*, $\mathbf{2}_2 \subset \mathbf{1}$ and $\mathbf{2} \cdot \mathbf{3} \subset \mathbf{1}$) provides an additional case useful for comparison. In particular, the limited computed acceleration seen in this case supports the evidence that molecular crowding within the cage cavity is a key player in the reactivity in the host-guest system.

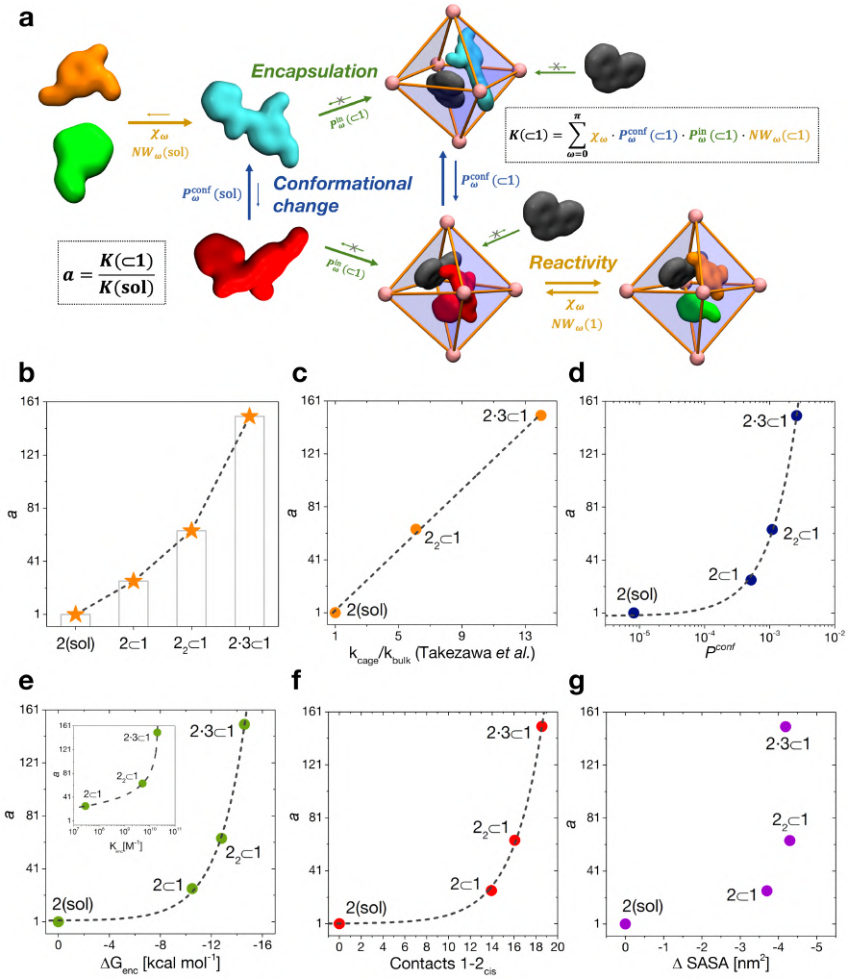


Fig. 7.5 **Reaction acceleration in a dynamic host-guest system**. (**a**) Full dynamic representation scheme, showing the processes that need to be taken into account to rationalize the reaction acceleration observed experimentally. (**b**) Computed reaction acceleration for the various investigated host-guest systems: the hydrolysis acceleration index, a, is expressed relative to the reactivity of the **2** guest alone in solution (see Equations 7.1-to-7.6). (**c**) Correlation between the reaction acceleration, a, computed from the simulations, and the acceleration measured experimentally[402] (linear fit reported by the dashed line). (**d-h**) Relationships between the computed reaction acceleration a and various characteristic parameters of the host-guest systems: (**d**) relative probability for finding the $\mathbf{2}_{cis}$ over $\mathbf{2}_{trans}$ conformer in solution *vs.* in the different host-guest complexes (dashed line: exponential fit); (**e**) encapsulation free energy ΔG_{enc} of the $\mathbf{2}_{cis}$ conformer in the different complexes (exponential fit reported by the dashed line); (**f**) the weighted number of contacts between the host and the guest, evaluated as the product between the peak position of the distribution of Figure 7.2c and its height; (**g**) the reduction in SASA of the $\mathbf{2}_{cis}$ guest in the different encapsulation complexes *vs.* when this is in solution (an indirect measure of solvophobic effect[429, 430], showing no clear correlation). *Reproduced with permission from Ref* [401].

In order to obtain an insight into the key molecular determinants controlling the reaction acceleration in these host-guest systems, in Figures 7.5d-g we plot the computed *a* parameters against some of their key constitutive terms. We have seen that the difference in affinity between *cis* and *trans* conformers among the different systems is the main factor determining the final reaction acceleration using Equation7.1, this being shown by the nearly perfect exponential correlation between *a* and the relative probability of finding the $\mathbf{2}_{cis}$ conformer with respect to the $\mathbf{2}_{trans}$ conformer in solution and in the different encapsulation complexes (Figure7.5d). The trend suggests that small incremental stabilizing effects on this conformation, *e.g.* by changing affinity and size of the co-guest, could result in potentially outstanding enhancements of reactivity for guest **2**, keeping all other parameters constant. Noteworthy, a quasi-exponential trend is observed between the computed reaction acceleration *a* and the encapsulation free energies (ΔG_{enc}) for $\mathbf{2}_{cis}$ in all systems (Figure 7.5e). In these systems, where the reactivity is observed to increase with the crowding inside the cage cavity, the *a* is also clearly related to the host-guest interaction (namely, to obtain a stable complexation, a strong host-guest affinity is necessary to compensate for the crowding penalty associated to the binding). This affinity, as evidenced by a qualitative investigation of how the different components interact with each other, both in our simulations and in the available experimental X-ray structures,[402] is mainly driven by π-π stacking between the four triazine ligands of cage **1** and the aromatic rings of amide **2**, particularly favorable in the *cis* conformation and by residual solvophobic interactions between such ring groups. A similar trend can be observed also when looking at the weighted number of contacts between the host and the guest (Figure 7.5f, evaluated from the distributions of Figure 7.2c, *i.e.* a proxy for the host-guest interaction energy). If we consider the interaction between **2** and cage **1** to be consistent among all the investigated systems, we can trace the trend back to the interaction between guest and co-guest (or the absence thereof in the $\mathbf{2} \subset \mathbf{1}$ case), with **3** showing a greater stabilizing effect for $\mathbf{2}_{cis}$ within the cage cavity compared to another **2** co-guest. Noteworthy, as revealed by the obtained trends of Figures 7.5e-f, such favorable affinity can stabilize the reactive conformer of amide **2** to a higher extent, which results in a remarkable increase in the reaction acceleration *a*. The remarkable stabilizing effect of co-guest **3** is driven by a balance between having an aromatic structure able to participate in π-π stacking with guest **2**, its considerable solvophobicity that makes its encapsulation within the cage cavity very stable, and a size (volume = 172 Å^3)[401] that allows the amide to reside

comfortably in the cavity in a *cis* (reactive) conformation. Another **2** co-guest, while still being able to engage in favorable interactions (Table 2: host-guest complexation on average more favorable than that of a single **2**), has a larger volume (241 $Å^3$)[401] which constraints the first guest in less favorable configurations and produce a less stable complex.

To obtain information on how much of the host-guest interaction is due to solvophobic effects, we calculated the reduction in the Solvent Accessible Surface Area (SASA) of the $\mathbf{2}_{cis}$ conformer when this is encapsulated in the cavity of the various complexes *vs.* when this is alone in solution.[429, 409, 430] While a correlation with the computed reaction acceleration is observed (see Figure 7.5g), the trend becomes less neat. The trend is respected while moving from the amide in solution to mono-guest ($\mathbf{2} \subset \mathbf{1}$) and double-guest complexes ($\mathbf{2}_2 \subset \mathbf{1}$ and $\mathbf{2} \cdot \mathbf{3} \subset \mathbf{1}$). However, the differences in acceleration between the various systems do not correlate in a neat manner with the ΔSASA calculated for the various cases. This reveals that (i) non-specific hydrophobic effects alone are not sufficient to grasp the complexity of these reactive systems, and suggests that (ii), like in most of receptor-ligand complexes in Nature, specific molecular interactions are probably relevant in controlling the host-guest affinity.

7.3 Conclusions

Using multi-scale modeling and metadynamics simulations, we study the host-guest system's intrinsic dynamics and the amide hydrolysis reaction. This allows us to (i) characterize hydrolysis barriers as a function of amide conformation, (ii) study the amide guest conformations and their probabilities inside and outside the cage, and (iii) analyze guest encapsulation/expulsion dynamics. Our approach estimates a reaction acceleration score, comparing system variants and identifying the stabilization of reactive amide conformers due to crowding effects in the cage.

In our case study, we compared the amide guest alone, encapsulated in the cage, and co-encapsulated with other guests. We found that encapsulation stabilizes reactive amide conformers, with higher encapsulation constants favoring this stabilization. When co-encapsulated, increased crowding further stabilizes the reactive *cis* conformer. Our computational results align well with experimental findings by Takezawa *et al.*[402], emerging from a detailed study of molecular dynamics and equilibria in

the host-guest system. This general approach explains confinement-induced reaction acceleration (or deceleration) by considering all dynamic processes within the system, providing a high-resolution framework for building structure-property relationships. It supports the rational design of host-guest systems with molecular-level control of reactivity, exploring the effects of guest and cage properties, and tuning host-guest dynamics for reactivity control.

While our method effectively reconstructs dynamic reactivity, it has limitations, such as not considering possible active roles of intermediate states or the explicit effect of the cage on hydrolysis. The cage's effect is implicitly included as encapsulation favors reactive conformers or limits solvent access. Methodological extensions, like hybrid QM/MM approaches, could include the cage explicitly if necessary. Despite simplicity, our approach links the computational chemistry of chemical reactions to complex dynamic processes typical of chemical engineering, highlighting the importance of controlling system dynamics for reactivity.

7.4 Methods

7.4.1 Reparametrization of ω dihedral potential terms of amide

Peptide bonds in protein structures are mainly found in *trans* conformation with a torsion angle ω close to π. Due to the small population of the *cis* conformation [418], the kinetics of isomerization of peptide bond has always been difficult to characterize experimentally and, as a consequence, only limited data on the free energy barriers separating the two isomers are available.[431] Being generally derived from quantum mechanical calculations on model compounds, or based on experimental data from thermodynamic and kinetic studies, the AMBER force field shows limitations in the estimation of the accurate free energy difference between the *cis* and *trans* conformations in peptide bond. Herein, this was evidenced and proven by preliminary metadynamics tests in this sense. As showed in section 7.3, the description of the *cis-trans* equilibrium for the amide **2** has a central role in the study of its reactivity, and thereby, our aim is to evaluate the accuracy of the potential terms involved in the process of isomerization of the ω dihedral angle. In the AMBER force field, the torsional potential energy term is expressed as the Pitzer potential[432], a Fourier

series term given by:

$$\sum \frac{V_n}{2}[1 + cos(n\mu - \gamma)] \tag{7.8}$$

where V_n, n, μ, and γ are, respectively, the dihedral force constants, periodicity, torsional angle, and phase angle. To reproduce the torsional energy profile of the peptide bond, only the first two terms in this series are relevant: V_1 (n=1, $\gamma = 0$), and V_2 (n=2, $\gamma = 180$); the former describes the *cis-trans* equilibra, while the latter is responsible for the barriers to rotation about C-N bond. In order to test the accuracy of the dihedral amide parameters, we compared the free-energy surface (FES) for the torsion of an N-methylacetamide (NMA) amide – the simplest analog of the peptide bond within amides, that here we use as a reference to optimize the amide force field parameters – obtained performing WT-MetaD simulation with the experimental energy profile obtained with NMR studies.[431] In these WT-MetaD simulations, as the collective variable (CV) we chose the ω dihedral angle of the amide, with a bias factor of 30 deposited every 500 steps using Gaussians of initial height of 1.2 kcal/mol and a σ of 0.35 rad.

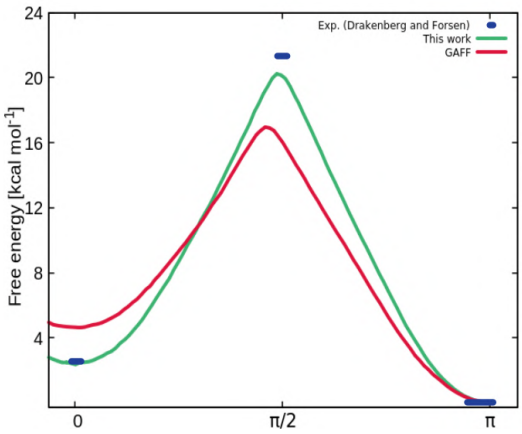


Fig. 7.6 **Comparison of Free Energy Surfaces for the torsion of ω for NMA.** We compared the FES for the torsion of an NMA amide obtained with our reparametrized dihedral potential terms (in green) or with the original GAFF parameters (in red) *vs.* the experiments (blue)[431]. The experimental free energy barriers for the transitions are 21.3 kcal/mol for trans to cis and 18.8 kcal/mol for cis to trans. Using the original GAFF parameters, the barriers are higher. Our reparametrized dihedral potential terms yield barriers of 20.3 kcal/mol for trans to cis and 18.0 kcal/mol for cis to trans, showing improved agreement with experimental values. *Adapted with permission from* Ref[401].

