ACS Partner Journal

ACCOUNTS of materials research

Editors' Choice

pubs.acs.org/amrcda

Article

Interpretable Machine Learning for Catalytic Materials Design toward Sustainability

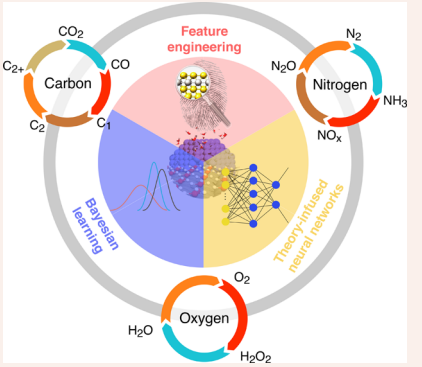
Hongliang Xin,* Tianyou Mou, Hemanth Somarajan Pillai, Shih-Han Wang, and Yang Huang

Cite This: https://doi.org/10.1021/accountsmr.3c00131

Read Online

ACCESS | Metrics & More | Article Recommendations

CONSPECTUS: Finding catalytic materials with optimal properties for sustainable chemical and energy transformations is one of the pressing challenges facing our society today. Traditionally, the discovery of catalysts or the philosopher's stone of alchemists relies on a trial-and-error approach with physicochemical intuition. Decades-long advances in science and engineering, particularly in quantum chemistry and computing infrastructures, popularize a paradigm of computational science for materials discovery. However, the brute-force search through a vast chemical space is hampered by its formidable cost. In recent years, machine learning (ML) has emerged as a promising approach to streamline the design of active sites by learning from data. As ML is increasingly employed to make predictions in practical settings, the demand for domain interpretability is surging. Therefore, it is of great importance to provide an in-depth review of our efforts in tackling this challenging issue in computational heterogeneous catalysis.



In this Account, we present an interpretable ML framework for accelerating catalytic materials design, particularly in driving sustainable carbon, nitrogen, and oxygen cycles. By leveraging the linear adsorption-energy scaling and Brønsted−Evans−Polanyi (BEP) relationships, catalytic outcomes (i.e., activity, selectivity, and stability) of a multistep reaction can often be mapped onto one or two kinetics-informed descriptors. One type of descriptor of great importance is the adsorption energies of representative species at active site motifs that can be computed from quantum-chemical simulations. To complement such a descriptor-based design strategy, we delineate our endeavors in incorporating domain knowledge into a data-driven ML workflow. We demonstrate that the major drawbacks of black-box ML algorithms, e.g., poor explainability, can be largely circumvented by employing (1) physics-inspired feature engineering, (2) Bayesian statistical learning, and (3) theory-infused deep neural networks. The framework drastically facilitates the design of heterogeneous metal-based catalysts, some of which have been experimentally verified for an array of sustainable chemistries. We offer some remarks on the existing challenges, opportunities, and future directions of interpretable ML in predicting catalytic materials and, more importantly, on advancing catalysis theory beyond conventional wisdom. We envision that this Account will attract more researchers' attention to develop highly accurate, easily explainable, and trustworthy materials design strategies, facilitating the transition to the data science paradigm for sustainability through catalysis.

1. INTRODUCTION

Catalysis, being fundamental in enabling chemical and energy transformations, holds an indispensable role in our modern society that is predominantly sustained by limited fossil reserves. Historically, human exploration of catalytic materials was purely experimental and sometimes mythical, for example, in search of the philosopher's stone by alchemists (Figure 1). The approach sometimes leads to commercially viable catalytic processes, albeit with a significant environmental footprint. To transition toward a sustainable, circular economy, a paradigm shift is needed in finding highly efficient catalytic materials preferably made of earth-abundant elements. By tapping renewable feedstocks for fuels and value-added chemicals, we can reduce our dependence on diminishing fossil resources and mitigate the impact of climate change. The design of high-performance catalysts is also integral to developing disrupting technologies for energy storage and utilization. For example, finding efficient electrocatalysts for the oxygen reduction reaction (ORR)[1] and the oxygen evolution reaction (OER)[2] is critical for large-scale commercialization of fuel cells, metal-air batteries, and water electrolyzers. One of the key challenges is to identify the optimal composition and structure of catalytic materials that are highly active and selective toward the desired products while being long-lasting under reaction conditions.

