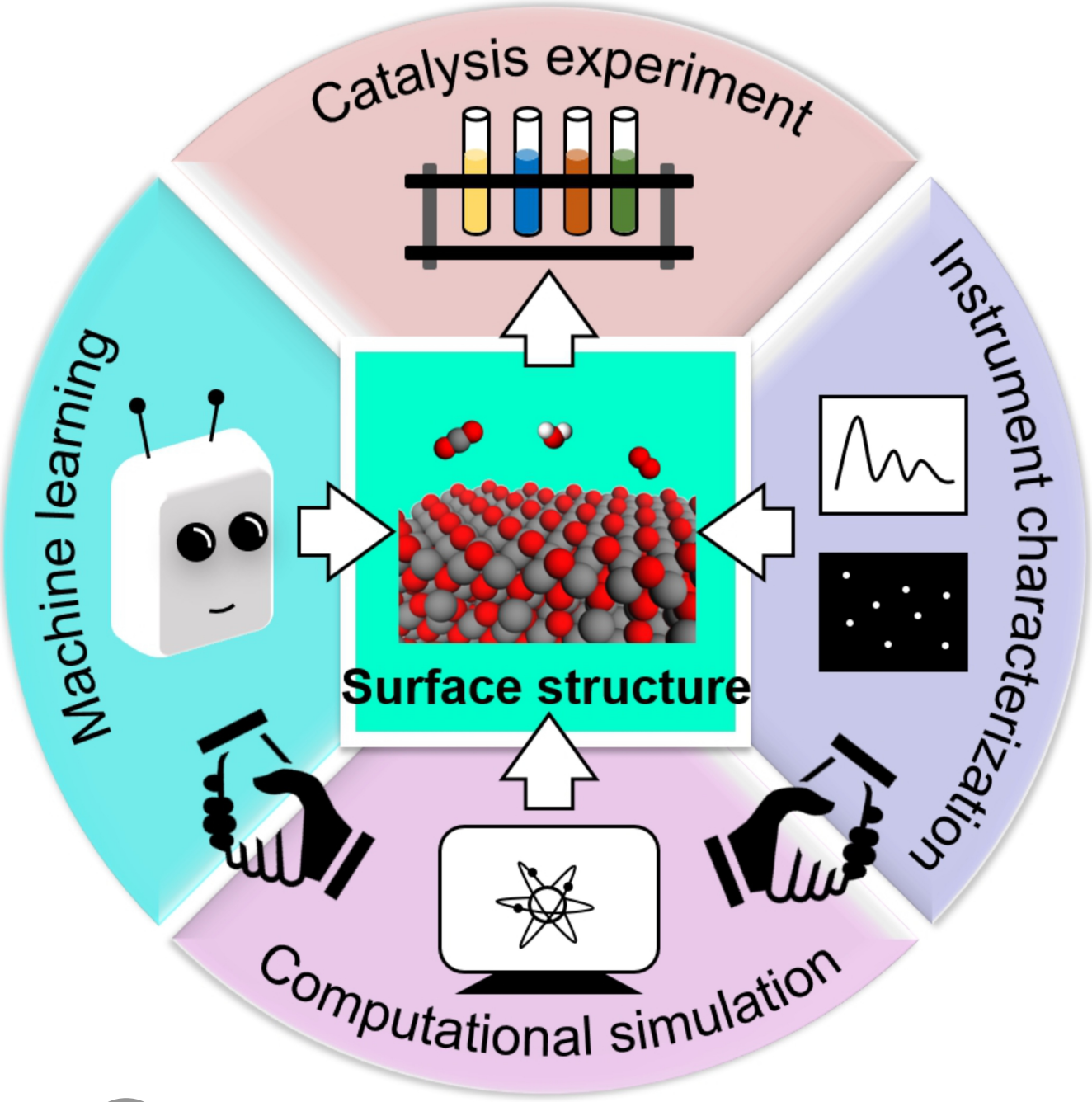
How to cite: *Angew. Chem. Int. Ed.* **2023,** *62,* e202216383
International Edition: doi.org/10.1002/anie.202216383
German Edition: doi.org/10.1002/ange.202216383

Catalyst Design

Data-Driven Machine Learning for Understanding Surface Structures of Heterogeneous Catalysts

*Haobo Li, Yan Jiao, Kenneth Davey, and Shi-Zhang Qiao**



Abstract: The design of heterogeneous catalysts is necessarily surface-focused, generally achieved via optimization of adsorption energy and microkinetic modelling. A prerequisite is to ensure the adsorption energy is physically meaningful is the stable existence of the conceived active-site structure on the surface. The development of improved understanding of the catalyst surface, however, is challenging practically because of the complex nature of dynamic surface formation and evolution under in-situ reactions. We propose therefore data-driven machine-learning (ML) approaches as a solution. In this Minireview we summarize recent progress in using machine-learning to search and predict (meta)stable structures, assist operando simulation under reaction conditions and micro-environments, and critically analyze experimental characterization data. We conclude that ML will become the new norm to lower costs associated with discovery and design of optimal heterogeneous catalysts.

1. Introduction

Heterogeneous catalysis is important in the chemical industry, especially in the context of energy conversion, carbon neutrality, and environment protection, and underscores sustainable development. A consensus amongst researchers is that catalytic reactions occur on the surface of heterogeneous catalysts. The surface structure of the catalyst at the gas/solid, or liquid/solid, interface during the catalytic reaction has therefore been subject to extensive investigation to understand reaction mechanisms and design catalyst materials.

Surface science began to develop rapidly in the 1970s.[1] Technical approaches to observe catalyst surface structure resulted rapidly in the development of catalysis. With the development of experimental technology, research in electrochemical surface science began in the 1990s.[2] Theoretical developments lagged behind experimental developments, until the advent of quantum chemical computations based on density functional theory (DFT) and application of the generalized gradient approximation (GGA) functional in the mid-1980s.[3] DFT computations are powerfully advantageous in computing the energy of surface structures from first principles and, importantly, to corroborate experiments and characterizations. This combined interdisciplinary fusion of experimental techniques and theoretical computations has become a distinguishing feature of the field.

A growing understanding of catalyst surface structures, computational design and screening of catalysts based on the structure of catalytic active sites led to a number of developments beginning in the early 2000s, for example, the work of Nørskov et al.[4] A key question, however, was whether theoretical computations could be relied on to successfully predict experimental outcomes. Surface structure consists of several atomic layers on the surface, and there are significant geometrically possible configurations and elemental compositions that will vary with reaction condition(s) and micro-environment. Surface structure complexity therefore can lead to significant computational costs and practical difficulties in operando theoretical simulation.

Machine learning (ML) has been applied at the intersection of multiple disciplines because of its capability to handle complex systems and make testable predictions. Although ML methods, including neural networks (NN), were proposed as early as the 1950s to 1970s, significant knowledge barriers existed to application in other disciplines. However, over the past decade the development of practical tools including the Torch library, Scikit-Learn,[5] and the Tensorflow library has reduced professional thresholds for users and boosted the application of ML in fields of physics and chemistry (Figure 1).[6] As a result, machine learning interatomic potentials (MLIP)[7] and machine learning force field (MLFF)[8] emerged and have been used to accelerate DFT computation and gain increased precision in large systems. Interpretable ML methods[9] are being used to understand multiple physical quantities on target properties and have led to suitable descriptors to make predictions. Additionally, the inverse design approach[10] is being applied to derive theoretical structure(s) from experimental characterizations with improved accuracy and efficiency.

In this Minireview, we summarize state-of-the-art research progress in data-driven ML applications to study heterogeneous catalyst surface structure. For "ideal" surface structures, ML can be applied to predict and explain the thermodynamic and kinetic stability of complex structures. For "actual" surface structures, ML can be combined with DFT, molecular dynamics (MD) and Monte Carlo (MC) simulations to enable molecular scale modelling of large and complex systems. From an experimental view, ML can be used advantageously to analyze and interpret experimentally obtained surface structure information and to construct increasingly accurate surface models for theoretical computation. ML will significantly accelerate conventional theoretical and experimental methods to rationally design surface structures in heterogeneous catalytic reactions.