Our simulations indicate that the default ω AMBER potential parameters do not accurately reproduce the *cis-trans* equilibrium and free energy barriers for

isomerization, as compared to experimental estimates (Figure 7.6). To improve the description of ω isomerization, we replaced the standard AMBER force field dihedral potential values with those pre-optimized by Doshi and Hamelberg [419], and further optimized them to align with experimental data (Figure 7.6). Specifically, we adjusted the dihedral force field parameters to match our *cis-trans* isomerization FES with the experimental free energy profile for NMA (Figure 7.6). This involved modifying the main torsional parameters for ω bond rotation, focusing on the V_2 term for the general X-C-N-C torsion, and the V_1 and $V_{2'}$ potential parameters for the H-N-C-O torsion, while keeping all other non-bonded parameters unchanged.

7.4.2 Conformational Metadynamic simulations

In order to characterize the ΔG associated to the *trans*-to-*cis* transition of **2**, we employed WT-MetaD simulations, biasing the 3 dihedral angles – ω, ψ and ϕ for the four systems, **2** in solution, $\mathbf{2} \subset \mathbf{1}$, $\mathbf{2}_2 \subset \mathbf{1}$ and $\mathbf{2} \cdot \mathbf{3} \subset \mathbf{1}$. The starting model conformations used in the WT-MetaD simulations of these complexes have been obtained after 1 μs of preliminary NPT MD equilibration. During this WT-MetaD run, in all simulated systems amide **2** was seen to undergo transition *trans*-to-*cis*, recrossing between the isomers multiple times. For all simulated complexes, we ran at least 1 μs of WT-MetaD simulation to reach convergence in the calculated *trans*-to-*cis* ΔG. In these WT-MetaD simulations the bias was deposited every 1000 steps (2 ps of simulation time) using Gaussians of initial height of 1.2 kcal/mol, σ of 0.33 rad, with bias factor of 25-35 depending on the system. For the characterization of the barriers and kinetics of the *trans*-to-*cis* isomerization, we turned to infrequent WT-MetaD simulations. We ran 51 infrequent WT-MetaD simulations, focusing on the isomerization along ω and ψ dihedral angles for each of our host-guest systems. This allowed us to compare how the encapsulation, and then the molecular crowding inside the cage cavity, affect the dynamics of isomerization of **2**. In these infrequent WT-MetaD runs, the bias was deposited every 5000 steps (10 ps of simulation time) using Gaussians of initial height of 1.2 kcal/mol, σ of 0.23 rad, with bias factors between 6-16 depending on the system.

Chapter 8

Conclusions

This thesis summarized and described various computational studies conducted during my PhD, highlighting the critical role of dynamics in determining the physical and chemical properties of materials. In the works presented here, we integrated advanced computational methods, particularly ML algorithms combined with molecular simulation techniques, in order to develop a general and robust framework for analyzing the atomic structural dynamics of metal surfaces and NPs. This integration is essential for managing the complex, high-dimensional data generated in these studies, providing unprecedented insights into the atomic-level behavior of these materials. The initial focus was on Cu-FCC surfaces. To fully and accurately characterize these systems, we trained a neural network potential with DFT-level data. This approach allows us to characterize the complex microscopic dynamic behavior of these surfaces, leading to the introduction of the concept of "*statistical identity*" for our metallic systems. In this context high-dimensional descriptors like SOAP were crucial in characterizing atomic environments, essential for both the *bottom-up* and *top-down* analyses described in Chapters 3, 4, and 5[172, 171, 342].

Our research extended to Au nanoparticles, demonstrating the versatility of our framework. We created a comprehensive dictionary of atomic environments, enabling continuous enrichment and detailed analysis of individual atoms in MD simulations. Then, collaborating with experimental groups, we aligned our models with practical scenarios, providing a comprehensive understanding of the system's dynamics and structural properties. Our methodology enabled us to reconstruct the structural dynamics of metal NPs with Å spatial resolution and *ps* time resolution over 6 *s* intervals [342, 331]. By using local descriptors like LENS[203], we captured

rare dynamic events, such as the collective sliding of atomic terrace on the NP face, providing a dynamic view of atomic NPs useful for interpreting properties under experimental conditions. The necessity of complementing a structural descriptor like SOAP with a dynamic one, LENS, is also demonstrated in Chapter 6. Relying solely on structure-based descriptors, these approaches have limitations due to the inherent nature of the descriptors themselves. For example, SOAP is permutationally invariant, which restricts its ability to detect local events that do not involve significant structural rearrangements but are still important for understanding the physical properties of the system. Moreover, while this type of descriptor effectively detects dominant structural environments, it may fail to capture sparsely observed, local time-dependent events within the trajectory. These transitions, although statistically infrequent, play a crucial role in the overall behavior and properties of the system, as we observed in Chapter 6. Integrating dynamic descriptors allows for a more comprehensive understanding of these fleeting yet significant transitions, ensuring a more complete and deeper prediction of system properties.

This thesis also investigates how the dynamics of a (host-guest) system control reactivity. This has been demonstrated in a case of amide hydrolysis in the cavity of a coordination cage[401]. This approach is affine to our work on metals: in both cases, we demonstrate how dynamic behaviors are crucial for understanding systems properties. In particular, in this specific host-guest system, we revealed how dynamic processes, such as guest encapsulation and expulsion, significantly influence reactivity. By coupling dynamic equilibrium with reaction constants, we obtained a comprehensive understanding of how structural environments dictate dynamic behaviors. As with metals, static analyses based only on structure fail to capture these essential dynamic phenomena, highlighting the limitations of structure-based approaches in fully understanding the behavior of complex systems.

Overall, this thesis discusses a general approach to unravel and characterize the intricate dynamic nature of complex systems, demonstrating the limitations of studying them as static structures, even at low temperatures. By analyzing atomic dynamics, we determine energy barriers, reconstruct the free energy surface, and establish probability distributions and lifetimes of various states. This direct examination of microscopic dynamics allows us to accurately determine macroscopic properties, ensuring, for example, a deeper understanding of metal surface behaviors under relevant conditions. In general, the works discussed herein represent a new viewpoint (or, in a sense a paradigm shift), moving the main subject from static

structure and energy to microscopic dynamics. This generally advances our understanding of how complex dynamical systems, from atomic crystals and metals to supramolecular systems, work, providing insights that are not possible to attain by employing solely traditional structural and energy-based analysis.

Future Perspectives

The results obtained from this research open several avenues for future investigations. One key area is the study of the reactivity and activity of atomic environments, as it has also been recently shown by others[185, 201, 400], which is particularly important for any property that is finely controlled by the available surface or structural sites over time.

In general, different atomic sites and environments exhibit distinct properties and behaviors,[161, 163, 250, 251] making it crucial to obtain a structural/dynamical map of how long all AEs in the NPs live (τ_i) and how quickly they interconvert ($\tau_{i\to j}$). Understanding these dynamics is key to assessing their effective performance in various applications. Statistically, if an AE's average lifetime τ_i is shorter than the timescale for a process $\tau_{process}$, the probability of utilizing that AE is proportional to $\tau_i/\tau_{process}$. This ratio indicates how often a process must interact with the same AE to be effective.

Estimating this in practical scenarios involves focusing on realistic cases and evaluating the behavior of all visited AEs in the NP. Although this is not the main focus of this thesis, it is certainly feasible and highlights the potential of the approach for future research. Moreover, it is important to note that such a probabilistic interpretation is valid as long as the interacting species do not significantly alter the dynamics and features of the AEs present on the NP—*e.g.*, no or negligible chemisorption (if this condition does not hold, a proper reactive parametrization and simulation of the system is needed, where new AEs may appear on the NP surface upon interaction with the species [341, 185, 201]). The methodologies developed can also be extended to other metal systems, allowing for broader application of the findings. Our approach has proven to be adaptable to different types of metals, and developing comprehensive SOAP dictionaries for various metals will enable a more detailed understanding of their dynamic behavior. Creating a unified dictionary that includes AEs from different systems could be particularly beneficial for comparing a wide range of metals. This would facilitate the identification of common patterns

and unique behaviors across different metallic systems, enhancing our ability to design materials with specific desired properties. Moreover, these approaches are so abstract and general that they can also be applied to completely different systems on different scales[205, 203, 433, 363]. This opens interesting new perspectives for revisiting traditional methods of classifying condensed matter systems and achieving a deeper level of understanding in materials science overall.

References

[1] W. D. Callister, D. G. Rethwisch, A. Blicblau, K. Bruggeman, M. Cortie, J. Long, J. Hart, R. Marceau, and R. Mitchell. *Materials science and engineering: an introduction*, volume 7. John wiley & sons, 2007.

[2] R. E. Smallman and R. J. Bishop. *Modern physical metallurgy and materials engineering*. Butterworth-Heinemann, 1999.

[3] D. D. Pollock. *Physical properties of materials for engineers*. CRC press, 2020.

[4] G. Gottstein. *Physical foundations of materials science*, volume 3. Springer, 2004.

[5] R. J. Tilley. *Understanding solids: the science of materials*. John Wiley & Sons, 2004.

[6] L. Solymar, D. Walsh, and R. R. Syms. *Electrical properties of materials*. Oxford university press, 2014.

[7] J. Singleton. *Band theory and electronic properties of solids*, volume 2. OUP Oxford, 2001.

[8] A. Abrikosov. *Fundamentals of the Theory of Metals*. Courier Dover Publications, 2017.

[9] J. C. Slater. The electronic structure of metals. *Rev. Mod. Phys.*, 6(4):209, 1934.

[10] V. K. Grigorovich. *The metallic bond and the structure of metals*. Commack, NY (USA); Nova Science Publishers, Inc., 1989.

[11] B. Cantor, P. Grant, and C. Johnson. *Aerospace Materials*. CRC Press, 2001.

[12] X. Zhang, Y. Chen, and J. Hu. Recent advances in the development of aerospace materials. *Prog. Aerosp. Sci.*, 97:22–34, 2018.

[13] L. Zhu, N. Li, and P. Childs. Light-weighting in aerospace component and system design. *Propuls. Power Res.*, 7(2):103–119, 2018.

[14] R. Boyer, J. Cotton, M. Mohaghegh, and R. Schafrik. Materials considerations for aerospace applications. *MRS Bull.*, 40(12):1055–1066, 2015.

[15] Richard O. Davis and Donald S. Olcott. *Properties of Copper and Copper Alloys: Electrical and Mechanical*. International Copper Research Association, 1967.

[16] F. Stelzer and A. Puschmann. Electrical contacts made of silver and its alloys for power application. *J. Mater. Sci.*, 46(3):717–722, 2011.

[17] D. Heller et al. Advanced high strength steel (ahss) applications and challenges in automotive industry. *JMEP*, 28(3):1236–1248, 2019.

[18] Markus Gehring and Matthias Fiebig. The role of high-strength steels in vehicle safety. *Automot. Eng. Int.*, 126:30–35, 2018.

[19] T. Sherman. *Applications of High-Strength Steels in Automotive Safety Components*, pages 121–135. Springer, 2015.

[20] M. Vaněk and J. C. Dunlop. Metals for solar photovoltaic cells and their market impacts. *Sol. Energy*, 159:57–66, 2016.

[21] N. Kumar. *Steel and Copper in Renewable Energy Applications*, pages 171–186. Elsevier, 2017.

[22] I. Pimenta and S. T. Pinho. Materials selection for sustainable wind turbine blades. *Renew. Sustain. Energy Rev.*, 92:306–318, 2018.

[23] William F. Smith. *Principles of Materials Science and Engineering*. McGraw-Hill, 1986.

[24] E. Nes and K. Marthinsen. Modeling the evolution in microstructure and properties during plastic deformation of fcc-metals and alloys–an approach towards a unified model. *Mater. Sci. Eng. A.*, 322(1-2):176–193, 2002.

[25] G. Meyers and K. Chawla. Plastic deformation in metals: A review of fundamental mechanisms. *Int. J. Plast.*, 20(4):341–355, 1984.

[26] N. F. Mott and W. D. Nix. The deformation of metals under stress. *Prog. Mater. Sci.*, 37(3):123–134, 1992.

[27] A. Brown. Surface effects in plastic deformation of metals. *Adv. Phys.*, 1(4):427–479, 1952.

[28] R. J. Asaro. Crystal Plasticity. *J. Appl. Mech.*, 50(4b):921–934, 1983.