Received: July 16, 2023
Revised: October 13, 2023
Accepted: October 18, 2023

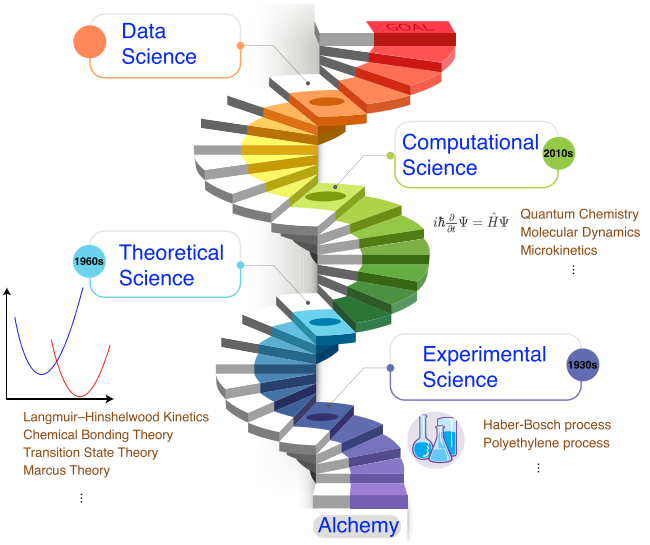


Figure 1. Evolution of scientific paradigms toward the goal of catalytic materials by design.

Over the past few decades, computational methods[3] including density functional theory (DFT), molecular dynamics, microkinetics, and continuum modeling have become powerful tools for understanding and predicting catalytic behaviors of materials, significantly accelerating the discovery of new catalysts to this day.

Exploring the vast chemical space of catalytic materials is a challenging task. The properties of a catalyst depend on intrinsic factors such as its composition and structure, which can be further complicated by defects, impurities, static and dynamic operating conditions, and external stimuli, e.g., electrochemical potentials, photons, and magnetic fields. Moreover, the tremendous cost and time-consuming nature of experimental synthesis and characterization of catalytic materials make it impossible to directly probe the entire chemical space. Within the current mainstream paradigm (Figure 1), computational science has been widely used to predict the properties of catalytic materials and screen for potential candidates.[4] However, atomistic simulations with the generalized gradient approximation (GGA) of the exchange-correlation effects of many-body systems, being reasonably accurate for describing catalytic processes on transition and noble metals, are too costly in exploring a huge design space.[5] Another challenge in this design paradigm is the need for a comprehensive understanding of reaction mechanisms as *a priori*, including the nature of active sites and occurring reaction pathways therein, which requires an integration of theoretical approaches and *in situ/operando* spectroscopic techniques.

In recent years, machine learning (ML) has risen to prominence in scientific discovery by harnessing the power of data,[6−8] signifying a paradigm shift toward a data science approach in catalytic materials design (Figure 1). One overarching theme of data-driven catalysis is to first leverage the linear adsorption-energy scaling and Brønsted−Evans−Polanyi (BEP) relationships in reaction kinetics, \understanding general activity trends of the known materials across the descriptor space spanned by adsorption energies of one or two representative species at active sites.[9] Then, ML can be employed to rapidly predict new materials with optimal descriptor values on performance maps. To enable this design workflow, we have developed an interpretable ML framework to predict reactivity properties of atomically tailored active sites.[10] We will discuss three aspects in this endeavor including physics-inspired feature engineering, Bayesian statistical learning, and theory-infused deep neural networks, which ultimately allow for interpretable catalytic materials design by harnessing the power of data. The ML models were trained using *ab initio* adsorption properties of solid surfaces, and can predict site reactivity at a diverse range of materials with physicochemical features of different fidelity levels ranging from easily accessible properties to high-level representations generated from domain engineering or extracted by learning algorithms.[11−13]