2. Ideal Catalyst Surface Structures

Molecular modelling is important because it can be relied on to provide atomic level insight for experimental measurements in surface science. Significantly, computational and experimental studies have achieved good correspondence in model systems with (relatively) simple structures, including

[*] Dr. H. Li, Prof. Y. Jiao, Prof. K. Davey, Prof. S.-Z. Qiao
School of Chemical Engineering and Advanced Materials
The University of Adelaide
Adelaide, SA 5005 (Australia)
E-mail: s.qiao@adelaide.edu.au

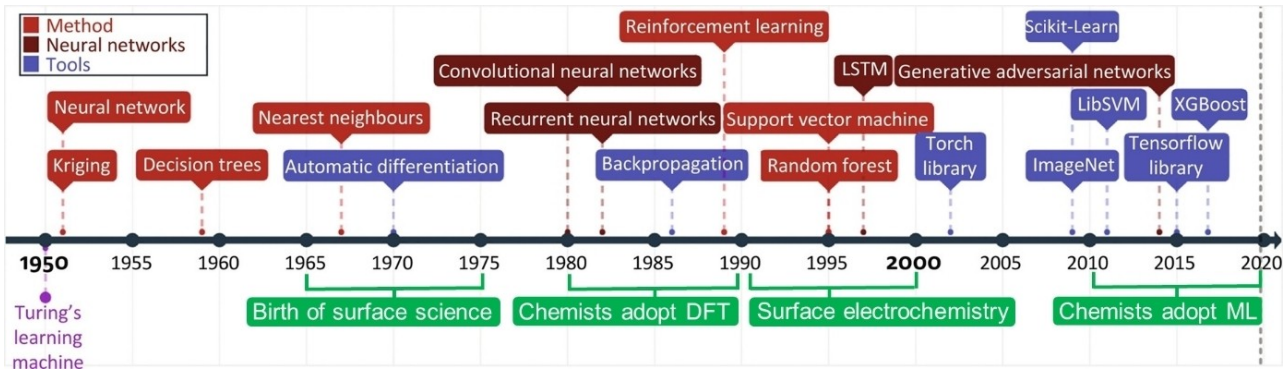


Figure 1. A crude timeline for developments in machine learning (ML), surface science, and computational chemistry. The ML part is reproduced with permission.[11] Copyright 2020, Elsevier. Developments in surface science and computational chemistry (marked in green) are summarized in this work.

metal single-crystal surfaces. However, the complexity of actual catalyst structures presents a potential practical difficulty for theoretical modelling.[12] With alloy catalysts, for example, as the number of metal elements increases from binary, ternary, to high-entropy alloys with five or more components, the number of possible surface structures increases exponentially (Figure 2a). Compared with the highly symmetrical crystalline bulk, the surface structure has significantly increased possibilities because of its symmetry breaking at the edges, and the interaction of surface atomic layers.[13] Given that stable structures with minimum energy, and metastable structures with energy at the local minima, can be obtained experimentally, it is necessary to compare many possibilities to ensure all key data. Additionally, ionic compounds with non-metallic element components and composite structures at the interfaces of different species, and amorphous phases add further complication (Figure 2a). In practice, however, modelling software including Atomic Simulation Environment (ASE)[14] can be used to build a significantly large number of models. However, the massive computational cost to calculate these models is hardly affordable. It is expected that ML can obviate these difficulties.

2.1. Thermodynamic Stability of Surface Structures

Thermodynamic surface free energy (γ) is a direct descriptor for comparing structures, i.e. the surface structure with lowest energy is selected. DFT is a quantum chemical

Haobo Li received her Ph.D. in 2017 from Dalian Institute of Chemical Physics (DICP). In 2018 she moved to Technical University of Munich (TUM) as an Alexander von Humboldt postdoctoral fellow. She is currently a Research Fellow at the School of Chemical Engineering and Advanced Materials, The University of Adelaide (Australia). Her research interest focus on artificial intelligence assisted design of catalysts.

Yan Jiao received her Ph.D. in Chemical Engineering from the University of Queensland (Australia). She is currently an Associate Professor and Australian Research Council Future Fellow at the School of Chemical Engineering and Advanced Materials, The University of Adelaide. Her research expertise is computational electrochemistry and developing catalyst materials for clean energy conversion reactions by molecular modelling.

A/Professor Kenneth Davey FIChemE CEng FAIFST CSci received his Ph.D. from Melbourne University (Australia). He then worked as a Postdoctoral Researcher and Senior Research Scientist with CSIRO and subsequently at The University of Adelaide. He is an interdisciplinary researcher in (bio)chemical engineering, with application to the development of new risk assessments for improved process efficiencies, reliability and safety, and energy conversions and storage.

Shi-Zhang Qiao received his Ph.D. degree in Chemical Engineering from Hong Kong University of Science and Technology in 2000. He is currently a Chair Professor and Australian Laureate Fellow at the School of Chemical Engineering and Advanced Materials, The University of Adelaide (Australia). His research expertise is nanostructured materials for electrocatalysis, energy storage, and conversion technologies.

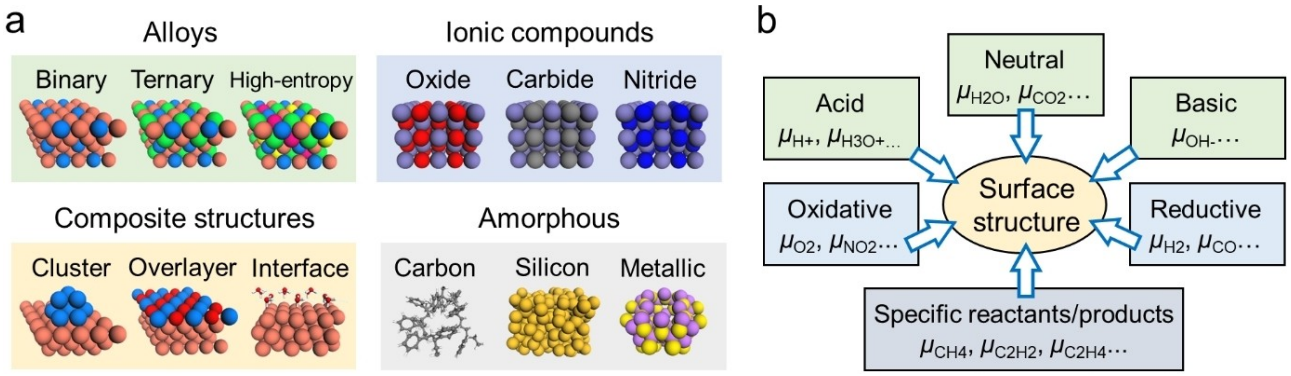


Figure 2. a) Catalyst surface structures with significant complexity. Numerous possibilities for surface morphology, alloying or non-metallic element composition, and crystalline state, present practical difficulties in computational modeling and simulation. b) Because different reactions have different references to chemical potential (μ), these affect the stability of the surface structure under reaction.

computation method with a good balance between speed and accuracy, and has therefore has been widely used. The Gibbs free energies of the surface structures (G_{surf}) are approximated by the respective total energies (E_{surf})[15] [Eq. (1)],

$$G_{surf} \approx E_{surf} \tag{1}$$

where surface free energy is computed from total energy, surface area (A), and chemical potentials (μ) for each component I [Eq. (2).

$$\gamma = \frac{1}{A}(E_{surf} - \sum_i \mu_i) \tag{2}$$

Thermodynamic selection of stable and metastable surface can then be simplified to directly compare the DFT-computed atomic potential energy value. It should be noted that for different reactions, the reference molecules of the chemical potential μ vary (Figure 2b), so different surface structure stability relationships can be obtained through computations using Equation (2). In this way, the effect of reactions on the ideal surface structure is determined.

A comparison between conventional and data-driven research methodology is shown in Figure 3. The conventional method is to perform theoretical modelling of empirical structures based on experimental findings. This method is limited, however, by experimental working efficiency, computational speed, and possible human error. With the introduction of emerging ML methods, the data-driven programming and sampling algorithms provide improved efficiency and accuracy. This is achieved by replacing human experience in evaluating the entire dataset, examining all local minimum data points, and determining stable and metastable structure(s). In this way, new knowledge about the physical meaning of structural stability is gained, adding understanding to the discipline of surface science.

Several ML methods have been applied to identify stable or metastable surface structures. To search for energy local minima on the potential energy surface from a large

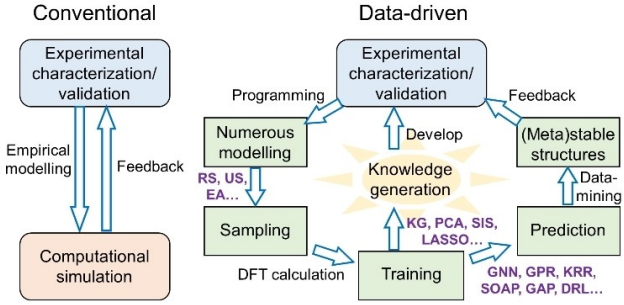


Figure 3. A comparison of workflow for conventional and data-driven methods to assess surface structure(s) of heterogeneous catalysts. Conventional methods are relatively empirical, whilst data-driven methods permit processing significant data with higher efficiency and potential for new knowledge.

modelling dataset, random sampling (RS) and uncertainty sampling (US)[16] from active learning[17] methods are used to select structures for computation. Bio-inspired sampling genetic algorithms (GA)[18] and evolutionary algorithms (EA)[19] have been used to find optimized structures along the potential energy surface. These sampling methods can be combined with ML regression to accelerate energetic data generation. Jennings et al.[20] used Gaussian process regression (GPR)-accelerated GA on PtAu alloy nanoparticles to give a 50-fold reduction in the number of required energy computations (Figure 4a). Jacobsen et al.[21] used kernel ridge regression (KRR)-guided EA to analyze the stability of local motifs in SnO_2(110) surface reconstruction and reported significant speedier searching. Sampling methods, including EA, have been combined with neural networks (NNs) to investigate multicomponent alloy nanoparticles.[22] The GOFEE[23] algorithm based on GPR has been developed to sample from the energy landscape and to generate structural candidates that correspond to the results of STM and AFM experiments for oxide overlayer structures on Pt_3Sn(111)[24] (Figure 4b).