[29] J. R. Davis. *Alloying: understanding the basics*. ASM international, 2001.

[30] P. Gilman and J. Benjamin. Mechanical alloying. *Annu. Rev. Mater. Sci.*, 13(1):279–300, 1983.

[31] C. Suryanarayana. Mechanical alloying and milling. *Prog. Mater. Sci.*, 46(2):1184, 2001.

[32] C. Suryanarayana, E. Ivanov, and V. Boldyrev. The science and technology of mechanical alloying. *Mater. Sci. Eng. A.*, 304:151–158, 2001.

[33] Serope Kalpakjian and Steven Schmid. *Manufacturing Processes for Engineering Materials*. Pearson, 2010.

[34] M. P. Groover. *Fundamentals of modern manufacturing: materials, processes, and systems*. John Wiley & Sons, 2010.

[35] B. Bhushan and P. L. Ko. Introduction to tribology. *Appl. Mech. Rev.*, 56(1):6–7, 2003.

[36] Y. Meng, J. Xu, Z. Jin, B. Prakash, and Y. Hu. A review of recent advances in tribology. *Friction*, 8:221–300, 2020.

[37] A. Rosenkranz, M. C. Righi, A. V. Sumant, B. Anasori, and V. N. Mochalin. Perspectives of 2d mxene tribology. *Adv. Mater.*, 35(5):2207757, 2023.

[38] P. Restuccia and M. C. Righi. Tribochemistry of graphene on iron and its possible role in lubrication of steel. *Carbon*, 106:118–124, 2016.

[39] G. Ertl, Helmut Knözinger, Jens Weitkamp, et al. *Handbook of heterogeneous catalysis*, volume 2. VCH Weinheim, 1997.

[40] R. H. Crabtree. *The organometallic chemistry of the transition metals*. John Wiley & Sons, 2009.

[41] D. Astruc. *Nanoparticles and catalysis*. John Wiley & Sons, 2008.

[42] R. Jin. The impacts of nanotechnology on catalysis by precious metal nanoparticles. *Nanotechnol. Rev.*, 1(1):31–56, 2012.

[43] M. Haruta and Masakazu Daté. Advances in the catalysis of Au nanoparticles. *Appl. Catal. A Gen.*, 222(1-2):427–437, 2001.

[44] G. Chinchen, K. Waugh, and D. Whan. The activity and state of the copper surface in methanol synthesis catalysts. *Appl. Catal.*, 25(1-2):101–107, 1986.

[45] G. Chinchen and K. Waugh. Chemical state of copper during methanol synthesis. *J. Catal.*, 97(1), 1986.

[46] M. Bowker, R. Hadden, H. Houghton, J. Hyland, and K. Waugh. The mechanism of methanol synthesis on copper/zinc oxide/alumina catalysts. *J. Catal*, 109(2):263–273, 1988.

[47] S. Natesakhawat, J. W. Lekse, J. P. Baltrus, P. R. Ohodnicki Jr, B. H. Howard, X. Deng, and C. Matranga. Active sites and structure–activity relationships of copper-based catalysts for carbon dioxide hydrogenation to methanol. *ACS Catal.*, 2(8):1667–1676, 2012.

[48] K. Stangeland, H. Li, and Z. Yu. Co2 hydrogenation to methanol: the structure–activity relationships of different catalyst systems. *Energy Ecol. Environ.*, 5(4):272–285, 2020.

[49] I. P. Beletskaya and A. V. Cheprakov. The heck reaction as a sharpening stone of palladium catalysis. *Chem. Rev.*, 100(8):3009–3066, 2000.

[50] N. Lopez, T. Janssens, B. Clausen, Y. Xu, M. Mavrikakis, T. Bligaard, and Jens Kehlet Nørskov. On the origin of the catalytic activity of gold nanoparticles for low-temperature co oxidation. *J. Catal*, 223(1):232–235, 2004.

[51] J. Spivey. *Metal nanoparticles for catalysis: advances and applications*. Royal society of chemistry, 2014.

[52] C. Gao, F. Lyu, and Y. Yin. Encapsulated metal nanoparticles for catalysis. *Chem. Rev.*, 121(2):834–881, 2020.

[53] Y. Khan, H. Sadia, S. Shah, M. Khan, and A. Shah. Classification, synthetic, and characterization approaches to nanoparticles, and their applications in various fields of nanotechnology: a review. *Catalysts*, 12(11):1386, 2022.

[54] F. Baletto and R. Ferrando. Structural properties of nanoclusters: Energetic, thermodynamic, and kinetic effects. *Rev. Mod. Phys.*, 77(1):371, 2005.

[55] L. Guo, P. Du, X. Fu, C. Ma, and J. Zeng. Contributions of distinct gold species to catalytic reactivity for carbon monoxide oxidation. *Nat. Commun.*, 7:13481, 2016.

[56] C. Tseng, T. Yang, H. Wu, and H. Chiang. Catalysis of oxidation of carbon monoxide on supported gold nanoparticle. *J. Hazard. Mater.*, 166(2):686–694, 2009.

[57] S. Overbury, N. Dudney, and M. Li. Gold nanoparticles supported on carbon nitride: influence of surface hydroxyls on low temperature carbon monoxide oxidation. *ACS Catal.*, 2(6):1138–1146, 2014.

[58] P. Sabatier. *La catalyse en chimie organique*, volume 3. C. Béranger, 1920.

[59] F. Vines, J. Gomes, and F. Illas. Understanding the reactivity of metallic nanoparticles: beyond the extended surface model for catalysis. *Chem. Soc. Rev.*, 43(14):4922–4939, 2014.

[60] D. Uzio and G. Berhault. Factors governing the catalytic reactivity of metallic nanoparticles. *Catal. Rev.*, 52(1):106–131, 2010.

[61] G. Pacchioni and H. Freund. Controlling the charge state of supported nanoparticles in catalysis: lessons from model systems. *Chem. Soc. Rev.*, 47(22):8474–8502, 2018.

[62] F. Quddus, A. Shah, F. Iftikhar, N. Shah, and A. Haleem. Environmentally benign nanoparticles for the photocatalytic degradation of pharmaceutical drugs. *Catalysts*, 13(3):511, 2023.

[63] S. K. Murthy. Nanoparticles in modern medicine: state of the art and future challenges. *Int J Nanomedicine*, 2(2):129–141, 2007.

[64] S. Lanone and J. Boczkowski. Biomedical applications and potential health risks of nanomaterials:molecular mechanisms. *Curr. Mol. Med.*, 6(6):651–663, 2006.

[65] R. Varma. Journey on greener pathways: from the use of alternate energy inputs and benign reaction media to sustainable applications of nano-catalysts in synthesis and environmental remediation. *Green Chem.*, 16(5):2027–2041, 2014.

[66] D. Wang, D. Kim, K. Choi, J. Seo, and S. Im. Enhancement of donor–acceptor polymer bulk heterojunction solar cell power conversion efficiencies by addition of Au nanoparticles. *Angew. Chem. Int. Ed.*, 123(24):5633–5637, 2011.

[67] M. Alkhalayfeh, A. Aziz, and M. Pakhuruddin. An overview of enhanced polymer solar cells with embedded plasmonic nanoparticles. *Renew. Sustain. Energy Rev.*, 141:110726, 2021.

[68] L. Liang, M. Liu, Z. Jin, Q. Wang, H. Wang, et al. Optical management with nanoparticles for a light conversion efficiency enhancement in inorganic γ-cspbi3 solar cells. *Nano Lett.*, 19(3):1796–1804, 2019.

[69] U. Chadha, S. Selvaraj, H. Ashokan, et al. Complex nanomaterials in catalysis for chemically significant applications: from synthesis and hydrocarbon processing to renewable energy applications. *Mater. sci. eng.*, 2022(1):1552334, 2022.

[70] M. A. Albrecht, C. W. Evans, and C. L. Raston. Green chemistry and the health implications of nanoparticles. *Green Chem.*, 8(5):417–432, 2006.

[71] F. Zaera. In-situ and operando spectroscopies for the characterization of catalysts and of mechanisms of catalytic reactions. *J. Catal*, 404:900–910, 2021.

[72] Y. Li, D. Zakharov, S. Zhao, R. Tappero, and U. Jung. Complex structural dynamics of nanocatalysts revealed in operando conditions by correlated imaging and spectroscopy probes. *Nat. Commun.*, 6:7583, 2015.

[73] Y. Nian, X. Huang, M. Liu, J. Zhang, and Y. Han. Insight into the dynamic evolution of supported metal catalysts by In situ/operando techniques and theoretical simulations. *ACS Catal.*, 13:11164–11171, 2023.

[74] D. Goodman. Correlations between surface science models and "real-world" catalysts. *J. Phys. Chem.*, page 13090–13102, 1996.

[75] C. Jayanthi, E. Tosatti, and L. Pietronero. Surface melting of copper. *Phys. Rev. B*, 31(6):3456, 1985.

[76] M. Bernasconi and E. Tosatti. Reconstruction, disordering and roughening of metal surfaces. *Surf. Sci. Rep.*, 17(7-8):363–422, 1993.

[77] R. Dovesi, B. Civalleri, C. Roetti, V. R. Saunders, and R. Orlando. Ab initio quantum simulation in solid state chemistry. *Reviews in Comp. Chem.*, 21:1–125, 2005.

[78] H. Cheng, S. Wang, G. Chen, Z. Liu, D. Caracciolo, M. Madiou, S. Shan, J. Zhang, H. He, R. Che, et al. Insights into heterogeneous catalysts under reaction conditions by in situ/operando electron microscopy. *Adv. Energy Mater.*, 12(38):2202097, 2022.

[79] X. Shi, X. Lin, R. Luo, S. Wu, L. Li, Z. Zhao, and J. Gong. Dynamics of heterogeneous catalytic processes at operando conditions. *JACS Au*, 1(12):2100–2120, 2021.

[80] K. F Kalz, R. Kraehnert, M. Dvoyashkin, R. Dittmeyer, R. Gläser, U. Krewer, K. Reuter, and J. Grunwaldt. Future challenges in heterogeneous catalysis: understanding catalysts under dynamic reaction conditions. *ChemCatChem*, 9(1):17–29, 2017.

[81] V. Colic and A. Bandarenka. Pt alloy electrocatalysts for the oxygen reduction reaction: from model surfaces to nanostructured systems. *ACS Catal.*, 6(8):5378–5385, 2016.

[82] S. Matera, W. Schneider, A. Heyden, and A. Savara. Progress in accurate chemical kinetic modeling, simulations, and parameter estimation for heterogeneous catalysis. *ACS Catal.*, 9(8):6624–6647, 2019.

[83] S. Chee, J. Arce-Ramos, W. Li, and A. Genest. Structural changes in noble metal nanoparticles during co oxidation and their impact on catalyst activity. *Nat. Commun.*, 11:2133, 2020.

[84] H. Mao, B. Chen, J. Chen, K. Li, and J. Lin. Recent advances in high-pressure science and technology. *MRE*, 1(1):59–75, 2016.

[85] Y. Han, H. Zhang, Y. Yu, and Z. Liu. In situ characterization of catalysis and electrocatalysis using apxps. *ACS Catal.*, 11(3):1464–1484, 2021.

[86] K. Jensen, C. Tyrsted, and M. Bremholm. In situ studies of solvothermal synthesis of energy materials. *ChemSusChem*, 7(6):1594–1611, 2014.

[87] Y. Yang, Y. Xiong, R. Zeng, X. Lu, M. Krumov, X. Huang, W. Xu, H. Wang, F. J. DiSalvo, J. D. Brock, et al. Operando methods in electrocatalysis. *ACS Catal.*, 11(3):1136–1178, 2021.

[88] A. Gross. *Theoretical surface science*. Springer, 2003.

[89] D. E. Fowler and J. V. Barth. Structure and dynamics of the Cu(001) surface investigated by medium-energy ion scattering. *Phys. Rev. B*, 52:2117–2124, Jul 1995.

[90] S. Å. Lindgren, L. Walldén, J. Rundgren, and P. Westrin. Low-energy electron diffraction from Cu(111): Subthreshold effect and energy-dependent inner potential; surface relaxation and metric distances between spectra. *Phys. Rev. B*, 29:576–588, Jan 1984.

[91] W. Lomer. Defects in pure metals. *Progress in metal physics*, 8:255–320, 1959.

[92] X. Li and K. Lu. Playing with defects in metals. *Nat. Mater.*, 16:700–701, 2017.

[93] CP Flynn. Point defect reactions at surfaces and in bulk metals. *Phys. Rev. B.*, 71(8), 2005.

[94] Y. Cao, S. Ni, X. Liao, M. Song, and Y. Zhu. Structural evolutions of metallic materials processed by severe plastic deformation. *Mater. Sci. Eng. R Rep.*, 133:1–59, 2018.

[95] Y. Sheng, Y. Hua, X. Wang, X. Zhao, L. Chen, and H. Zhou. Application of high-density electropulsing to improve the performance of metallic materials: Mechanisms, microstructure and properties. *Materials*, 11(2):185, 2018.