2. INTEGRATING DOMAIN KNOWLEDGE INTO MACHINE LEARNING

In heterogeneous catalysis, the development of ML models for a broad range of materials design tasks has become increasingly popular. These models are typically trained on data sets of easily accessible features of surface sites and the properties of interest, e.g., binding energies of simple adsorbates at active sites.[14] The idea is to use these pretrained models to predict reactivity descriptors of new sites and guide the design of more efficient catalysts prior to highly accurate albeit expensive quantum-chemical simulations and/or experiments. One important step in developing these ML models is the selection of features to numerically represent the materials or active sites. These features should include the relevant physical properties that govern the catalytic outcomes of the solid surfaces. Traditionally, data-driven supervised ML algorithms heavily rely on data to learn the underlying patterns and correlations between input features and output targets. However, those models might not be able to capture the underlying working mechanisms of the system being studied, i.e., casual relationships could be missing. This can lead to inaccurate predictions and biased models for truly out-of-sample systems. Incorporating scientific knowledge into ML algorithms is crucial for developing predictive models that are generalizable toward complex systems.[15] By infusing scientific principles from the domain discipline into ML algorithms, the models can be designed to unravel the underlying feature interactions that govern the system behavior. Moreover, feature selection can help to reduce the dimensionality of the problem and improve the model efficiency. Arguably, there is a trade-off between the accuracy and explainability of modern ML models.[16] By inheriting the merits of both worlds, it becomes possible to construct ML models that have data-driven precision while being scientifically informed and physically interpretable. Below we discuss our efforts along this direction by integrating domain explainability into ML algorithms.

2.1. Physics-Inspired Feature Engineering

Feature engineering is a crucial step in developing ML models for predicting the material properties. In the context of heterogeneous catalysis, the local environment of an adsorption site plays a significant role in determining its chemical reactivity. Therefore, it is essential to employ physically transparent features that can capture the reactivity properties of surface sites (Figure 2). Several types of features have been proposed, largely inspired by the theory of chemisorption, e.g., the *d*-band theory for *d*-metal catalysis.[17]

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

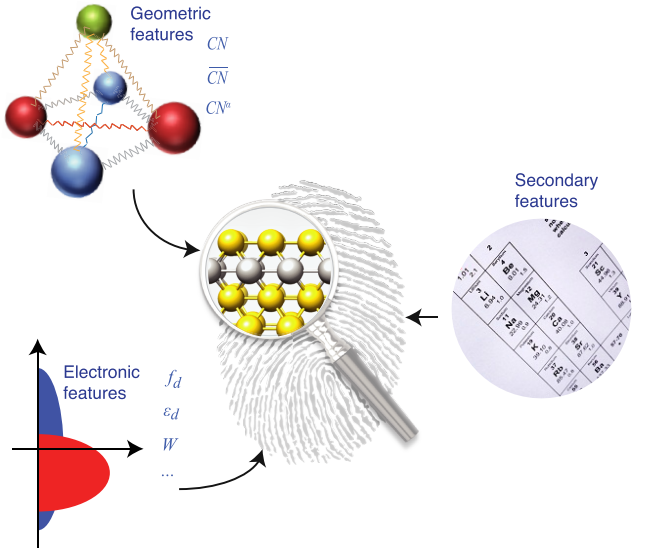


Figure 2. Physics-inspired feature engineering for predicting the catalytic properties of solid surfaces. Reproduced with permission from ref 13. Copyright 2017 Elsevier.