The generation of new knowledge as feedback for experiments is another significant advantage with ML. This requires the use of interpretable ML models and the

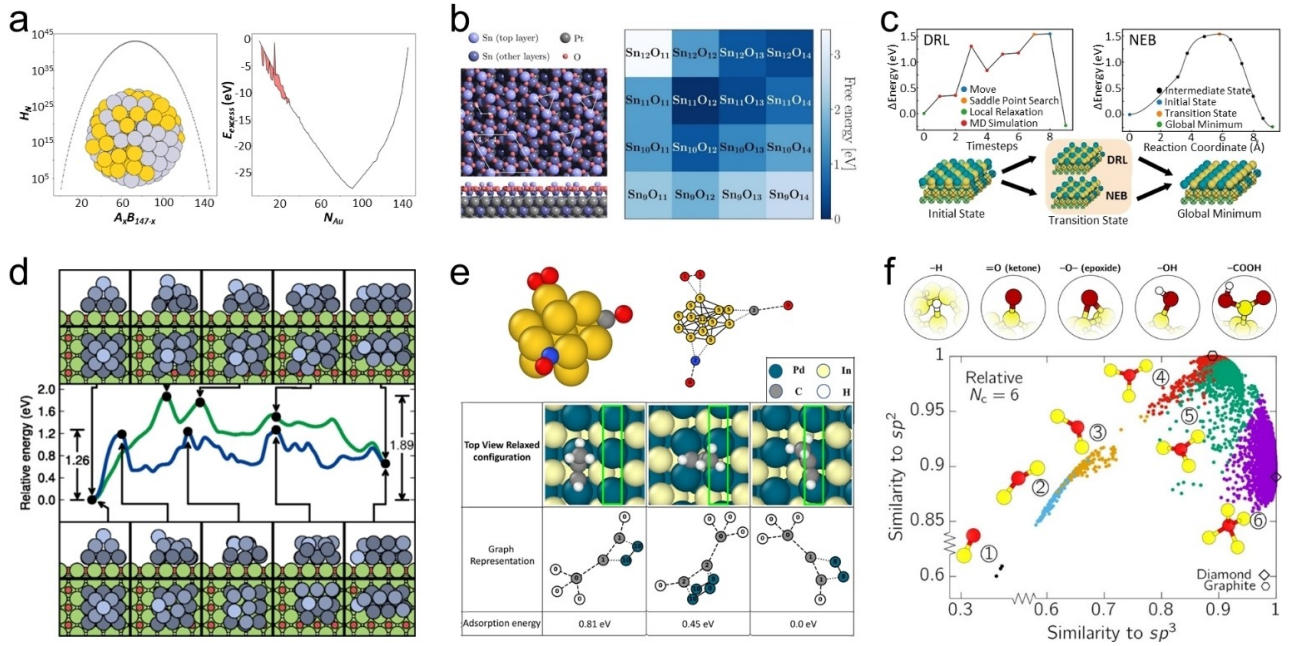


Figure 4. a) Left: Number of homotops as a function of composition; inset: randomly ordered PtAu 147-atom icosahedron. Right: Convex hull located with ML-accelerated genetic algorithm (GA) using DFT calculations. Reproduced with permission.[20] Copyright 2019, Springer Nature. b) Left: Model for lowest-energy $Sn_{11}O_{12}$ structure corresponding to the observed (4×4) phase. Right: Computed free energy for different structures under experimental conditions (10^{-5} mbar, 600 °C) relative to $Sn_{11}O_{12}$ structure. Reproduced with permission.[24] Copyright 2022, Wiley-VCH. c) Left: Example energy pathway to global minimum from deep reinforcement learning (DRL). Right: A minimum energy pathway created by NEB to same global minimum. Reproduced with permission.[34] Copyright 2021, IOP Publishing. d) Transformation of a hollow pyramid to an elongated structure. Lowest-energy path is shown in blue, whilst the pathway resulting from minimizing the 2-norm mapping is shown as reference in green. Top and bottom panels show images of side and top view of the initial state, transition states, and final state for the two pathways. Reproduced with permission.[36] Copyright 2018, American Physical Society. e) Graph theory based algorithm to generate graphs for a given atomic model. Determining unique structures for multidentate adsorbate $CHCH_3$ on a PdIn(021) stepped surface. Reproduced with permission.[40] Copyright 2020, Springer Nature. f) Clustering analysis with six target clusters and relative coherence criterion with functional groups explored. Reproduced with permission.[44] Copyright 2018, American Chemical Society.

selection of appropriate physical quantities as features. One set of physical quantities, including the cohesive energy (ΔE), lattice parameters (a and c), bulk modulus (B), and vacancy formation energy (E_f) was used as features for a KRR model to evidence relationships between the metal surface energy and physical parameters.[25] Another set of physical quantities, the Mendeleev Python Library,[26] provides a series of quantifiable features from physical and chemical properties classified by elements. For selection of physical quantities/features, the least absolute shrinkage and selection operator (LASSO)[27] and sure independence screening (SIS)[28] are used to rank coefficients of different features to boost t accuracy of prediction and interpretability of the resulting statistical models. Feature extraction and classification principal component analysis (PCA)[29] are used as a dimensional reduction tool to accelerate the generation of new knowledge. These methods are integrated into CatLearn[30] for graphical-based atomic structure enumeration in surface science and catalysis. To summarize the features selected by the ML process and the descriptors composed of features, knowledge graph (KG)[31] is a potential method to extract, organize, and represent physical-inspired knowledge from data across surface structures[32] for use with advanced applications.

2.2. Dynamic Stability of Surface Structures

In addition to comparing thermodynamic surface energies, the kinetic energy barrier for interconversion between different structures is an important indicator for the stability of surface structures. The nudged elastic band (NEB)[33] method, based on the DFT computation, is suitable for searching transition states during transformation of different surface structures because it is more accurate and reliable than dynamic methods based on classical force field. However, because of the need to search for multiple points on the potential energy surface, the highly significant computations needed are a practical difficulty in applying this method to complex structures.

ML acceleration is applied to push the limits in the search for kinetic transition states and energy barriers. Yoon et al.[34] applied deep reinforcement learning (DRL)[35] to iteratively change the positions of atoms in the near-surface region to generate kinetic pathways to accessible local minima involving changes in the surface compositions (Figure 4c). As the surface atoms move in each time-step, the energy and atomic coordinates of the structure are used as features and fed into the actor network of DRL to determine the next surface structure. The highest energy

point obtained between the two metastable structures is the ML-predicted transition state structure. Kinetic pathways in surface reconstruction obtained with this method agree well with NEB computational findings, and demonstrate the accuracy of ML based methods. Kolsbjerg et al.[36] used an NN-enhanced EA sampling to find kinetic transition states between different configurations of Pt nanoparticles supported on an MgO surface (Figure 4d). Through EA parent for certain structural candidates, the success rate for locating the global minimum rises to a satisfactory level to screen for "best" pathway mapping between different particle shapes to replace high-cost optimizations for the transition barriers. Good computational speed enabled the screening of hundreds of different pathways.

2.3. Representing Complex Surface Structures

For complex surface structures, especially composite or amorphous surfaces, conventional modeling and identification are difficult in practice because of surface irregularities. ML methods can determine specific atomic structure descriptors to address this, however. A typical method uses graph neural networks (GNN)[37] to transform the atomic structure into graphs with atoms as the nodes, and bonds as the edges. The physical/chemical properties for each atom are added to each node as features to represent different elements. In this way, each surface structure can be transformed into a matrix of specific data, which, together with its computed surface energy, can be used as a training set for neural network (NN) algorithms to predict the surface energy of new structures. Based on GNN, Palizhati et al.[38] used a modified crystal graph convolutional neural network (CGCNN) on a dataset of ca. 3000 surface structures to predict surface energies of intermetallic alloys across 36 elements and 47 space groups. The graph theory (GT) was also applied to predict the initial structures with surface adsorbates[39] and with different intermediate coverages[40] to significantly reduce computational time for geometry optimization (Figure 4e).

Another method to read and identify the atomic structures for complex surfaces is smooth overlap of atomic position (SOAP).[41] This provides an "intuitive" measure of dissimilarity between atomic environments and transforms it into a numerical descriptor that can then be fed to ML algorithms. This can combine Gaussian approximation potential (GAP)[49] frameworks to learn and predict the energy of an arbitrary atomic configuration. In this way, more irregular and amorphous surface structures than flat alloy surfaces are resolved. Bartók et al.[43] used SOAP-GAP to predict silicon surfaces with distortions during the reconstruction. Caro et al.[44] applied SOAP-GAP to conduct a comprehensive and systematic assessment of various atomic motifs for carbon material surfaces to resolve flexibility and amorphous matters (Figure 4f).

3. Actual Catalyst Surface Structures

In addition to intrinsic stability, surface structure can undergo significant change under experimental conditions, especially during reactions. In-situ experiments in catalysis are nearly always difficult to perform because of the interference of temperature and atmospheric pressure surrounding the actual system,[2,45] together with voltage and solvent in the electrochemical system.[46] With conventional DFT computations, the incorporation of these factors is possible via numerical correction. Ab initio thermodynamics adds temperature (T) and atmospheric pressure (P) to chemical potential (μ) computation [Eq. (3)].[47]

$$\mu(P, T) = \mu_0 + k_B T \ln\left(\frac{P}{P^0}\right) \tag{3}$$

With the computational hydrogen electrode (CHE),[48] potential (U) and pH are introduced to the electrochemical potential of protons [Eq. (4)].

$$\mu_H(U, \mathrm{pH}) = E_H + eU - k_B T \log_{10}(\mathrm{pH}) \tag{4}$$

In addition, the simulation of the solvent at the solid–liquid interface is difficult in practice because of the significantly large number of atoms that need to be considered. Therefore, operando computational methods that build experimental conditions and micro-environment into modelling, whilst meeting the need for greater accuracy rather than simple numerical methods, have significant practical difficulty in theoretical research.