[96] L. Sun, G. Wu, Q. Wang, and J. Lu. Nanostructural metallic materials: Structures and mechanical properties. *Mater. Today*, 38:114–135, 2020.

[97] G. Gourdin and V. Doan-Nguyen. In situ, operando characterization of materials for electrochemical devices. *Cell Rep. Phys. Sci.*, 2(12), 2021.

[98] S. Nayak, C. Hung, V. Sharma, and S. Alpay. Insight into point defects and impurities in titanium from first principles. *npj Comput. Mater.*, 4:11, 2018.

[99] E. Lu, J. Zhao, I. Makkonen, et al. Enhancement of vacancy diffusion by C and N interstitials in the equiatomic femnnicocr high entropy alloy. *Acta Mater.*, 215:117093, 2021.

[100] Z. Zhang, A. Ghasemi, N. Koutná, et al. Correlating point defects with mechanical properties in nanocrystalline tin thin films. *Mater. Des.*, 207:109844, 2021.

[101] J Čížek. Characterization of lattice defects in metallic materials by positron annihilation spectroscopy: A review. *J. Mater. Sci. & Technology*, 34(4):577–598, 2018.

[102] K. Saarinen, P. Hautojärvi, and C. Corbel. Positron annihilation spectroscopy of defects in semiconductors. In *Semiconductors and Semimetals*, volume 51, pages 209–285. Elsevier, 1998.

[103] T. Broom. Lattice defects and the electrical resistivity of metals. *Adv. Phys.*, 3(9):26–83, 1954.

[104] J. Martin. The electrical resistivity due to structural defects. *Philos. Mag.*, 24(189):555–566, 1971.

[105] A. Karolik and A. Luhvich. Calculation of electrical resistivity produced by dislocations and grain boundaries in metals. *J. Phys. Condens. Matter*, 6(4):873, 1994.

[106] Zoltan Balogh and Guido Schmitz. Diffusion in metals and alloys. In *Physical metallurgy*, pages 387–559. Elsevier, 2014.

[107] L. Lu, Q. Pan, K. Hattar, and B. L Boyce. Fatigue and fracture of nanostructured metals and alloys. *MRS Bulletin*, 46:258–264, 2021.

[108] Y. Xu and L. Wang. In situ characterization techniques for understanding point defects in metals. *Mater. Today*, 11(3):1464–1484, 2020.

[109] Jin Fu, Heng Li, Xu Song, and MW Fu. Multi-scale defects in powder-based additively manufactured metals and alloys. *J. Mater. Sci. Technol.*, 122:165–199, 2022.

[110] G. Somorjai. Surface reconstruction and catalysis. *Annu. Rev. Phys. Chem.*, 45(1):721–751, 1994.

[111] F. Ercolessi, E. Tosatti, and M. Parrinello. Au (100) surface reconstruction. *Phys. Rev. Lett.*, 57(6):719, 1986.

[112] E Crljen, P. Lazić, Damir Šokčević, and Radovan Brako. Relaxation and reconstruction on (111) surfaces of Au, Pt, and Cu. *Phys. Rev. B*, 68(19):195411, 2003.

[113] F. Hanke and Jonas Björk. Structure and local reactivity of the Au (111) surface reconstruction. *Phys. Rev. B*, 87(23):235422, 2013.

[114] P. Carnevali, F. Ercolessi, and E. Tosatti. Melting and nonmelting behavior of the Au (111) surface. *Phys. Rev. B*, 36(12):6701, 1987.

[115] F. Polo-Garzon, Z. Bao, X. Zhang, W. Huang, and Z. Wu. Surface reconstructions of metal oxides and the consequences on catalytic chemistry. *ACS Catal.*, 9(6):5692–5707, 2019.

[116] H. Topsøe. Developments in operando studies and in situ characterization of heterogeneous catalysts. *Journal of Catalysis*, 216(2):155–164, 2003.

[117] J. H. Pfisterer, Y. Liang, O. Schneider, and A. S. Bandarenka. Direct instrumental identification of catalytically active surface sites. *Nature*, 549:74–77, 2017.

[118] C. Barroo, Z Wang, R. Schlögl, and M. Willinger. Imaging the dynamics of catalysed surface reactions by in situ scanning electron microscopy. *Nat. Catal.*, 3:30–39, 2020.

[119] R. Yuan, H. Jiao, H. Zhu, D. Fang, and S. Jiao. In situ characterization techniques and methodologies for high-temperature electrochemistry. *Chem*, 9(9):2481–2508, 2023.

[120] Maoyu Wang, Líney Árnadóttir, Zhichuan J Xu, and Zhenxing Feng. In situ X-ray absorption spectroscopy studies of nanoscale electrocatalysts. *Nano-Micro Letters*, 11:18, 2019.

[121] J. Timoshenko and B. Roldan Cuenya. In situ/operando electrocatalyst characterization by X-ray absorption spectroscopy. *Chem. Rev.*, 121(2):882–961, 2020.

[122] Christina HM van Oversteeg, Hoang Q Doan, Frank MF de Groot, and Tanja Cuk. In situ X-ray absorption spectroscopy of transition metal based water oxidation catalysts. *Chem. Soc. Rev.*, 46(1):102–125, 2017.

[123] P. Sedigh Rahimabadi and M. Khodaei. Review on applications of synchrotron-based X-ray techniques in materials characterization. *X-Ray Spectrom.*, 49(3):348–373, 2020.

[124] J. Schnadt, J. Knudsen, and N. Johansson. Present and new frontiers in materials research by ambient pressure x-ray photoelectron spectroscopy. *J. Phys. Condens. Matter*, 32(41):413003, 2020.

[125] J. B. Wagner, F. Cavalca, C. D. Damsgaard, L. D. Duchstein, and T. W. Hansen. Exploring the environmental transmission electron microscope. *Micron*, 43(11):1169–1175, 2012.

[126] Z. L. Wang. New developments in transmission electron microscopy for nanotechnology. *Adv. Mater.*, 15(18):1497–1514, 2003.

[127] D. A. Muller. Structure and bonding at the atomic scale by scanning transmission electron microscopy. *Nat. Mater.*, 8:263–270, 2009.

[128] S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy. High-throughput computational materials scouting. *Nat. Mater.*, 12(3):191–201, 2013.

[129] J. Mianroodi, P. Shanthraj, C. Liu, and S. Vakili. Modeling and simulation of microstructure in metallic systems based on multi-physics approaches. *npj Comput. Mater.*, 8:93, 2022.

[130] J. Keith, V. Vassilev-Galindo, and B. Cheng. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.*, 121(16):9816–9872, 2021.

[131] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864, 1964.

[132] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):1133, 1965.

[133] R. Jones and O. Gunnarsson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61(3):689, 1989.

[134] J. Hafner. Ab-initio simulations of materials properties. *Rep. Prog. Phys.*, 51(3):451, 1987.

[135] J. Rogal, M. W. Finnis, and A. Gross. Density-functional theory in materials science. *Comput. Mater. Sci.*, 158:112–119, 2019.

[136] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55(22):2471, 1985.

[137] S. R. Billeter, A. Curioni, and W. Andreoni. Ab initio molecular dynamics simulation of the melting of aluminum. *Phys. Rev. B*, 67(6):064209, 2003.

[138] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.*, 45(14):1196, 1980.

[139] S. M. Foiles, M. Baskes, and M. S. Daw. Calculation of the vibrational properties of grain boundaries in metals. *Phys. Rev. B*, 33(12):7983, 1986.

[140] M. S. Daw and M. I. Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B*, 29(12):6443, 1984.

[141] M. I. Baskes. Modified embedded-atom potentials for cubic materials and impurities. *Phys. Rev. B*, 46:2727–2742, 1992.

[142] M. Finnis and J. Sinclair. A simple empirical N-body potential for transition metals. *Phil. Mag. A*, 50(1):45–55, 1984.

[143] A. Carlsson. Beyond pair potentials in elemental transition metals. *Solid State Phys.*, 43:1–91, 1990.

[144] P. Schlexer, V. Degirmenci, M. Xu, et al. First-principles-based multiscale modelling of heterogeneous catalysis. *Nat. Rev. Mater.*, 3:144–164, 2018.

[145] Y. Zhai, D. A. Lados, and J. L. LaGoy. Additive manufacturing: making imagination the major limitation. *JOM*, 66:808–816, 2014.

[146] T. Gatsos, K. A. Elsayed, Y. Zhai, and D. A. Lados. Review on computational modeling of process–microstructure–property relationships in metal additive manufacturing. *JOM*, 72:403–419, 2020.

[147] Sophie Loehlé and Maria Clelia Righi. Ab initio molecular dynamics simulation of tribochemical reactions involving phosphorus additives at sliding iron interfaces. *Lubricants*, 6(2):31, 2018.

[148] S. G. Louie, Y. Chan, F. H. da Jornada, Z. Li, and D. Y. Qiu. Discovering and understanding materials through computation. *Nat. Mater.*, 20:728–735, 2021.

[149] K. D. Vogiatzis, M. V. Polynski, J. K. Kirkland, J. Townsend, A. Hashemi, C. Liu, and E. A. Pidko. Computational approach to molecular catalysis by 3d transition metals: challenges and opportunities. *Chem. Rev*, 119(4):2453–2523, 2018.

[150] J. N. Harvey, F. Himo, F. Maseras, and L. Perrin. Scope and challenge of computational methods for studying mechanism and reactivity in homogeneous catalysis. *ACS Catal.*, 9(8):6803–6813, 2019.

[151] S. D. Senanayake, J. A. Rodriguez, D. Stacchiola, J. Hrbek, and E. Fujita. In situ and in operando characterization of catalytic materials under realistic reaction conditions. *ACS Catal.*, 1(10):1047–1064, 2011.

[152] F. Tao and R. M. Crooks. In situ and operando techniques for characterization of catalysts. *ACS Catal.*, 10:1047–1054, 2010.

[153] A. C. van Duin, S. P. Dasgupta, C. D. Lorenz, and W. A. Goddard III. Chemistry at the metal–water interface: How surface electrons affect the structure of water layers. *J. Phys. Chem. A*, 102(38):7364–7372, 1998.

[154] W. D. Jones. Computational studies of catalytic reactions in nanomaterials. *Acc. Chem. Res.*, 48(3):676–685, 2015.

[155] S. Bahn and K. Jacobsen. Beyond self-diffusion: Molecular dynamics studies of diffusion in multicomponent systems using the green-kubo formalism. *Phys. Rev. Lett.*, 87(26):266101, 2002.

[156] J. Rossmeisl, Z. Qu, H. Zhu, G. Kroes, and Jens K Nørskov. Electrolysis of water on oxide surfaces. *J. Electroanal. Chem.*, 607(1):83–89, 2008.

[157] R. Lavroff, H. Morgan, Z. Zhang, P. Poths, et al. Ensemble representation of catalytic interfaces: soloists, orchestras, and everything in-between. *Chem. Sci.*, 13:2225–2241, 2022.

[158] Z. Zhang, B. Zandkarimi, and A. N. Alexandrova. Ensembles of metastable states govern heterogeneous catalysis on dynamic interfaces. *Acc. Chem. Res.*, 53(2):447–458, 2020.

[159] W. Li et al. Critical role of thermal fluctuations for co binding on electrocatalytic metal surfaces. *JACS Au*, 1(10):1708–1718, 2021.

[160] L. Grajciar et al. Towards operando computational modeling in heterogeneous catalysis. *Chem. Soc. Rev.*, 47(22):8307–8348, 2018.

[161] E. Gazzarrini, K. Rossi, and F. Baletto. Born to be different: the formation process of Cu nanoparticles tunes the size trend of the activity for CO_2 to CH_4 conversion. *Nanoscale*, 13(11):5857–5867, 2021.

[162] V. Van Speybroeck, M. Bocus, P. Cnudde, and L. Vanduyfhuys. Operando modeling of zeolite-catalyzed reactions using first-principles molecular dynamics simulations. *ACS Catal.*, 13(9):11455–11493, 2023.

[163] Y. Wang, D. Mei, V. Glezakou, J. Li, and R. Rousseau. Dynamic formation of single-atom catalytic active sites on ceria-supported gold nanoparticles. *Nat. Commun.*, 6:6511, 2015.

[164] J. Mortensen, L. Hansen, B. Hammer, and J. Nørskov. Nitrogen adsorption and dissociation on Fe (111). *J. Catal.*, 182(2):479–488, 1999.

[165] R. Pucci and G. Angilella. Density functional theory, chemical reactivity, and the fukui functions. *Found. Chem.*, 24(1):59–71, 2022.

[166] L. Piccolo. Surface studies of catalysis by metals: nanosize and alloying effects. In *Nanoalloys: synthesis, structure and properties*, pages 369–404. Springer, 2012.

[167] C. R. Henry. Surface studies of supported model catalysts. *Surf. Sci. Rep.*, 31(7-8):231–325, 1998.