In the early days of developing ML models for the design of heterogeneous catalytic materials, one feature set that was employed is the *d*-band moments of transition and noble metal sites.[18] These moments represent statistical parameters characterizing the filling, center, width, and shape of the *d*-states distribution, and have been shown to be important factors for understanding reactivity trends of *d*-metal surfaces upon perturbation, such as by alloying, applying lattice strains, or adding promoters or poisons. Additionally, one important knowledge in chemisorption at metal surfaces is that the adsorbate−substrate bond length is largely governed by the delocalized, *sp*-electron density of local metal sites.[19] To quantify the driving force of charge transfer, if any in metal alloys, the local electronegativity of an adsorption site has been proposed as an essential feature.[19] For adsorbates with almost fully filled valence states, their binding on *d*-metals with fewer *d*-holes is not solely governed by the site *d*-states distribution.[20] In such instances, the Pauli repulsion energy contribution, which characterizes the energy penalty incurred to orthogonalize the interacting orbitals, can become dominant. The interatomic coupling matrix element squared, modulated by the *sp*-electron density, emerges as a key factor to capture these interactions. The coordination number of an adsorption site and its variants have been proposed as another type of easily accessible features. It is in essence a bond counting scheme with a varying degree of complexities for understanding reactivity trends of metal sites,[21−23] such as regular *CN*, generalized $\overline{CN}$, and orbitalwise CN^{α}. The regular *CN* refers to the count of first nearest neighboring atoms directly coordinated with the atoms of interest at a site. Widely utilized in the bond-order conservation principle in surface chemistry, this concept was later used to correlate with the surface reactivity of different metal facets.[21] The generalized $\overline{CN}$ was conceived to capture subtle variations in local environments, acknowledging that not all first nearest neighbors contribute equally to coordination.[22] To transcend the simplicity of this bond-counting scheme, the concept of orbitalwise CN^{α} was introduced that considers the interatomic interactions within a cutoff radius and can be naturally applied to alloy systems.[23] Contrasted with regular and generalized counterparts, orbitalwise CN^{α} was found to be crucial in defining the strength of the metal−adsorbate bond. Beyond primary electronic and geometric features, secondary features such as ionization energy, electron affinity, work function, atomic radius, electronegativity, and *d*-orbital radius of constituent metal atoms are also extensively employed to describe the reactivity properties of solid surfaces. These features represent the inherent properties of the metal atoms at an adsorption site or site motifs. The selection of physical features to depict the electronic and geometric representations