Molecular dynamics (MD) and Monte Carlo (MC) are methods to extend the simulation from the microscopic to mesoscopic scale (Figure 5a). However, the accuracy of the classical MD and MC computation is limited because quantum effects, such as tunneling and zero-point vibrational energy, are neglected when the classical force field is used. Another significant limitation with MD is inaccuracy in modelling the making/breaking of chemical bonds. Ab initio molecular dynamics (AIMD) uses quantum force fields based on electronic structure, whilst ab initio Monte Carlo requires parametrization of a significantly large number of high-accuracy computations. Development is limited because of high computational cost(s). The important relevance of ML is in its ability to combine advantages of DFT with MD and MC through predicting energy via machine learning potential (MLP)[49] and force via machine learning force field (MLFF),[50] for an accurate description of chemical systems. In this way, ML can be used to obtain a more accurate simulation for surface structure evolution in larger time-/length-scale (Figure 5b, c). The ML method is accurate for interpolation but less so for extrapolation. The accuracy of MLP and MLFF frequently depends on similarities in the training data and the simulation system. Δ-machine learning (Δ-ML)[51] with confidence interval can be used to solve this. Via learning the difference between the basis potential and the reference to determine Δ-MLP,[52] and similarly, for Δ-MLFF, MLP and MLFF can be updated

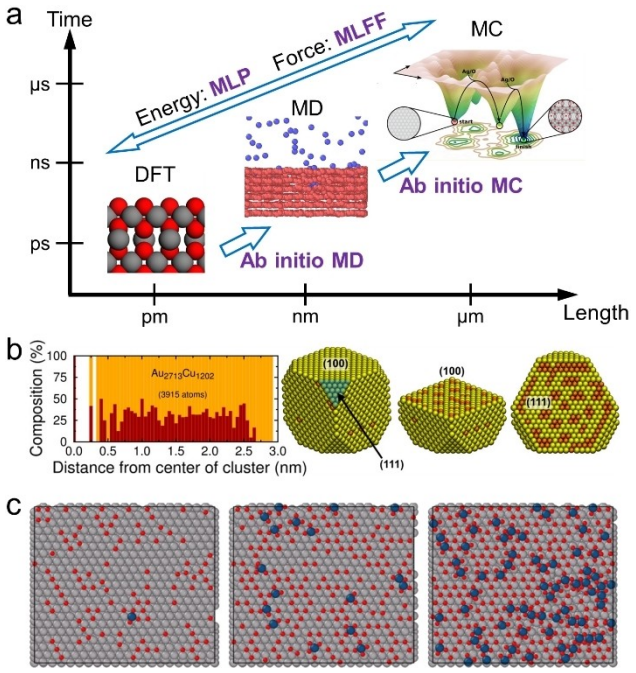


Figure 5. a) Operando computational methods at different length- and time-scales in combination with ML. Insert for MC is reprinted with permission.[54] Copyright 2019, American Chemical Society. b, c) Exammples of surface structure study with greater accuracy using ML on a large time-/length-scale: b) Cu/Au nanoparticle with 3915 atoms. c) $p(20\times20)$ Pt(111) surface with differing O coverage. (b) Reprinted with permission.[70] Copyright 2014, American Chemical Society. (c) Reprinted with permission.[87] Copyright 2022, American Chemical Society.

instantly to simulate accurately the system with a 10-fold increase in speed.[53]

Addressed in the following sections is how ML can: 1) incorporate temperature and pressure into the reaction system, 2) include the solvation environment on the catalyst surface, and 3) build surface phase diagrams of the catalyst in operando.

3.1. Effect(s) of Temperature and Atmosphere

Both ab initio MD and MC have been used to model heterogeneous catalyst surface structures for temperatures greater than 0 K, where the addition of ML can make calculations faster with guaranteed accuracy. AIMD permits simulation of dynamic structures at higher temperatures, including surface melting and transformation. Faraji et al.[55] used a charge equilibration neural network technique (CENT) to generate the MLP for MD on $CaF_2(100)$ surface structures over a wide temperature range of 300–1200 K. The CENT potential applies to ionic materials because it describes charge transfer for ionic bonding,[56] and its predicted structure was verified by DFT computation. Chapman et al.[57] applied MLFF for MD simulation of Al surface melting at up to 1000 K. Halim et al.[58] used the Gaussian process (GP) force field to reduce computational demand of AIMD to simulate the complex process of migration and structural transformation of Cu–Zn surface alloying on Cu(997), and reported comparable results with STM experiment.

MC can be applied to component segregation of particle surface structure because it can accomplish larger size and time scale simulation. Neural network potentials trained with DFT-computed energies are alternatives for empirical parameters in MC simulations to improve accuracy. Elias et al.[59] used artificial neural network potential Monte Carlo simulations to simulate experimental Cu/Ce oxide nanoparticles with sizes up to $Cu_{54}Ce_{405}O_{834}$ and annealing temperatures from 5000 to 300 K. Yang et al.[60] trained a NN framework from >5000 DFT-computed slabs for a larger-scale Monte Carlo simulation to investigate surface segregation in a ternary Cu–Pd–Au alloy.

Gas-phase atmosphere can influence surface structure at gas–solid interfaces. Dynamic simulation methods try to bridge any "structure gap" between well-ordered metal facets in surface science and defect-laden, real catalysts. ML-based piecewise embedded atomic neural network (PEANN) has been used to generate potential energy surface for AIMD in molecule–surface systems, with numerical savings of the order of 10^5.[61] Zhou et al.[62] applied this for dissociative chemisorption of CH_4 on defective Ir surface structures. Liu et al.[63] performed a Semi-Grand Canonical Monte Carlo (SGCMC)[64] simulation to assess surface segregation and aggregation of bimetallic metal alloys induced by surrounding acrolein molecules. Cheng et al.[65] used MD with neural network potential (NN-MD) for the oxygen-derived process for reduction of $Cu_2O(111)$ to Cu surface with square sites. In addition, structures of additional species in the micro-environment at the interface, including Re and Cs promoters, have been investigated with MLP-accelerated MD.[66]

3.2. Effect(s) of Solvation Structure

The micro-environment at the solid–liquid interface of the electrocatalysts has interactions between the surface and solvent molecules. With an increase in complexity of solvent components and surface structures, an explicit solvent model that directly places solvent molecules on the catalyst surface is needed for higher accuracy, rather than an implicit solvent model. This inclusion of solvation, in principle, can be modeled via traditional first-principle methods, such as DFT for static, and AIMD for dynamic structures. However, DFT and AIMD are computationally demanding because of complex and variable liquid structure and interactions including hydrogen bonding and chemical adsorption.

MLP significantly accelerates computation of solid–liquid interface systems,[67] and has therefore gradually become more widely used. Natarajan et al.[68] applied a neural network potential (NNP) to provide the required energies and forces at DFT level for MD simulation to assess different water-molecule arranged-structures on a series of Cu surface facets. Quaranta et al.[69] subsequently extended this to more hydrophilic ZnO surfaces. As a

larger-scale computation system, MC simulation based on NNP has been used to investigate solvent effect(s) on nanoparticles. Artrith et al.[70] combined DFT and accurate NNP with MC to compute water-solvated Au/Cu alloy nanoparticles. Additionally, a combination of SOAP structural descriptor and sketch map ML algorithm[71] has been used to predict solvation energy to quantify the number of explicit solvent molecules around local solvation environments.[72]

3.3. Significance of the Surface Phase Diagram

The surface phase diagram of a catalyst system can be obtained via summarizing a series of surface structures and selecting the most stable. Surface phase diagrams are very important for heterogeneous catalyst research, as they can directly correlate first-principles computations with experimentally available conditions, such as in-situ reaction conditions and catalyst components. Therefore, it is an important component of operando computation to establish direct correspondence between theoretical and experimental value(s). Despite having accuracy to match experimental measurement, a significantly large number of possibilities need to be considered so as to not miss key surface structures. This requires high computational cost(s).[73] ML is advantageous in predicting data because of reduced computational cost, and could produce surface phase diagrams for wide application.

One way to address the significantly large number of surface structures is to use ML methods to train DFT-computed surface free energy datasets to predict the surface phase diagram. Ulissi et al.[74] reconstructed the Pourbaix diagram for IrO_2(110), which relates differing surface terminations to potential (U) and pH, from 20 DFT relaxations using GPR. This compared with ca. 90 using typical search methods. Ghanekar et al.[75] used adsorbate chemical environment-based graph convolution neural network (ACE-GCN) to present high *OH coverage on Pt(111) and (211) surfaces, and trained on a fraction ($\approx$10 %) of the total DFT-relaxed configurations to obtain the complete Pourbaix diagram (Figure 6a). Another is to perform stochastic surface walking (SSW) on the global potential energy surface obtained from NNP training from DFT computations to search for local minima for energy. Li et al.[76] applied this SSW-NN[77] method to search for the low surface free energy for Pd_1Ag_3 surface with a series of H coverages (Figure 6b). By including explicit charge on the surface during DFT computation for the training dataset, Fang et al.[78] used this method to simulate potential-dependent Pt(110) surface phases under electrochemical condition. This work provides a positive example of the role of ML in combining DFT with operando computation.