[168] S. Satoh, H. Fujimoto, and H. Kobayashi. Theoretical study of NH3 adsorption on Fe(110) and Fe(111) surfaces. *J. Phys. Chem. B*, 110(10):4846–4852, 2006.

[169] S. C. Yeo, S. S. Han, and H. M. Lee. Mechanistic investigation of the catalytic decomposition of ammonia NH3 on an Fe(100) surface: a DFT study. *J. Phys. Chem. C*, 118(10):5309–5316, 2014.

[170] Jacques C Védrine. Heterogeneous catalysis on metal oxides. *Catalysts*, 7(11):341, 2017.

[171] D. Rapetti, M. Delle Piane, M. Cioni, D. Polino, R. Ferrando, and G. M. Pavan. Machine learning of atomic dynamics and statistical surface identities in gold nanoparticles. *Commun. Chem.*, 6:143, 2023.

[172] M. Cioni, D. Polino, D. Rapetti, L. Pesce, M. Delle Piane, and G. M. Pavan. Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments. *J. Chem. Phys.*, 158(12):124701, 2023.

[173] D. Nelli. Central vacancy creation in icosahedral nanoparticles induced by the displacement of large impurities. *Eur. Phys. J. Appl. Phys.*, 97:18, 2022.

[174] H. E. Sauceda, L. E Gálvez González, S. Chmiela, L. Paz-Borbón, K. Müller, and A. Tkatchenko. Bigdml—towards accurate quantum machine learning force fields for materials. *Nat. Commun.*, 13:3733, 2022.

[175] V. Botu, R. Batra, J. Chapman, and R. Ramprasad. Machine learning force fields: construction, validation, and outlook. *J. Phys. Chem. C*, 121(1):511–522, 2017.

[176] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schutt, A. Tkatchenko, and K. Muller. Machine learning force fields. *Chem. Rev.*, 121(16):10142–10186, 2021.

[177] C. Zeni, K. Rossi, A. Glielmo, A. Fekete, N. Gaston, F. Baletto, and A. De Vita. Building machine learning force fields for nanoclusters. *J. Chem. Phys.*, 148(24), 2018.

[178] Y. Mishin. Machine-learning interatomic potentials for materials science. *Acta Mater.*, 214:116980, 2021.

[179] M. Yang, L. Bonati, D. Polino, and M. Parrinello. Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water. *Catal. Today*, 387:143–149, 2022.

[180] H. Wang, L. Zhang, J. Han, and W. E. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.*, 228:178–184, 2018.

[181] F. Noé, A. Tkatchenko, K. Müller, and C. Clementi. Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.*, 71:361–390, 2020.

[182] I. Poltavsky and A. Tkatchenko. Machine learning force fields: Recent advances and remaining challenges. *J. Phys. Chem. Lett.*, 12(28):6551–6564, 2021.

[183] J. Behler. Machine learning potentials for molecular simulation: Applications to atoms and molecules. *J. Chem. Phys.*, 145(17):170901, 2017.

[184] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, Kristof T Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.*, 3(5), 2017.

[185] L. Bonati, D. Polino, C. Pizzolitto, P. Biasi, R. Eckert, S. Reitmeier, R. Schlögl, and M. Parrinello. The role of dynamics in heterogeneous catalysis: Surface diffusivity and N2 decomposition on Fe (111). *Proc. Natl. Acad. Sci. U.S.A.*, 120(50):e2313023120, 2023.

[186] J. Behler and M. Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, 2007.

[187] J. Behler. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.*, 145(17):170901, 2016.

[188] J. S. Smith, O. Isayev, and A. E. Roitberg. Ani-1: An extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.*, 8(4):3192–3203, 2017.

[189] T. Mueller, A. Hernandez, and C. Wang. Machine learning for interatomic potential models. *J. Chem. Phys.*, 152(5):050902, 2020.

[190] R. Kondor A. P. Bartók, M. C. Payne and G. Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104:136403, 2010.

[191] C. R. Trott S. M. Foiles A. P. Thompson, L. P. Swiler and G. J. Tucker. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.*, 285:316–330, 2015.

[192] A. Shapeev. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Model. Simul.*, 14(3):1153–1173, 2016.

[193] J. Chapman S. Krishnan L. Chen T. D. Huan, R. Batra and R. Ramprasad. A universal strategy for the creation of machine learning-based atomistic force fields. *npj Comput. Mater.*, 3:37, 2017.

[194] N. Artrith and A. Urban. An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for tio2. *Comput. Mater. Sci.*, 114:135–150, 2016.

[195] J. R. K. Z. Li and A. De Vita. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.*, 114(9):096405, 2015.

[196] H. Wang R. Car L. Zhang, J. Han and E. Weinan. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.*, 120(14):143001, 2018.

[197] R. M. Balabin and E. I. Lomakina. Support vector machine regression (ls-svm)—An alternative to artificial neural networks (anns) for the analysis of quantum chemistry data? *Phys. Chem. Chem. Phys.*, 13(24):11710–11718, 2011.

[198] G. Pilania A. Mannodi-Kanakkithodi R. Ramprasad, R. Batra and C. Kim. Machine learning in materials informatics: Recent applications and prospects. *npj Comput. Mater.*, 3(1):54, 2017.

[199] C. Poelking N. Bernstein J. R. Kermode G. Csányi A. P. Bartók, S. De and M. Ceriotti. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.*, 3(12):e1701816, 2017.

[200] M. Spencer. Stable and metastable metal surfaces in heterogeneous catalysis. *Nature*, 323:685–687, 1986.

[201] S. Perego, L. Bonati, S. Tripathi, and M. Parrinello. How dynamics changes ammonia cracking on iron surfaces. *ChemRxiv*, 2024.

[202] A. P. Bartók, R. Kondor, and G. Csányi. On representing chemical environments. *Phys. Rev. B*, 87:184115, 2013.

[203] M. Crippa, A. Cardellini, C. Caruso, and G. M. Pavan. Detecting dynamic domains and local fluctuations in complex molecular systems via timelapse neighbors shuffling. *Proc. Natl. Acad. Sci. U.S.A.*, 120(30):e2300565120, 2023.

[204] P. Gasparotto, D. Bochicchio, M. Ceriotti, and G. M. Pavan. Identifying and tracking defects in dynamic supramolecular polymers. *J. Phys. Chem. B*, 124(3):589–599, 2019.

[205] M. Crippa, C. Perego, A. L. de Marco, and G. M. Pavan. Molecular communications in complex systems of dynamic supramolecular polymers. *Nat. Commun.*, 13:2162, 2022.

[206] W. Kohn. Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.*, 71(5):1253, 1999.

[207] T. Helgaker, P. Jorgensen, and J. Olsen. *Molecular electronic-structure theory*. John Wiley & Sons, 2013.

[208] V. Heine. The pseudopotential concept. *Solid State Phys.*, 24:1–36, 1970.

[209] M. L. Cohen and V. Heine. The fitting of pseudopotentials to experimental data and their subsequent application. In *Solid state physics*, volume 24, pages 37–248. Elsevier, 1970.

[210] J. Ihm, A. Zunger, and M. L. Cohen. Momentum-space formalism for the total energy of solids. *J. Phys. Condens. Matter*, 12(21):4409, 1979.

[211] H. J. Monkhorst and J. D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188, 1976.

[212] H. A. Lorentz. Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase. *Annalen der physik*, 248(1):127–136, 1881.

[213] C. L. Kong. Combining rules for intermolecular potential parameters. ii. rules for the lennard-jones (12–6) potential and the morse potential. *J. Chem. Phys.*, 59(5):2464–2467, 1973.

[214] P. M. Morse. Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Phys. Rev.*, 34(1):57, 1929.

[215] R. A. Buckingham. The classical equation of state of gaseous Helium, Neon and Argon. *Proc. R. Soc. Lond. A.*, 168(933):264–283, 1938.

[216] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der physik*, 369(3):253–287, 1921.

[217] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comput. Phys.*, 73(2):325–348, 1987.

[218] M. J. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. Stewart. Development and use of quantum mechanical molecular models. 76. am1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107(13):3902–3909, 1985.

[219] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995.

[220] B. R. Brooks et al. Charmm: the biomolecular simulation program. *J. Comput. Chem.*, 30(10):1545–1614, 2009.

[221] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *J. Comput. Chem.*, 25(13):1656–1676, 2004.

[222] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174, 2004.

[223] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *J. Phys. Chem.*, 97(40):10269–10280, 1993.

[224] C. Liu, J. Cohen, J. Adams, and A. Voter. Eam study of surface self-diffusion of single adatoms of fcc metals Ni, Cu, Al, Ag, Au, Pd, and Pt. *Surf. Sci.*, 253(1):334–344, 1991.

[225] M.I. Mendelev, M.J. Kramer, C.A. Becker, and M. Asta. Analysis of semi-empirical interatomic potentials appropriate for simulation of crystalline and liquid Al and Cu. *Philos. Mag.*, 88(12):1723–1750, 2008.

[226] Y. Mishin, M. Mehl, D. Papaconstantopoulos, A. Voter, and J. Kress. Structural stability and lattice defects in copper: Ab initio, tight-binding, and embedded-atom calculations. *Phys. Rev. B*, 63(22):224106, 2001.

[227] R. P. Gupta. Lattice relaxation at a metal surface. *Phys. Rev. B*, 23:6265–6270, 1981.

[228] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20):12562–12566, 2002.

[229] A. Barducci, M. Bonomi, and M. Parrinello. Metadynamics. *WIREs Comput. Mol. Sci.*, 1(5):826–843, 2011.

[230] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100(2):020603, 2008.

[231] P. Tiwary and M. Parrinello. A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B*, 119(3):736–742, 2015.

[232] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti. Bond-orientational order in liquids and glasses. *Phys. Rev. B*, 28(2):784–805, 1983.

[233] W. Lechner and C. Dellago. Accurate determination of crystal structures based on averaged local bond order parameters. *J. Chem. Phys.*, 129(11):114707, 2008.

[234] M. Crippa, A. Cardellini, M. Cioni, G. Csányi, and G. M Pavan. Machine learning of microscopic structure-dynamics relationships in complex molecular systems. *Mach. learn.: sci. technol.*, 4(4):045044, 2023.

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

[236] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen. Towards the computational design of solid catalysts. *Nat. Chem.*, 1(1):37–46, 2009.

[237] Jens K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard. Density functional theory in surface chemistry and catalysis. *Proc. Natl. Acad. Sci. U.S.A.*, 108(3):937–943, 2011.

[238] F. Calle-Vallejo, J. Tymoczko, V. Colic, Q. H. Vu, M. D. Pohl, K. Morgenstern, D. Loffreda, P.e Sautet, W. Schuhmann, and A.r S. Bandarenka. Finding optimal surface sites on heterogeneous catalysts by counting nearest neighbors. *Science*, 350(6257):185–189, 2015.

[239] B. W. J. Chen, L. Xu, and M. Mavrikakis. Computational methods in heterogeneous catalysis. *Chem. Rev.*, 121(2):1007–1048, 2021.

[240] A. Trayanov, A. Levi, and E. Tosatti. Anisotropic roughening theory of the (110) faces of Cu, Ni, Pd, and Ag. *Surf. Sci.*, 233(1-2):184–207, 1990.

[241] H Dürr, R Schneider, and Th Fauster. Anomalous dynamical behavior of the Cu (110) surface. *Phys. Rev. B*, 43(15):12187, 1991.

[242] J. Merikoski, H Häkkinen, Matti Manninen, J Timonen, and Kimmo Kaski. Disordering mechanisms of the Cu (110) surface. *Int. J. Mod. Phys. B*, 8(23):3175–3204, 1994.

[243] O. Zeybek. Thermal behavior of the roughened Cu (110) surface. *Solid State Commun.*, 139(7):339–344, 2006.

[244] V. Yamakov, D. Wolf, S. Phillpot, A. Mukherjee, and H. Gleiter. Deformation-mechanism map for nanocrystalline metals by molecular-dynamics simulation. *Nat. Mat.*, 3:43–47, 2004.

[245] L. A. Zepeda-Ruiz, A. Stukowski, T. Oppelstrup, and V. V. Bulatov. Probing the limits of metal plasticity with molecular dynamics simulations. *Nature*, 550:492–495, 2017.

[246] X. Wang, S. Zheng, S. Shinzato, Z. Fang, Y. He, L. Zhong, C. Wang, S. Ogata, and S. X. Mao. Atomistic processes of surface-diffusion-induced abnormal softening in nanoscale metallic crystals. *Nat. Commun.*, 12:5237, 2021.

[247] D. Nelli, F. Pietrucci, and R. Ferrando. Impurity diffusion in magic-size icosahedral clusters. *J. Chem. Phys.*, 155(14):144304, 2021.

[248] A. Gardin, C. Perego, G. Doni, and G. M. Pavan. Classifying soft self-assembled materials via unsupervised machine learning of defects. *Commun. Chem.*, 5:82, 2022.

[249] D. Bochicchio and G. M. Pavan. From cooperative self-assembly to water-soluble supramolecular polymers using coarse-grained simulations. *ACS nano*, 11(1):1000–1011, 2017.

[250] S. Xie, S. Choi, X. Xia, and Y. Xia. Catalysis on faceted noble-metal nanocrystals: both shape and size matter. *Curr. Opin. Chem. Eng.*, 2(2):142–150, 2013.

[251] F. Dattila, R. Garcıá-Muelas, and N. López. Active and selective ensembles in oxide-derived copper catalysts for CO_2 reduction. *ACS Energy Lett.*, 5(10):3176–3184, 2020.

[252] S. Doerr, M. Majewski, A. Pérez, A. Krämer, C. Clementi, F. Noe, T. Giorgino, and G. De Fabritiis. Torchmd: A deep learning framework for molecular simulations. *J. Chem. Theory Comput.*, 17(4):2355–2363, 2021.

[253] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. Schnet – A deep learning architecture for molecules and materials. *J. Chem. Phys.*, 148(24):241722, 2018.

[254] O. T. Unke and M. Meuwly. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.*, 15(6):3678–3693, 2019.

[255] Z. Li, J. R. Kermode, and A. De Vita. Molecular dynamics with On-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.*, 114:096405, 2015.

[256] G. P. Purja Pun, R. Batra, R. Ramprasad, and Y. Mishin. Physically informed artificial neural networks for atomistic modeling of materials. *Nat. Commun.*, 10:2339, 2019.

[257] 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. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A*, 124(4):731–745, 2020.

[258] S. De, A. P. Bartók, G. Csányi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.*, 18(20):13754–13769, 2016.

[259] R. Capelli, F. Muniz-Miranda, and G. M. Pavan. Ephemeral ice-like local environments in classical rigid models of liquid water. *J. Chem. Phys.*, 156(21):214503, 2022.

[260] T. Bian, A. Gardin, J. Gemen, L. Houben, C. Perego, B. Lee, N. Elad, Z. Chu, G. M. Pavan, and R. Klajn. Electrostatic co-assembly of nanoparticles with oppositely charged small molecules into static and dynamic superstructures. *Nat. Chem.*, 13:940–949, 2021.

[261] R. Capelli, A. Gardin, C. Empereur-Mot, G. Doni, and G. M. Pavan. A data-driven dimensionality reduction approach to compare and classify lipid force fields. *J. Phys. Chem. B*, 125(28):7785–7796, 2021.

[262] C. Zeni, K. Rossi, T. Pavloudis, J. Kioseoglou, S. de Gironcoli, R. E. Palmer, and F. Baletto. Data-driven simulation and characterisation of gold nanoparticle melting. *Nat. Commun.*, 12:6056, 2021.

[263] 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. Progress and perspectives of electrochemical CO_2 reduction on copper in aqueous electrolyte. *Chem. Rev.*, 119(12):7610–7672, 2019.

[264] K. Manthiram, B. J. Beberwyck, and A. P. Alivisatos. Enhanced electrochemical methanation of carbon dioxide with a dispersible nanoscale copper catalyst. *J. Am. Chem. Soc.*, 136(38):13319–13325, 2014.

[265] A. J. Hickman and M. S. Sanford. High-valent organometallic copper and palladium in catalysis. *Nature*, 484:177–185, 2012.

[266] S. Magdassi, M. Grouchko, and A. Kamyshny. Copper nanoparticles for printed electronics: Routes towards achieving oxidation stability. *Materials*, 3(9):4626–4638, 2010.

[267] Lei L., Y. Shen, X. Chen, L. Qian, and K. Lu. Ultrahigh strength and high electrical conductivity in copper. *Science*, 304(5669):422–426, 2004.

[268] M. Behrens, F. Studt, I. Kasatkin, S. Kühl, M. Hävecker, F. Abild-Pedersen, S. Zander, F. Girgsdies, P. Kurr, B. Kniep, et al. The active site of methanol synthesis over $Cu/ZnO/Al_2O_3$ industrial catalysts. *Science*, 336(6083):893–897, 2012.

[269] M. D. Higham, M. G. Quesne, and C. R. A. Catlow. Mechanism of CO_2 conversion to methanol over Cu (110) and Cu (100) surfaces. *Dalton Trans.*, 49(25):8478–8497, 2020.

[270] S. M. Foiles, M. I. Baskes, and M. S. Daw. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. *Phys. Rev. B*, 33:7983–7991, 1986.

[271] V. Rosato, M. Guillope, and B. Legrand. Thermodynamical and structural properties of f.c.c. transition metals using a simple tight-binding model. *Philos. Mag. A*, 59(2):321–336, 1989.

[272] B. Lee, J. Shim, and M. I. Baskes. Semiempirical atomic potentials for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, Al, and Pb based on first and second nearest-neighbor modified embedded atom method. *Phys. Rev. B*, 68:144112, 2003.

[273] L. Zhang, J. Han, H. Wang, R. Car, and W. E. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.*, 120:143001, 2018.

[274] V. L. Deringer and G. Csányi. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B*, 95:094203, 2017.

[275] Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang, and Weinan E. Dp-gen: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput. Phys. Commun.*, 253:107206, 2020.

[276] C. Kittel and P. McEuen. *Introduction to solid state physics*. John Wiley & Sons, 2018.

[277] E. A. Brandes and G. Brook. *Smithells metals reference book*. Elsevier, 2013.

[278] W.R. Tyson and W.A. Miller. Surface free energies of solid metals: Estimation from liquid surface tension measurements. *Surf. Sci.*, 62(1):267–276, 1977.

[279] J. Merikoski, I. Vattulainen, J. Heinonen, and T. Ala-Nissila. Effect of kinks and concerted diffusion mechanisms on mass transport and growth on stepped metal surfaces. *Surf. Sci.*, 387(1-3):167–182, 1997.

[280] J. Merikoski and T. Ala-Nissila. Diffusion processes and growth on stepped metal surfaces. *Phys. Rev. B*, 52(12):R8715, 1995.

[281] L. S. Perkins and A. E. DePristo. Self-diffusion mechanisms for adatoms on fcc (100) surfaces. *Surf. Sci.*, 294(1-2):67–77, 1993.

[282] G. Boisvert and L. J. Lewis. Self-diffusion of adatoms, dimers, and vacancies on Cu (100). *Phys. Rev. B*, 56(12):7643, 1997.

[283] M. Scheffler. Physical origin of exchange diffusion on fcc (100) metal surfaces. *Phys. Rev.-series B-*, 56:R15–569, 1997.

[284] H Dürr, JF Wendelken, and J-K Zuo. Island morphology and adatom energy barriers during homoepitaxy on Cu(001). *Surf. Sci.*, 328(1-2):L527–L532, 1995.

[285] H. Ernst, F. Fabre, and J. Lapujoulade. Nucleation and diffusion of Cu adatoms on Cu (100): A helium-atom-beam scattering study. *Phys. Rev. B*, 46(3):1929, 1992.

[286] L. Hansen, P. Stoltze, K. W. Jacobsen, and J. K. Nørskov. Self-diffusion on copper surfaces. *Phys. Rev. B*, 44:6523–6526, 1991.

[287] F. Montalenti and R. Ferrando. Jumps and concerted moves in Cu, Ag, and Au (110) adatom self-diffusion. *Phys. Rev. B*, 59(8):5881, 1999.

[288] K. H. Chae, H. C. Lu, and T. Gustafsson. Medium-energy ion-scattering study of the temperature dependence of the structure of Cu(111). *Phys. Rev. B*, 54:14082–14086, 1996.

[289] A. N. Al-Rawi, A. Kara, and T. S. Rahman. Comparative study of anharmonicity: Ni(111), Cu(111), and Ag(111). *Phys. Rev. B*, 66:165439, 2002.

[290] L. Yang and T. S. Rahman. Enhanced anharmonicity on Cu(110). *Phys. Rev. Lett.*, 67:2327–2330, 1991.

[291] P. Ditlevsen, P. Stoltze, J. No, et al. Anharmonicity and disorder on the Cu (110) surface. *Phys. Rev. B*, 44(23):13002, 1991.

[292] A. Baddorf and E. Plummer. Enhanced surface anharmonicity observed in vibrations on Cu (110). *Phys. Rev. Lett.*, 66(21):2770, 1991.

[293] A. Trayanov, A. C. Levi, and E. Tosatti. Missing-row roughening of (110) surfaces. *Europhysics Letters*, 8(7):657, 1989.

[294] G. Witte, J. Braun, D. Nowack, L. Bartels, B. Neu, and G. Meyer. Oxygen-induced reconstructions on Cu(211). *Phys. Rev. B*, 58:13224–13232, 1998.

[295] R.E. Kirby, C.S. McKee, and L.V. Renny. Faceting of Cu(210) and Ni(210) by activated nitrogen. *Surf. Sci.*, 97(2):457–477, 1980.

[296] R. J. Campello, D. Moulavi, and J. Sander. Density-based clustering based on hierarchical density estimates. *Lect. Notes Comput. Sci.*, 7819 LNAI(PART 2):160–172, 2013.

[297] R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander. Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Trans. Knowl. Discov. Data*, 10(1):1–51, 2015.

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

[299] Dmitrii E. Makarov. *Single Molecule Science: Physical Principles and Models*. CRC Press, 2015.

[300] Piyush B. Gupta, C. Fillmore, G Jiang, S D. Shapira, K. Tao, C. Kuperwasser, and Eric S. Lander. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146(4):633–644, 2011.

[301] Zhang J.n, M. Fei, and X. Ke-Wei. Calculation of the surface energy of fcc metals with modified embedded-atom method. *Chin. Phys.*, 13(7):1082–1090, 2004.

[302] J. Wang and S. Q. Wang. Surface energy and work function of fcc and bcc crystals: Density functional study. *Surf. Sci.*, 630:216–224, 2014.

[303] S. Titmuss, A. Wander, and D. A. King. Reconstruction of clean and adsorbate-covered metal surfaces. *Chem. Rev.*, 96(4):1291–1306, 1996.

[304] R. Koch, M. Borbonus, O. Haase, and K. H. Rieder. Reconstruction behaviour of fcc(110) transition metal surfaces and their vicinals. *Appl. Phys. A*, 55(5):417–429, 1992.

[305] R. Koch, M. Sturmat, and J.J. Schulz. High-temperature stm investigation of Au(110), Pt(110) and Ag(110). *Surf. Sci.*, 454:543–551, 2000.

[306] P. Grosse, D. Gao, F. Scholten, I. Sinev, H. Mistry, and B. Roldan Cuenya. Dynamic changes in the structure, chemical state and catalytic selectivity of Cu nanocubes during CO_2 electroreduction: size and support effects. *Angew. Chem. Int. Ed.*, 57(21):6192–6197, 2018.

[307] Y. Li, D. Kim, S. Louisia, C. Xie, Q. Kong, S. Yu, T. Lin, S. Aloni, S. C. Fakra, and P. Yang. Electrochemically scrambled nanocrystals are catalytically active for CO_2-to-multicarbons. *Proc. Natl. Acad. Sci. U.S.A.*, 117(17):9194–9201, 2020.

[308] G. Wang, L. Jiang, Y. Morikawa, J. Nakamura, Z. Cai, Y. Pan, and X. Zhao. Cluster and periodic DFT calculations of adsorption and activation of CO_2 on the Cu (hkl) surfaces. *Surf. Sci.*, 570(3):205–217, 2004.

[309] L. Delgado-Callico, K. Rossi, R. Pinto-Miles, P. Salzbrenner, and F. Baletto. A universal signature in the melting of metallic nanoparticles. *Nanoscale*, 13(2):1172–1180, 2021.

[310] HD Wanzenböck, N Weissenbacher, and R Kellner. Surface enhanced infrared absorption (seira) observed on different metal surface structures. In *AIP Conference Proceedings*, volume 430, pages 598–601. American Institute of Physics, 1998.

[311] Y. P. Zhu, C. Guo, Y. Zheng, and S. Qiao. Surface and interface engineering of noble-metal-free electrocatalysts for efficient energy conversion processes. *Acc. Chem. Res.*, 50(4):915–923, 2017.

[312] C. Li, M. Iqbal, J. Lin, X. Luo, B. Jiang, V. Malgras, K. C. Wu, J. Kim, and Y. Yamauchi. Electrochemical deposition: an advanced approach for templated synthesis of nanoporous metal architectures. *Acc. Chem. Res.*, 51(8):1764–1773, 2018.

[313] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, et al. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. Condens. Matter Phys.*, 21(39):395502, 2009.

[314] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865, 1996.

[315] C. M. Andolina, M. Bon, D. Passerone, and W. A. Saidi. Robust, multi-length-scale, machine learning potential for Ag–Au bimetallic alloys from clusters to bulk materials. *J. Phys. Chem. C*, 125(31):17438–17447, 2021.