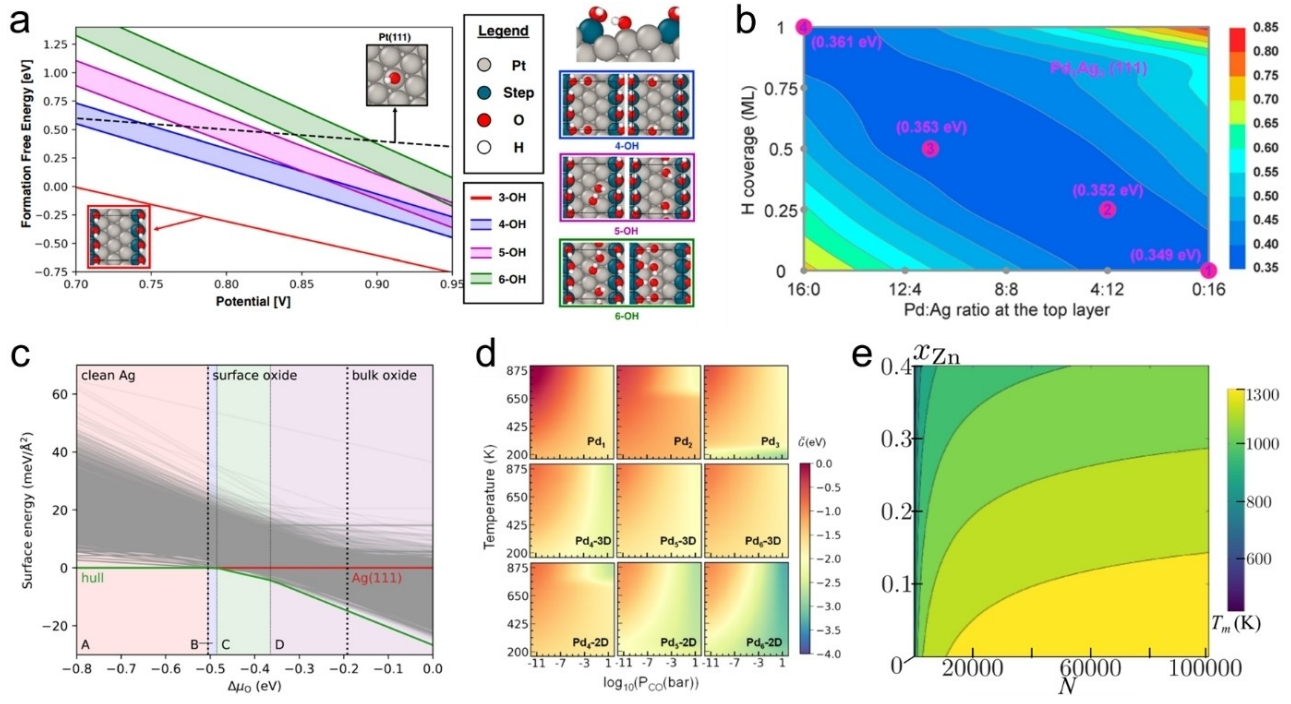


Figure 6. Summary for surface phase diagrams studied with ML. a) Ab initio Pourbaix diagram for OH* configurations on Pt(111) and (221) surfaces. Reproduced with permission.[75] Copyright 2020, Springer Nature. b) Pd–Ag–H surface contour maps for Pd–Ag–H/Pd_1Ag_3(111) at 25 °C and $P(H_2) = 0.05$ atm. Reprinted with permission.[76] Copyright 2021, American Chemical Society. c) Surface phase diagram for Ag(111) exposed to O_2 atmosphere, generated via GCMC. Reprinted with permission.[83] Copyright 2022, American Chemical Society. d) Free energy per Pd atom $\bar{G}$(eV) as a function of temperature and CO pressure. Colour indicates absolute value for $\bar{G}$ for stability. Reprinted with permission.[54] Copyright 2019, American Chemical Society. e) Size- and composition-dependent melting phase diagram for large CuZn alloy nanoparticles as a function of content of Zn (x_{Zn}) and number of atoms (N). Reprinted with permission.[86] Copyright 2021, American Chemical Society.

MD simulation, which can model dynamic changes at high temperature and at solid–liquid interfaces, was used for computation of the surface phase diagram following combination with ML with improved accuracy. Timmermann et al.[79] trained an ML-GAP for high-temperature MD simulation for annealing across all symmetry-inequivalent low-index surfaces for IrO_2 to get a surface phase diagram with dependence on temperature that was confirmed with STM, LEED, and XPS experiments. Artrith et al.[80] used artificial neural networks (ANNs) to interpolate the atomic energy from DFT for MD simulation and constructed first-principles phase diagrams for amorphous Li_xSi_{1-x} associated with Li atomic ratio x, based on $\approx$45000 structures. Li et al.[81] reported that an MLFF-based MD-simulated surface phase diagram for $Fe_3O_4(111)$ achieved results similar to DFT and simulated the structure of the aqueous solvent environment at the interface concurrently. Rosenbrock et al.[82] proposed using SOAP descriptor on liquid-phase structure for an active learning dataset for MTP to perform MD simulation of the liquid–solid transition phase diagram for Ag–Pd alloy.

MC simulation was used to construct a phase diagram for larger surface structures to automate discovery of realistic surfaces to overcome the insufficient exploration of all possible surfaces and the bias of human-selected structures pool. Wexler et al.[54] combined DFT and grand canonical Monte Carlo (GCMC) into an "ab initio GCMC" to determine oxide overlayer structures on Ag(111). More than 6000 structures were sampled to obtain the surface phase diagram for oxidation (Figure 6c), and random forests (RF) regression was used to determine the structural features that govern surface stability. Wang et al.[83] performed this ab initio GCMC simulation on CO-adsorbed Pt clusters supported on $CeO_2(111)$ surface (Figure 6d), and applied a cluster genetic algorithm (CGA)[84] to predict low-energy structures to obtain the surface phase diagram as a function of temperature and CO pressure. Additionally, because of the advantage in simulating large nanoparticles, MC-based high-dimensional neural network potentials (HDNNPs)[85] have been used to compute the melting phase diagram for CuZn alloy nanoparticles with up to $N=100000$ atoms, corresponding to an approximate diameter of 12 nm[86] (Figure 6e). Recently, Xu et al.[87] developed a pipeline to model metal surface oxidation during catalysis using large-scale MLP-based GCMC. These large-scale simulations were demonstrated as useful tools to study formation mechanism(s) of complex surface systems.

4. Characterization of Surface Structures

In addition to accelerating theoretical computation, ML is increasingly being used in conjunction with experimental characterizations to analyze experimental data and to improve the nexus between experiment and computation. For the conventional process, from theoretical design of catalyst structures to experimental characterization and confirmation, it is practically difficult to find designed structures in experiments. Additionally, there are practical difficulties to an efficient analysis of experimentally observed surface structures to provide timely feedback to model constructions. Advanced ML will reduce the human work-component and improve experimental design efficiency via automation of artificial intelligence, and shift the research workflow to permit theoretical models directly from experimental findings. Significantly, this will bridge the "material gap"[88] between experiment and theory, and deepen understanding of the mechanism(s). This has significant potential for development of surface science experiments, especially for in-situ experimental techniques that can continuously generate significant amounts of data during catalytic reaction(s).

4.1. Identifying Microscopic Images

Researchers have developed ML applications for automatically recognizing images. These can be combined with experimental characterizations that generate images to give automated structure identification. The main benefits are the: 1) savings in manpower and reductions in human error, and 2) potential to obtain more rational segmentation results than with conventional methods.

Microscopic images of rock-minerals have been automatically classified and identified though deep learning NN because of its automatic extraction feature in image analyses.[93] More advanced transmission X-ray microscopy (TXM) was applied to make phase contrast tomography for active particles, and a mask regional convolutional neural network (Mask R-CNN)[94] has been used in identification and segmentation of every individual particle for reconstruction of a (3D) volume.[95] Atomic precision scanning probe microscopy (SPM), including scanning tunneling microscopy (STM) and force in atomic force microscopy (AFM), with ML architecture based on ANNs was used to detect nematic order in STM images and other forms for symmetry breaking.[96] Krull et al.[89] demonstrated DeepSPM, an autonomous system capable of continuous SPM image acquisition, to distinguish a "good" from a "bad" probe (Figure 7a). It was trained using a classifier convolutional neural network (CNN) model from a dataset of 7589 constant-current STM topography images of MgPc molecules on Ag(100) surface classified by human operation.

4.2. Spectra Analyses

The combination of ML and spectral characterization experiments generally uses the spectra dataset determined from theoretical simulation as an intermediate bridge to establish a relationship between the spectra and the structure, and to achieve improved analyses and understanding of the experimental data. ML: 1) better reflects the mathematical relationship between theoretical simulations and experimental spectra, 2) provides a more rational theoretical basis for the assignment and interpretation of spectral peaks, and 3) is expected to find a reasonable structure for uncertain signals. Current interdisciplinary

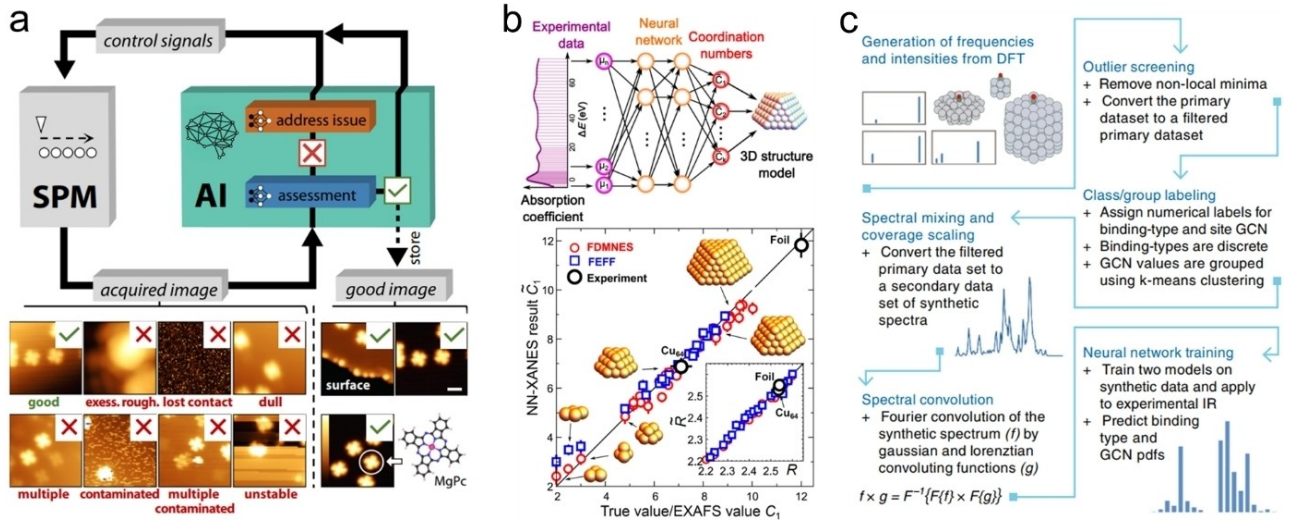


Figure 7. ML in experimental characterization of surface structures of heterogeneous catalysts. a) Schematic for DeepSPM, a ML-based autonomous artificial intelligence system for autonomous-scanning probe microscopy. Reproduced with permission.[89] Copyright 2020, Springer Nature. b) Top: Workflow to solve structure of a metal catalyst from experimental XANES. Bottom: Confirmation of NN accuracy with theoretical particle-averaged XANES spectra, and with experimental data. Reprinted with permission.[90] Copyright 2017, American Chemical Society. Reprinted with permission.[91] Copyright 2018, American Chemical Society. c) Workflow combining expert knowledge, ML, and spectroscopic data to close materials gap via physics- and data-driven surrogate models for application in experimental IR spectra. Reproduced with permission.[92] Copyright 2020, Springer Nature.

research with ML on surface structure of catalysts focuses on X-ray absorption near-edge structure (XANES) spectroscopy, and vibrational spectroscopy including infrared (IR) and Raman spectra. However, wider application of more spectral methods combined with ML can be expected.

XANES spectroscopy findings establish the coordination environment around atoms, such as coordination number (CN), which can be simulated by ab initio codes including FEFF[97] and FDMNES[98] in addition to experimental measurement. Timoshenko et al.[90] trained ANN with theoretical XANES data to predict the CN for Pt, which was then related to the 3D structure of supported Pt nanoparticles (Figure 7b). This methodology permits solution of the structure of a metal catalyst from experimental XANES via reconstructing average size, shape, and morphology of well-defined nanoparticles. The structure for Cu nanoparticles supported on a ZrO_x surface,[91] and CN of Cu in CuO_x catalyst[99] was solved in this way.

To address increased structural complexity of transition-metal oxides, Carbone et al.[100] proposed a robust method combining a CNN model and PCA analysis to establish relationships between XANES spectra and the local chemical environment surrounding an atom. Guda et al.[101] used LASSO and ridge regression in an interpretable ML assessment of descriptors for XANES spectra of $Fe(SiO_4)_N$ clusters based on structural parameters.

Vibrational spectroscopy captures details of most surface vibrational modes and can be simulated with frequencies and intensities through DFT computations of phonons. Lansford et al.[92] computed a dataset for IR spectra for CO adsorbed on Pt nanoparticles and implemented multinomial regression via neural network ensembles to infer detailed surface microstructure, including adsorption sites and coverage from experimental spectra (Figure 7c). Wang et al.[102] established a quantitative spectrum–property relationship based on sure independence screening and sparsifying operator (SISSO)[103] predicted descriptors from simulated IR/Raman spectra, from 70500 adsorption configurations for CO/NO on metal and alloy. To establish the dynamic structure of larger molecules and water solvent environment on the surface, Hu et al.[104] developed a random, forest protocol based on a large data set of AIMD computations for surface-enhanced Raman spectroscopy (SERS) for a *trans*-1,2-bis(4-pyridyl)ethylene molecule adsorbed on Au surface.

4.3. Inverse Construction

Via application of the concept of "inverse design"[10] in ML, the conventional, theoretical simulation of a characterization image/spectrum can be inversed to deduce an atomic structure. Its significance is to reverse the order from theoretical model to experimental observation and confirmation, and change it to obtain a theoretical model from experimental observation. This significantly speeds up the exploration of mechanisms and materials design by significantly reducing the number of candidates. Advanced ML-based processing can exhibit faster convergence and better stability in relatively difficult problems.[105] The workflow is to train the ML model with theoretical simulated datasets that have been benchmarked via experiment, so that it can generate atomic structure from experimental characterized image/spectrum. Application of this method to a series of images/spectra obtained from in-situ experiments can yield the operando dynamic structure of catalyst formation and

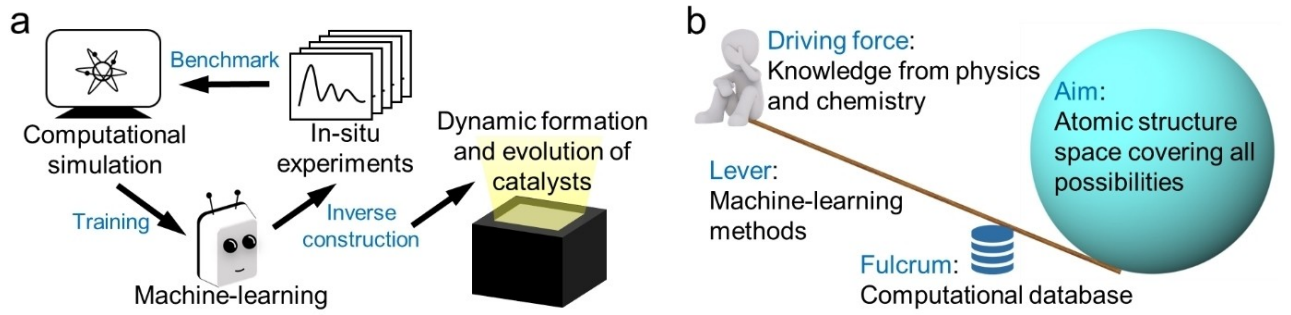


Figure 8. Outlook schematic for combined ML and experiment. a) Dynamic mechanism for catalyst structure formation and evolution through ML-based, inverse construction of atomic structure model from in-situ characterization experiments. b) "Give me a dataset, and with machine-learning I will explore the structural world." Quoted and modified from Archimedes.

evolution, thereby solving "black box" problems in catalysis (Figure 8a). An "inverting" of XANES spectra to catalytic activity descriptors has been reported by Timoshenko et al.[106] Additionally, practical tools to generate atomic structures with Cartesian atomic coordinates were developed to push this inverse reconstruction to wider application.[107]

5. Summary and Outlook

ML is increasingly applied to study the surface structure of heterogeneous catalysts. From accelerating and combining multi-scaling modelling methods, including DFT, MD, and MC, to identifying, analyzing, and mining experimental characterization data, advanced ML tools are replacing empirical approaches and turning them into data-driven methods. This dynamic is expected to address the present poor understanding of the catalyst surface structure especially that for in-situ dynamic structural change(s) under reaction conditions, and boost developments in catalysis. Based on this, we conclude with the following outlook:

1. Although experimental characterization can give accurate and intuitive cognition, it is not efficient in producing the sufficiently big dataset required for ML. Theoretical computation and simulations based on experimental findings can grow the database needed, however, for ML training. Theoretical computation therefore will remain important to assist ML and drive the growing understanding of the whole structure from initial physics and chemistry knowledge (Figure 8b). In addition, there is room for improvement in agreement between computational simulation and experiment under reaction conditions such as light, electricity, plasma, and micro-environments including solvents and promoters. ML methods will provide a bridge between theoretical simulation and experiment, and aid the development of operando computation methodology.
2. The continuing study of catalyst surface stability will serve catalyst development and become part of the workflow for catalyst design. Designed catalysts need to be assessed for stability during reaction. Conventional computing methods are, however, not adequate to meet the demands for significantly numerous modelling and operando computations. ML will address this as part of the workflow to predict and combine different simulation scales. There are a number of reports[108] on using ML to first screen stable surface structures, and then proceed to catalyst design from these structures. This is expected to become the established ML method and become widespread in predicting catalyst surface structure. Importantly, when a catalyst stability window[109] does not match the reaction conditions, physics-driven interpretable ML can be combined to tune the stability/activity window, for example, via incorporating oxidative/reductive heteroatoms, selecting suitable substrates or support materials, or changing reaction conditions with solvents/micro-environments to rationally modify the catalyst to maintain stability of the active site during reaction.
3. Innovative artificial-intelligence and ML methods will be needed to develop catalysis and surface science. In particular methods that provide design guidance. For example, natural language processing (NLP)[110] can extract knowledge of surface science from significant literature via a search of (considered) keywords. It is expected to underpin therefore more generalized approaches to stabilize a specific active site on catalyst surfaces. Establishing a suitable transformation language[111] for surface microstructures is practically promising. Abstract grammars[112] will be developed that underpin theoretical level guidance. With discovery of structure–activity relationships and reaction mechanisms, rational ML design will become the new norm to describe surface behaviour and to design efficient heterogeneous catalysts