[316] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. Joannopoulos. Optimized pseudopotentials. *Phys. Rev. B*, 41(2):1227, 1990.

[317] N. Marzari, D. Vanderbilt, A. De Vita, and M. C. Payne. Thermal Contraction and Disordering of the Al(110) Surface. *Phys. Rev. Lett.*, 82:3296–3299, 1999.

[318] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[319] D. P. Kingma and J. Ba. Adam:A method for stochastic optimization. *arXiv:1412.6980*, 2014.

[320] M. Stratakis and H. Garcia. Catalysis by Supported Gold Nanoparticles: Beyond Aerobic Oxidative Processes. *Chem. Rev.*, 112(8):4469–4506, 2012.

[321] R. Meyer, C. Lemire, S. K. Shaikhutdinov, and H. . Freund. Surface chemistry of catalysis by gold. *Gold Bull.*, 37(1-2):72–124, 2004.

[322] P. Liu, E. A. Irmak, A. D. Backer, A. D. wael, I. Lobato, A. Béché, S. V. Aert, and S. Bals. Three-dimensional atomic structure of supported Au nanoparticles at high temperature. *Nanoscale*, 13(3):1770–1776, 2020.

[323] R. Sardar, A. M. Funston, P. Mulvaney, and R. W. Murray. Gold Nanoparticles: Past, Present, and Future. *Langmuir*, 25(24):13840–13851, 2009.

[324] M. Haruta, N. Yamada, T. Kobayashi, and S. Iijima. Gold catalysts prepared by coprecipitation for low-temperature oxidation of hydrogen and of carbon monoxide. *J. Catal.*, 115(2):301–309, 1989.

[325] K. Saha, S. S. Agasti, C. Kim, X. Li, and V. M. Rotello. Gold Nanoparticles in Chemical and Biological Sensing. *Chem. Rev.*, 112(5):2739–2779, 2012.

[326] Jorge Pérez-Juste, Isabel Pastoriza-Santos, Luis M. Liz-Marzán, and Paul Mulvaney. Gold nanorods: Synthesis, characterization and applications. *Coord. Chem. Rev.*, 249(17):1870–1901, 2005.

[327] Y. Zhou, C. Jin, Y. Li, and W. Shen. Dynamic behavior of metal nanoparticles for catalysis. *Nano Today*, 20:101–120, 2018.

[328] E. B. Sterk, A. Nieuwelink, M. Monai, J. N. Louwen, E. T. C. Vogt, I. A. W. Filot, and B. M. Weckhuysen. Structure Sensitivity of CO_2 Conversion over Nickel Metal Nanoparticles Explained by Micro-Kinetics Simulations. *JACS Au*, 2022.

[329] W. Xu, J. S. Kong, and P. Chen. Probing the catalytic activity and heterogeneity of Au-nanoparticles at the single-molecule level. *Phys. Chem. Chem. Phys.*, 11(15):2767–2778, 2009.

[330] A. Pedrazo-Tardajos, E. Arslan Irmak, V. Kumar, A. Sánchez-Iglesias, Q. Chen, M. Wirix, B. Freitag, W. Albrecht, S. Van Aert, L. M. Liz-Marzán, and S. Bals. Thermal activation of gold atom diffusion in Au@Pt nanorods. *ACS Nano*, 16(6):9608–9619, 2022.

[331] E. Arslan Irmak, P. Liu, S. Bals, and S. Van Aert. 3D atomic structure of supported metallic nanoparticles estimated from 2D adf stem images: A combination of atom-counting and a local minima search algorithm. *Small Methods*, 5(12):2101150, 2021.

[332] M. Settem, R. Ferrando, and A. Giacomello. Tempering of Au nanoclusters: capturing the temperature-dependent competition among structural motifs. *Nanoscale*, 14(3):939–952, 2022.

[333] Benjamin C. Curley, Roy L. Johnston, Neil P. Young, Z. Y. Li, Marcel Di Vece, Richard E. Palmer, and Andrew L. Bleloch. Combining Theory and Experiment to Characterize the Atomic Structures of Surface-Deposited Au_309 Clusters. *J. Phys. Chem. C*, 111:17846–17851, 2007.

[334] J. Palomares-Baez, E. Panizon, and R. Ferrando. Nanoscale Effects on Phase Separation. *Nano Lett.*, 17(9):5394–5401, 2017.

[335] E. Aprà, F. Baletto, R. Ferrando, and A. Fortunelli. Amorphization Mechanism of Icosahedral Metal Nanoclusters. *Phys. Rev. Lett.*, 93:065502, 2004.

[336] P. Mahler Larsen, S. Schmidt, and J.Schiotz. Robust structural identification via polyhedral template matching. *Model. Simul. Mat. Sci. Eng.*, 24, 2016.

[337] J. D. Honeycutt and H. C. Andersen. Molecular dynamics study of melting and freezing of small lennard-jones clusters. *J. Phys. Chem.*, 91(19):4950–4963, 1987.

[338] L. McInnes, J. Healy, and S. Astels. hdbscan: Hierarchical density based clustering. *J. of Open Source Softw.*, 2(11), 2017.

[339] F. Cyrot-Lackmann and F. Ducastelle. Binding energies of transition-metal atoms adsorbed on a transition metal. *Phys. Rev. B*, 4:2406–2412, 1971.

[340] D. Foster, T. Pavloudis, J. Kioseoglou, and R. Palmer. Atomic-resolution imaging of surface and core melting in individual size-selected Au nanoclusters on carbon. *Nat. Commun.*, 10:2583, 2019.

[341] M. Yang, U. Raucci, and M. Parrinello. Ammonia decomposition on lithium imide surfaces: A new paradigm in heterogeneous catalysis, 2022. ChemRxiv: 10.26434/chemrxiv-2022-qr7wt.

[342] M. Cioni, M. Delle Piane, D. Polino, D. Rapetti, M. Crippa, E. A. Irmak, S. Van Aert, S. Bals, and G. M. Pavan. Sampling real-time atomic dynamics in metal nanoparticles by combining experiments, simulations, and machine learning. *Adv. Sci.*, (2307261), 2024.

[343] D. Rapetti and R. Ferrando. bh++, 2022.

[344] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, Pieter J. in 't Veld, Axel Kohlmeyer, Stan G. Moore, Trung Dac Nguyen, Ray Shan, Mark J. Stevens, Julien Tranchida, Christian Trott, and Steven J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.*, 271:108171, 2022.

[345] L. Himanen, Marc OJ Jäger, Eiaki V Morooka, Filippo Federici Canova, Yashasvi S Ranawat, David Z Gao, Patrick Rinke, and Adam S Foster. Dscribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.*, 247:106949, 2020.

[346] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods*, 17:261–272, 2020.

[347] R. Ferrando. Determining the equilibrium structures of nanoalloys by computational methods. *J. Nanopart. Resh.*, 20(7):179, 2018.

[348] R. Ferrando, J. Jellinek, and R. L. Johnston. Nanoalloys: from theory to applications of alloy clusters and nanoparticles. *Chem. Rev*, 108(3):845–910, 2008.

[349] M. Dearg, C. Roncaglia, D. Nelli, R. Ferrando, T. J. Slater, R. E. Palmer, et al. Frame-by-frame observations of structure fluctuations in single mass-selected Au clusters using aberration-corrected electron microscopy. *Nanoscale Horizons*, 9(1):143–147, 2024.

[350] D. Bochicchio and R. Ferrando. Morphological instability of core-shell metallic nanoparticles. *Phys. Rev. B*, 87(16):165435, 2013.

[351] K. Rossi, T. Ellaby, LO Paz-Borbón, I Atanasov, L Pavan, and F Baletto. Melting of large Pt-MgO(100) icosahedra. *J. Phys. Condens. Matter*, 29(14):145402, 2017.

[352] A. Rimola, D. Costa, M. Sodupe, J. Lambert, and P. Ugliengo. Silica surface features and their role in the adsorption of biomolecules: computational modeling and experiments. *Chem. Rev.*, 113(6):4216–4313, 2013.

[353] C. Li, C. Koenigsmann, W. Ding, B. Rudshteyn, K. R. Yang, K. P. Regan, S. J. Konezny, V. S. Batista, G. W. Brudvig, C. A. Schmuttenmaer, et al. Facet-dependent photoelectrochemical performance of tio2 nanostructures: an experimental and computational study. *J. Am. Chem. Soc.*, 137(4):1520–1529, 2015.

[354] C. R. Zhu, D. Gao, J. Ding, D. Chao, and J. Wang. Tmd-based highly efficient electrocatalysts developed by combined computational and experimental approaches. *Chem. Soc. Rev.*, 47(12):4332–4356, 2018.

[355] T. Sperger, I. A. Sanhueza, and F. Schoenebeck. Computation and experiment: a powerful combination to understand and predict reactivities. *Acc. Chem. Res.*, 49(6):1311–1319, 2016.

[356] T. Ishida, H. Koga, M. Okumura, and M. Haruta. Advances in gold catalysis and understanding the catalytic mechanism. *TCR*, 16(5):2278–2293, 2016.

[357] P. Schlexer, A. B. Andersen, B. Sebok, I. Chorkendorff, J. Schiøtz, and T. W. Hansen. Size-Dependence of the Melting Temperature of Individual Au Nanoparticles. *Part. Part. Syst. Charact.*, 36(3):1800480, 2019.

[358] S. I. Sanchez, L. D. Menard, A. Bram, J. H. Kang, M. W. Small, R. G. Nuzzo, and A. I. Frenkel. The emergence of nonbulk properties in supported metal clusters: negative thermal expansion and atomic disorder in Pt nanoclusters supported on γ-al2o3. *J. Am. Chem. Soc.*, 131(20):7040–7054, 2009.

[359] M. A. Centeno, Tomás Ramírez Reina, Svetlana Ivanova, Oscar Hernando Laguna, and José Antonio Odriozola. Au/ceo2 catalysts: structure and co oxidation activity. *Catalysts*, 6(10):158, 2016.

[360] B. Zhu, J. Meng, W. Yuan, X. Zhang, H. Yang, Y. Wang, and Y. Gao. Reshaping of metal nanoparticles under reaction conditions. *Angew. Chem. Int. Ed.*, 59(6):2171–2180, 2020.

[361] K. Rossi, G. G. Asara, and F. Baletto. Correlating oxygen reduction reaction activity and structural rearrangements in mgo-supported platinum nanoparticles. *ChemPhysChem*, 20(22):3037–3044, 2019.

[362] S. Van Aert, K. J. Batenburg, M. D. Rossell, R. Erni, and G. Van Tendeloo. Three-dimensional atomic imaging of crystalline nanoparticles. *Nature*, 470:374–377, 2011.

[363] C. Caruso, A. Cardellini, M. Crippa, D. Rapetti, and G. M. Pavan. Timesoap: Tracking high-dimensional fluctuations in complex molecular systems via time variations of soap spectra. *J. Chem. Phys.*, 158(21), 2023.

[364] B. Hvolbæk, T. VW. Janssens, B. S. Clausen, H. Falsig, C. H. Christensen, and J. Nørskov. Catalytic activity of Au nanoparticles. *Nano Today*, 2(4):14–18, 2007.

[365] Q. Fu, H. Saltsburg, and M. Flytzani-Stephanopoulos. Active nonmetallic Au and Pt species on ceria-based water-gas shift catalysts. *Science*, 301(5635):935–938, 2003.

[366] S. A. Mavlyankariev, S. J. Ahlers, V. A. Kondratenko, D. Linke, and E. V. Kondratenko. Effect of support and promoter on activity and selectivity of gold nanoparticles in propanol synthesis from co2, c2h4, and H2. *ACS Catal.*, 6(5):3317–3325, 2016.

[367] S. L. Hemmingson, T. E. James, G. M. Feeley, A. M. Tilson, and C. T. Campbell. Adsorption and adhesion of Au on reduced ceo2(111) surfaces at 300 and 100 K. *J. Phys. Chem. C*, 120(22):12113–12124, 2016.

[368] M. Haruta. When gold is not noble: catalysis by nanoparticles. *TCR*, 3(2):75–87, 2003.

[369] R. M. Jones, K. Rossi, C. Zeni, M. Vanzan, I. Vasiljevic, A. Santana-Bonilla, and F. Baletto. Structural characterisation of nanoalloys for (photo) catalytic applications with the sapphire library. *Faraday Discuss.*, 242:326–352, 2023.

[370] A. Likas, N. Vlassis, and J. J. Verbeek. The global k-means clustering algorithm. *Pattern Recognit.*, 36(2):451–461, 2003.

[371] H. Ha, S. Yoon, K. An, and H. Y. Kim. Catalytic co oxidation over Au nanoparticles supported on ceo2 nanocrystals: effect of the Au–ceo2 interface. *ACS Catal.*, 8(12):11491–11501, 2018.

[372] T. Uchiyama, H. Yoshida, Y. Kuwauchi, S. Ichikawa, S. Shimada, M. Haruta, and S. Takeda. Systematic morphology changes of gold nanoparticles supported on ceo2 during co oxidation. *Angew. Chem. Int. Ed.*, 50(43):10157–10160, 2011.

[373] R. Si and M. Flytzani-Stephanopoulos. Shape and crystal-plane effects of nanoscale ceria on the activity of Au-ceo2 catalysts for the water–gas shift reaction. *Angewandte Chemie*, 120(15):2926–2929, 2008.

[374] S. Van Aert, A. De Backer, G. Martinez, B. Goris, S. Bals, G. Van Tendeloo, and A. Rosenauer. Procedure to count atoms with trustworthy single-atom sensitivity. *Phys. Rev. B*, 87(6):064107, 2013.

[375] I. Lobato, S. Van Aert, and J. Verbeeck. Progress and new advances in simulating electron microscopy datasets using multem. *Ultramicroscopy*, 168:17–27, 2016.

[376] F. F. Krause, M. Schowalter, T. Grieb, Knut Müller Caspary, Thorsten Mehrtens, and Andreas Rosenauer. Effects of instrument imperfections on quantitative scanning transmission electron microscopy. *Ultramicroscopy*, 161:146–160, 2016.

[377] H. Gao and L. Peng. Parameterization of the temperature dependence of the debye–waller factors. *Acta Crystallogr. A*, 55(5):926–932, 1999.

[378] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[379] C. R. Henry. Morphology of supported nanoparticles. *Prog. Surf. Sc.*, 80(3-4):92–116, 2005.

[380] M. Yu, A. B. Yankovich, A. Kaczmarowski, D. Morgan, and P. M. Voyles. Integrated computational and experimental structure refinement for nanoparticles. *ACS nano*, 10(4):4031–4038, 2016.

[381] F. Pietrucci and Roman Martoňák. Systematic comparison of crystalline and amorphous phases: Charting the landscape of water structures and transformations. *J. Chem. Phys.*, 142(10):104704, 2015.

[382] Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.*, 134(7):074106, 2011.

[383] R. Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B*, 99:014104, 2019.

[384] F. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento. Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem.*, 115(16):1094–1101, 2015.

[385] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti. Physics-Inspired Structural Representations for Molecules and Materials. *Chem. Rev.*, 121(16):9759–9815, 2021.

[386] J. Andrews, O. Gkountouna, and E. Blaisten-Barojas. Forecasting molecular dynamics energetics of polymers in solution from supervised machine learning. *Chem. Sci.*, 13(23):7021, 2022.

[387] P. Gasparotto and M. Ceriotti. Recognizing molecular patterns by machine learning: An agnostic structural definition of the hydrogen bond. *J. Chem. Phys.*, 141(17):174110, 2014.

[388] M. B. Davies, M. Fitzner, and A. Michaelides. Accurate prediction of ice nucleation from room temperature water. *Proc. Natl. Acad. Sci. U.S.A.*, 119(31):e2205347119, 2022.

[389] F. Noé, S. Olsson, J. Köhler, and H. Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

[390] A. Cardellini, M. Crippa, C. Lionello, S. P. Afrose, D. Das, and G. M. Pavan. Unsupervised data-driven reconstruction of molecular motifs in simple to complex dynamic micelles. *J. Phys. Chem. B*, 127(11):2595–2608, 2023.

[391] C. Lionello, C. Perego, A. Gardin, R. Klajn, and G. M. Pavan. Supramolecular semiconductivity through emerging ionic gates in ion–nanoparticle superlattices. *ACS Nano*, 17(1):275–287, 2023.

[392] B. Schölkopf, A. Smola, and K. Müller. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Comput.*, 10(5):1299–1319, 1998.

[393] L. van der Maaten and G. Hinton. Visualizing Data using t-SNE. *J. Mach. Learn. Res.*, 9(86):2579–2605, 2008.

[394] S. Lloyd. Least squares quantization in pcm. *IEEE Trans. Inf. Theory*, 28(2):129–137, 1982.

[395] Douglas Reynolds. *Gaussian Mixture Models*, pages 659–663. Springer US, 2009.

[396] E. Schubert, J Sander, M. Ester, H. Peter Kriegel, and X. Xu. Dbscan revisited, revisited: Why and how you should (still) use dbscan. *ACM Trans. Database Syst.*, 42(3):(3) 1–21, 2017.

[397] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36(8):1627–1639, 1964.

[398] C. Caruso, M. Crippa, A Cardellini, M. Cioni, M. Perrone, M. Delle Piane, and G. M. Pavan. Classification and spatiotemporal correlation of dominant fluctuations in complex dynamical systems. *arXiv:2409.18844v1*, 2024.

[399] K. Jorner, A. Tomberg, C. Bauer, Christian Sköld, and Per-Ola Norrby. Organic reactivity from mechanism to machine learning. *Nat. Rev. Chem.*, 5:240–255, 2021.

[400] S. Perego and L. Bonati. Data-efficient modeling of catalytic reactions via enhanced sampling and on-the-fly learning of machine learning potentials. *ChemRxiv*, 2024.

[401] M. Delle Piane, L. Pesce, M. Cioni, and G. M. Pavan. Reconstructing reactivity in dynamic host–guest systems at atomistic resolution: amide hydrolysis under confinement in the cavity of a coordination cage. *Chem. Sci.*, 13(37):11232–11245, 2022.

[402] H. Takezawa, K. Shitozawa, and M. Fujita. Enhanced reactivity of twisted amides inside a molecular cage. *Nat. Chem.*, 12(6):574–578, 2020.

[403] A. B. Grommet, M. Feller, and R. Klajn. Chemical reactivity under nanoconfinement. *Nat. Nanotechnol.*, 15(4):256–271, 2020.

[404] D. Fiedler, D. H. Leung, R. G. Bergman, and K. N. Raymond. Selective Molecular Recognition, C–H Bond Activation, and Catalysis in Nanoscale Reaction Vessels. *Acc. Chem. Res.*, 38(4):349–358, 2005.

[405] M. Maestri and E. Iglesia. First-principles theoretical assessment of catalysis by confinement: NO–O2 reactions within voids of molecular dimensions in siliceous crystalline frameworks. *Phys. Chem. Chem. Phys.*, 20(23):15725–15735, 2018.

[406] M. Yoshizawa, M. Tamura, and M. Fujita. Diels-Alder in Aqueous Molecular Hosts: Unusual Regioselectivity and Efficient Catalysis. *Science*, 312(5771):251 – 254, 2006.

[407] Y. Zou, Ilma Jahović, and Jonathan R Nitschke. Hydrolysis of Twisted Amides inside a Self-Assembled Coordination Cage. *Chem*, 6(6):1217–1218, 2020.

[408] V. Mouarrawis, R. Plessius, J. I. van der Vlugt, and J. N. H. Reek. Confinement Effects in Catalysis Using Well-Defined Materials and Cages. *Front. Chem.*, 6:623, 2018.

[409] L. Pesce, C. Perego, A. B. Grommet, R. Klajn, and G. M. Pavan. Molecular factors controlling the isomerization of azobenzenes in the cavity of a flexible coordination cage. *J. Am. Chem. Soc.*, 142(21):9792–9802, 2020.

[410] D. G. A. Johansson, G. Wallin, A. Sandberg, B. Macao, J. Åqvist, and T. Härd. Protein autoproteolysis: Conformational strain linked to the rate of peptide cleavage by the ph dependence of the N $\rightarrow$ O acyl shift reaction. *J. Am. Chem. Soc.*, 131(27):9475–9477, 2009.

[411] M. Han, R. Michel, B. He, Y. Chen, D. Stalke, M. John, and G. H. Clever. Light-triggered guest uptake and release by a photochromic coordination cage. *Angew. Chem. Int. Ed.*, 52(4):1319–1323, 2013.

[412] F. J. Rizzuto, L. K. von Krbek, and J. R. Nitschke. Strategies for binding multiple guests in metal–organic cages. *Nat. Rev. Chem.*, 3:204–222, 2019.

[413] A. L. L. East. On the hydrolysis mechanisms of amides and peptides. *Int. J. Chem. Kin.*, 50(10):705–709, 2018.

[414] K. L. Fleming, P. Tiwary, and J. Pfaendtner. New Approach for Investigating Reaction Dynamics and Rates with Ab Initio Calculations. *J. Phys. Chem. A*, 120(2):299–305, 2016.

[415] M. Fujita, D. Oguro, M. Miyazawa, H. Oka, K. Yamaguchi, and K. Ogura. Self-assembly of ten molecules into nanometre-sized organic host frameworks. *Nature*, 378:469–471, 1995.

[416] F. Thunecke, A. Kálmán, F. Kálmán, S. Ma, A. S. Rathore, and C. Horváth. Kinetic study on the cis-trans isomerization of peptidyl-proline dipeptides. *J. Chromatogr. A*, 744(1):259–272, 1996.

[417] J. Blumberger, B. Ensing, and M. L. Klein. Formamide Hydrolysis in Alkaline Aqueous Solution: Insight from Ab Initio Metadynamics Calculations. *Angew. Chem. Int. Ed.*, 45(18):2893–2897, 2006.

[418] P. Craveur, A. P. Joseph, P. Poulain, A. G. d. Brevern, and J. Rebehmed. Cis–trans isomerization of omega dihedrals in proteins. *Amino Acids*, 45(2):279–289, 2013.

[419] U. Doshi and D. Hamelberg. Reoptimization of the AMBER Force Field Parameters for Peptide Bond (Omega) Torsions Using Accelerated Molecular Dynamics. *J. Phys. Chem. B*, 113(52):16590–16595, 2009.

[420] V. V. Welborn, W. Li, and T. Head-Gordon. Interplay of water and a supramolecular capsule for catalysis of reductive elimination reaction from gold. *Nat. Commun.*, 11:415, 2020.

[421] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner. Construction of tight-binding-like potentials on the basis of density–functional theory: Application to carbon. *Phys. Rev. B*, 51:12947–12957, 1995.

[422] G. Seifert, D. Porezag, and T. Frauenheim. Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme. *Int. J. Quantum Chem.*, 58(2):185–192, 1996.

[423] Ying Wu, Adelina Ilie, and Simon Crampin. Self-consistent charge and dipole density functional tight binding method and application to carbon-based systems. *Comput. Mater. Sci.*, 134:206–213, 2017.

[424] M. Gruden, L. Andjeklović, A. K. Jissy, S. Stepanović, M. Zlatar, Q. Cui, and M. Elstner. Benchmarking density functional tight binding models for barrier heights and reaction energetics of organic molecules. *J. Comput. Chem.*, 38(25):2171–2185, 2017.

[425] C. W. Bauschlicher, T. Qi, E. J. Reed, A. Lenfant, J. W. Lawson, and T. G. Desai. Comparison of reaxff, DFTB, and DFT for phenolic pyrolysis. 2. elementary reaction paths. *J. Phys. Chem. A*, 117(44):11126–11135, 2013.

[426] A. Priyadarsini and B. S. Mallik. Comparative first principles-based molecular dynamics study of catalytic mechanism and reaction energetics of water oxidation reaction on 2d-surface. *J. Comp. Chem.*, 42(16):1138–1149, 2021.

[427] C. S. Pereira, R. L. Silveira, and M. S. Skaf. Qm/mm simulations of enzymatic hydrolysis of cellulose: Probing the viability of an endocyclic mechanism for an inverting cellulase. *J. Chem. Info. Model.*, 61(4):1902–1912, 2021.

[428] V. S. Bharadwaj, B. C. Knott, Jerry Ståhlberg, Gregg T Beckham, and Michael F Crowley. The hydrolysis mechanism of a gh45 cellulase and its potential relation to lytic transglycosylase and expansin function. *J. Bio. Chem.*, 295(14):4477–4487, 2020.

[429] D. Bochicchio, M. Salvalaglio, and G. M. Pavan. Into the dynamics of a supramolecular polymer at submolecular resolution. *Nat. Commun.*, 8:147, 2017.

[430] M. Garzoni, M. B. Baker, C. M. A. Leenders, I. K. Voets, L. Albertazzi, A. R. A. Palmans, E. W. Meijer, and G. M. Pavan. Effect of H-Bonding on Order Amplification in the Growth of a Supramolecular Polymer in Water. *J. Am. Chem. Soc.*, 138(42):13985–13995, 2016.

[431] T. Drakenberg and S. Forsén. The barrier to internal rotation in monosubstituted amides. *J. Chem. Soc. D*, (21):1404–1405, 1971.

[432] Kenneth S. Pitzer. Potential energies for rotation about single bonds. *Discuss. Faraday Soc.*, 10:66–73, 1951.

[433] M. Becchi, F. Fantolino, and G. M. Pavan. Layer-by-layer unsupervised clustering of statistically relevant fluctuations in noisy time-series data of complex dynamical systems. *Proc. Natl. Acad. Sci. U.S.A.*, 121(33):e2403771121, 2024.