Acknowledgements

The authors gratefully acknowledge financial support from the Australian Research Council (ARC) through Discovery Project programs (FL170100154, DP220102596 & FT190100636). Open Access publishing facilitated by The University of Adelaide, as part of the Wiley - The University of Adelaide agreement via the Council of Australian University Librarians.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Heterogeneous Catalysts · Machine Learning (ML) · Operando Computation · Surface Structures · In-Situ Characterization

[1] C. B. Duke, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3858–3864.
[2] a) M. J. Weaver, X. Gao, *Annu. Rev. Phys. Chem.* **1993**, *44*, 459–94; b) D. M. Kolb, *Angew. Chem. Int. Ed.* **2001**, *40*, 1162–1181; *Angew. Chem.* **2001**, *113*, 1198–1220.
[3] A. D. Becke, *J. Chem. Phys.* **2014**, *140*, 18A301.
[4] J. K. Nørskov, T. Bligaard, J. Rossmeisl, C. H. Christensen, *Nat. Chem.* **2009**, *1*, 37–46.
[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
[6] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, A. Tkatchenko, *Chem. Rev.* **2021**, *121*, 9816–9872.
[7] T. Zubatiuk, O. Isayev, *Acc. Chem. Res.* **2021**, *54*, 1575–1585.
[8] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, K.-R. Müller, *Chem. Rev.* **2021**, *121*, 10142–10186.
[9] J. A. Esterhuizen, B. R. Goldsmith, S. Linic, *Nat. Catal.* **2022**, *5*, 175–184.
[10] B. Sanchez-Lengeling, A. Aspuru-Guzik, *Science* **2018**, *361*, 360–365.
[11] J. S. Dramsch, *Adv. Geophys.* **2020**, *61*, 1–55.
[12] Y. Zheng, Y. Jiao, A. Vasileff, S.-Z. Qiao, *Angew. Chem. Int. Ed.* **2018**, *57*, 7568–7579; *Angew. Chem.* **2018**, *130*, 7690–7702.
[13] G. Li, C. Felser, *Appl. Phys. Lett.* **2020**, *116*, 070501.
[14] A. Hjorth Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. M. T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K. W. Jacobsen, *J. Phys. Condens. Matter* **2017**, *29*, 273002.
[15] K. Reuter, *Catal. Lett.* **2016**, *146*, 541–563.
[16] D. Lewis, W. Gale, *In Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval* **1994**, ACM/Springer, pp. 3–12.
[17] B. Settles, *Morgan & Claypool* **2012**.
[18] S. Katoch, S. S. Chauhan, V. Kumar, *Multimed Tools Appl.* **2021**, *80*, 8091–8126.
[19] a) A. R. Oganova, C. W. Glass, *J. Chem. Phys.* **2006**, *124*, 244704; b) L. B. Vilhelmsen, B. Hammer, *J. Chem. Phys.* **2014**, *141*, 044711.
[20] P. C. Jennings, S. Lysgaard, J. S. Hummelshøj, T. Vegge, T. Bligaard, *npj Comput. Mater.* **2019**, *5*, 46.
[21] T. L. Jacobsen, M. S. Jørgensen, B. Hammer, *Phys. Rev. Lett.* **2018**, *120*, 026102.
[22] S. Hajinazar, E. D. Sandoval, A. J. Cullo, A. N. Kolmogorov, *Phys. Chem. Chem. Phys.* **2019**, *21*, 8729.
[23] M. K. Bisbo, B. Hammer, *Phys. Rev. Lett.* **2020**, *124*, 086102.
[24] L. R. Merte, M. K. Bisbo, I. Sokolović, M. Setvín, B. Hagman, M. Shipilin, M. Schmid, U. Diebold, E. Lundgren, B. Hammer, *Angew. Chem. Int. Ed.* **2022**, *61*, e202204244; *Angew. Chem.* **2022**, *134*, e202204244.
[25] Y. Zhang, X. Xu, *Mater. Chem. Phys.* **2021**, *267*, 124622.
[26] L. M. Mentel, mendeleev—a Python resource for properties of chemical elements, ions and isotopes, 2014—. Available at: https://github.com/lmmentel/mendeleev.
[27] R. Tibshirani, *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288.
[28] J. Fan, J. Lv, *J. R. Stat. Soc. Ser. B* **2008**, *70*, 849–911.
[29] I. T. Jolliffe, J. Cadima, *Philos. Trans. R. Soc. A* **2016**, *374*, 20150202.
[30] M. H. Hansen, J. A. G. Torres, P. C. Jennings, Z. Wang, J. R. Boes, O. G. Mamun, T. Bligaard, arXiv:1904.00904v1.
[31] D. Fensel, U. Şimşek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk, I. Toma, J. Umbrich, A. Wahler, *Knowledge Graphs*, Springer, Berlin, **2020**.
[32] A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders, R. Fushimi, *ACS Catal.* **2018**, *8*, 7403–7429.
[33] G. Henkelman, B. P. Uberuaga, H. Jonsson, *J. Chem. Phys.* **2000**, *113*, 9901.
[34] J. Yoon, Z. Cao, R. K. Raju, Y. Wang, R. Burnley, A. J. Gellman, A. B. Farimani, Z. W. Ulissi, *Mach. Learn.: Sci. Technol.* **2021**, *2*, 045018.
[35] a) M.-P. V. Christiansen, H. L. Mortensen, S. A. Meldgaard, B. Hammer, *J. Chem. Phys.* **2020**, *153*, 044107; b) S. A. Meldgaard, H. L. Mortensen, M. S. Jørgensen, B. Hammer, *J. Phys. Condens. Matter* **2020**, *32*, 404005.
[36] E. L. Kolsbjerg, A. A. Peterson, B. Hammer, *Phys. Rev. B* **2018**, *97*, 195424.
[37] a) F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, *IEEE Trans. Neural Netw.* **2009**, *20*, 61–80; b) T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120*, 145301.
[38] A. Palizhati, W. Zhong, K. Tran, S. Back, Z. W. Ulissi, *J. Chem. Inf. Model.* **2019**, *59*, 4742–4749.
[39] J. R. Boes, O. Mamun, K. Winther, T. Bligaard, *J. Phys. Chem. A* **2019**, *123*, 2281–2285.
[40] S. Deshpande, T. Maxson, J. Greeley, *npj Comput. Mater.* **2020**, *6*, 79.
[41] A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, *87*, 184115.
[42] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, *Phys. Rev. Lett.* **2010**, *104*, 136403.
[43] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, M. Ceriotti, *Sci. Adv.* **2017**, *3*, e1701816.
[44] M. A. Caro, A. Aarva, V. L. Deringer, G. Csányi, T. Laurila, *Chem. Mater.* **2018**, *30*, 7446–7455.
[45] E. M. Stuve, A. Krasnopoler, D. E. Sauer, *Surf. Sci.* **1995**, *335*, 177–185.
[46] Y. Jiao, Y. Zheng, M. Jaroniecb, S. Z. Qiao, *Chem. Soc. Rev.* **2015**, *44*, 2060–2086.
[47] K. Reuter, M. Scheffler, *Phys. Rev. B* **2002**, *65*, 035406.
[48] J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, H. Jonsson, *J. Phys. Chem. B* **2004**, *108*, 17886–17892.
[49] J. Behler, *Chem. Rev.* **2021**, *121*, 10037–10072.
[50] R. Jinnouchi, F. Karsai, G. Kresse, *Phys. Rev. B* **2019**, *100*, 014105.
[51] a) R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. v. Lilienfeld, *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096; b) M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, K. Burke, *Nat. Commun.* **2020**, *11*, 5223.
[52] A. Nandi, C. Qu, P. L. Houston, R. Conte, J. M. Bowman, *J. Chem. Phys.* **2021**, *154*, 051102.
[53] J. Xu, X.-M. Cao, P. Hu, *J. Chem. Theory Comput.* **2021**, *17*, 4465–4476.
[54] R. B. Wexler, T. Qiu, A. M. Rappe, *J. Phys. Chem. C* **2019**, *123*, 2321–2328.
[55] S. Faraji, S. A. Ghasemi, B. Parsaeifard, S. Goedecker, *Phys. Chem. Chem. Phys.* **2019**, *21*, 16270–16281.
[56] S. A. Ghasemi, A. Hofstetter, S. Saha, S. Goedecker, *Phys. Rev. B* **2015**, *92*, 045131.

[57] J. Chapman, R. Ramprasad, *J. Phys. Chem. C* **2020**, *124*, 22127–22136.
[58] H. H. Halim, Y. Morikawa, *ACS Phys. Chem. Au* **2022**, *2*, 430–447.
[59] J. S. Elias, N. Artrith, M. Bugnet, L. Giordano, G. A. Botton, A. M. Kolpak, Y. Shao-Horn, *ACS Catal.* **2016**, *6*, 1675–1679.
[60] Y. Yang, Z. Guo, A. J. Gellman, J. R. Kitchin, *J. Phys. Chem. C* **2022**, *126*, 1800–1808.
[61] B. Jiang, J. Li, H. Guo, *J. Phys. Chem. Lett.* **2020**, *11*, 5120–5131.
[62] X. Zhou, Y. Zhang, H. Guo, B. Jiang, *Phys. Chem. Chem. Phys.* **2021**, *23*, 4376–4385.
[63] M. Liu, Y. Yang, J. R. Kitchin, *J. Chem. Phys.* **2021**, *154*, 134701.
[64] D. P. Landau, K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, 4th ed., Cambridge University Press, Cambridge, **2014**.
[65] D. Cheng, Z.-J. Zhao, G. Zhang, P. Yang, L. Li, H. Gao, S. Liu, X. Chang, S. Chen, T. Wang, G. A. Ozin, Z. Liu, J. Gong, *Nat. Commun.* **2021**, *12*, 395.
[66] B. W. J. Chen, B. Wang, M. B. Sullivan, A. Borgna, J. Zhang, *ACS Catal.* **2022**, *12*, 2540–2551.
[67] C. Schrana, F. L. Thiemanna, P. Rowea, E. A. Müller, O. Marsalekf, A. Michaelides, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2110077118.
[68] S. K. Natarajan, J. Behler, *Phys. Chem. Chem. Phys.* **2016**, *18*, 28704.
[69] V. Quaranta, J. Behler, M. Hellström, *J. Phys. Chem. C* **2019**, *123*, 1293–1304.
[70] N. Artrith, A. M. Kolpak, *Nano Lett.* **2014**, *14*, 2670–2676.
[71] S. De, A. P. Bartok, G. Csányi, M. Ceriotti, *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.
[72] Y. Basdogan, M. C. Groenenboom, E. Henderson, S. De, S. B. Rempe, John A. Keith, *J. Chem. Theory Comput.* **2020**, *16*, 633–642.
[73] C. Griesser, H. Li, E.-M. Wernig, D. Winkler, N. S. Nia, T. Mairegger, T. Götsch, T. Schachinger, A. Steiger-Thirsfeld, S. Penner, D. Wielend, D. Egger, C. Scheurer, K. Reuter, J. Kunze-Liebhäuser, *ACS Catal.* **2021**, *11*, 4920–4928.
[74] Z. W. Ulissi, A. R. Singh, C. Tsai, J. K. Nørskov, *J. Phys. Chem. Lett.* **2016**, *7*, 3931–3935.
[75] P. G. Ghanekar, S. Deshpande, J. Greeley, *Nat. Commun.* **2022**, *13*, 5788.
[76] X.-T. Li, L. Chen, C. Shang, Z.-P. Liu, *J. Am. Chem. Soc.* **2021**, *143*, 6281–6292.
[77] S.-D. Huang, C. Shang, P.-L. Kang, X.-J. Zhang, Z.-P. Liu, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2019**, *9*, e1415.
[78] Y.-H. Fang, D.-D. Song, H.-X. Li, Z.-P. Liu, *J. Phys. Chem. C* **2021**, *125*, 10955–10963.
[79] J. Timmermann, F. Kraushofer, N. Resch, P. Li, Y. Wang, Z. Mao, M. Riva, Y. Lee, C. Staacke, M. Schmid, C. Scheurer, G. S. Parkinson, U. Diebold, K. Reuter, *Phys. Rev. Lett.* **2020**, *125*, 206101.
[80] N. Artrith, A. Urban, G. Ceder, *J. Chem. Phys.* **2018**, *148*, 241711.
[81] X. Li, W. Paier, J. Paier, *Front. Chem.* **2020**, *8*, 601029.
[82] C. W. Rosenbrock, K. Gubaev, A. V. Shapeev, L. B. Pártay, N. Bernstein, G. Csányi, G. L. W. Hart, *npj Comput. Mater.* **2021**, *7*, 24.
[83] Y. Wang, Y.-Q. Su, E. J. M. Hensen, D. G. Vlachos, *Chem. Mater.* **2022**, *34*, 1611–1619.
[84] Y. Wang, Y.-Q. Su, E. J. M. Hensen, D. G. Vlachos, *ACS Nano* **2020**, *14*, 13995–14007.
[85] J. Behler, M. Parrinello, *Phys. Rev. Lett.* **2007**, *98*, 146401.
[86] J. Weinreich, M. L. Paleico, J. Behler, *J. Phys. Chem. C* **2021**, *125*, 14897–14909.
[87] J. Xu, W. Xie, Y. Han, P. Hu, *ACS Catal.* **2022**, *12*, 14812–14824.
[88] Y. Wang, C. Wöll, *Chem. Soc. Rev.* **2017**, *46*, 1875–1932.
[89] A. Krull, P. Hirsch, C. Rother, A. Schiffrin, C. Krull, *Commun. Phys.* **2020**, *3*, 54.
[90] J. Timoshenko, D. Lu, Y. Lin, A. I. Frenkel, *J. Phys. Chem. Lett.* **2017**, *8*, 5091–5098.
[91] J. Timoshenko, A. Halder, B. Yang, S. Seifert, M. J. Pellin, S. Vajda, A. I. Frenkel, *J. Phys. Chem. C* **2018**, *122*, 21686–21693.
[92] J. L. Lansford, D. G. Vlachos, *Nat. Commun.* **2020**, *11*, 1513.
[93] Y. Zhang, M. Li, S. Han, Q. Ren, J. Shi, *Sensors* **2019**, *19*, 3914.
[94] K. He, G. Gkioxari, P. Dollár, R. Girshick, *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397.
[95] Z. Jiang, J. Li, Y. Yang, L. Mu, C. Wei, X. Yu, P. Pianetta, K. Zhao, P. Cloetens, F. Lin, Y. Liu, *Nat. Commun.* **2020**, *11*, 2310.
[96] J. B. Goetz, Y. Zhang, M. J. Lawler, *SciPost Phys.* **2020**, *8*, 087.
[97] J. J. Rehr, J. J. Kas, F. D. Vila, M. P. Prange, K. Jorissen, *Phys. Chem. Chem. Phys.* **2010**, *12*, 5503–5513.
[98] O. Bunău, Y. Joly, *J. Phys. Condens. Matter* **2009**, *21*, 345501.
[99] Y. Liu, N. Marcella, J. Timoshenko, A. Halder, B. Yang, L. Kolipaka, M. J. Pellin, S. Seifert, S. Vajda, P. Liu, A. I. Frenkel, *J. Chem. Phys.* **2019**, *151*, 164201.
[100] M. R. Carbone, S. Yoo, M. Topsakal, D. Lu, *Phys. Rev. Mater.* **2019**, *3*, 033604.
[101] S. A. Guda, A. S. Algasov, A. A. Guda, A. Martini, A. N. Kravtsova, A. L. Bugaev, L. V. Guda, A. V. Soldatov, *J. Surf. Invest. X-Ray Synchrotron Neutron Tech.* **2021**, *15*, 934–938.
[102] X. Wang, S. Jiang, W. Hu, S. Ye, T. Wang, F. Wu, L. Yang, X. Li, G. Zhang, X. Chen, J. Jiang, Y. Luo, *J. Am. Chem. Soc.* **2022**, *144*, 16069–16076.
[103] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L. M. Ghiringhelli, *Phys. Rev. Mater.* **2018**, *2*, 083802.
[104] W. Hu, S. Ye, Y. Zhang, T. Li, G. Zhang, Y. Luo, S. Mukamel, J. Jiang, *J. Phys. Chem. Lett.* **2019**, *10*, 6026–6031.
[105] K. Xu, E. Darve, *J. Comput. Phys.* **2022**, *453*, 110938.
[106] J. Timoshenko, A. I. Frenkel, *ACS Catal.* **2019**, *9*, 10192–10211.
[107] V. Fung, S. Jia, J. Zhang, S. Bi, J. Yin, P. Ganesh, **2022**, arXiv:2207.13227.
[108] a) R. A. Flores, C. Paolucci, K. T. Winther, A. Jain, J. A. G. Torres, M. Aykol, J. Montoya, J. K. Nørskov, M. Bajdich, T. Bligaard, *Chem. Mater.* **2020**, *32*, 5854–5863; b) S. Ma, Z.-P. Liu, *Nat. Commun.* **2022**, *13*, 2716.
[109] H. Li, K. Reuter, *ACS Catal.* **2022**, *12*, 10506–10513.
[110] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, *Nature* **2019**, *571*, 95–98.
[111] C. Edwards, T. Lai, K. Ros, G. Honke, H. Ji, **2022**, arXiv:2204.11817.
[112] J. T. Margraf, Z. W. Ulissi, Y. Jung, K. Reuter, *J. Phys. Chem. C* **2022**, *126*, 2931–2936.

Manuscript received: November 7, 2022
Accepted manuscript online: December 12, 2022
Version of record online: January 9, 2023