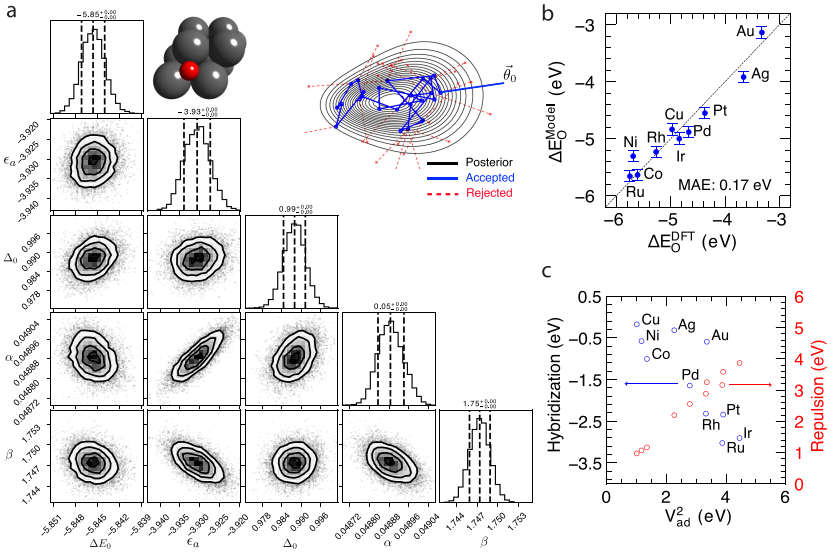


Figure 3. (a) Bayesian parametrization of *O adsorption on M(111) with posterior distributions. The inset shows the Markov chain Monte Carlo (MCMC) algorithm in parameter optimization.[24] (b) DFT-calculated *O adsorption energies (atomic O as the reference) at transition and noble metal surfaces vs model prediction using the posterior distribution of model parameters. Error bars represent the standard deviation of model predictions with 1000 random draws from converged trajectories. (c) Energy decomposition of *O adsorption on *d*-metal surfaces including hybridization and repulsion contributions. Reproduced with permission from ref 11, licensed under a Creative Commons License CC BY 4.0.

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

of surface sites is pivotal in the development of ML models and should be undertaken with due diligence.

2.2. Bayesian Statistical Learning

Bayeschem represents an interpretable ML approach, primarily developed to unravel the reactivity origin of active metal sites and to facilitate the design of catalytic materials guided by physical insights.[11] Built upon the principles of Bayesian inference, it utilizes physical models parametrized by learning from small data sets, thereby allowing for the integration of prior knowledge into the process through Bayes' rule. As a global optimization algorithm, Bayesian statistical learning explores relationships between variables (such as material composition, structure, and their high-level derivatives like *d*-band electronic factors) and target properties in a mathematically explicit manner. The *Bayeschem* model is conceived with our pre-existing understanding of these relationships as it pertains to surface reactivity knowledge from the domain. The Markov chain Monte Carlo (MCMC) sampling with the Metropolis−Hastings algorithm was used to approximate the posterior distribution of parameters by exploring the parameter space in a way that only requires the knowledge about their relative posterior densities. It avoids direct computation of complex normalizing constants by defining transition probabilities based on the likelihood functions. After sufficient samplings, the initial set of results is discarded and the remaining samples are thinned to extract converged values from the posterior distributions. Once optimized with a given data set, the Bayesian models can predict reactivity properties of unseen materials by leveraging their known attributes and estimating the probability distribution of the output target. It should be noted that this distribution represents the prediction uncertainty rather than the stochastic nature of the physical quantities and can be utilized to guide the sampling of the design space. A key advantage of the Bayesian approach lies in its capability to incorporate prior knowledge from diverse sources, such as experimental data, theoretical calculations, and expert knowledge, into the learning process. This enables *Bayeschem* to make more interpretable and comparably accurate predictions with fewer data points than conventional regression ML approaches. For instance, as depicted in Figure 3a, the posterior parameter distribution of *O chemisorption within the *d*-band theoretical framework can be learned from a set of *ab initio* data, accurately capturing the adsorption energies across transition and noble metal surfaces as shown in Figure 3b and insightful revelations about underlying contributions to their adsorption energies. As illustrated in Figure 3c, the hybridization energy contribution is linearly correlated with interatomic coupling matrix elements for late transition metals, while Pauli repulsion is of importance for noble metals with filled *d*-states (near zero hybridization energy). These capabilities make *Bayeschem* a potent tool for accelerating the discovery of novel materials with machine-learned insights for a range of catalytic applications.

2.3. Theory-Infused Deep Neural Networks

The rapid discovery of structural motifs with kinetically favorable descriptor values is undoubtedly desirable, yet it remains a challenging undertaking due to the computational demands for solving the many-electron Schrödinger equation accurately. Machine learning (ML) presents an alternative route to predicting chemical reactivity by learning the correlated interactions of atoms, ions, or molecules with a substrate, given a sufficient amount of *ab initio* data. However, black-box ML algorithms often fall short in terms of generalizing well outside the labeled data despite promising performance on both training and test samples. In response to this limitation, the theory-infused neural network (TinNet)[12] has been designed to predict the adsorption energies of simple molecules or their fragments at solid surfaces. TinNet is a novel ML approach that seamlessly blends domain theory into deep neural networks,[25,26] as demonstrated in Figure 4a, for

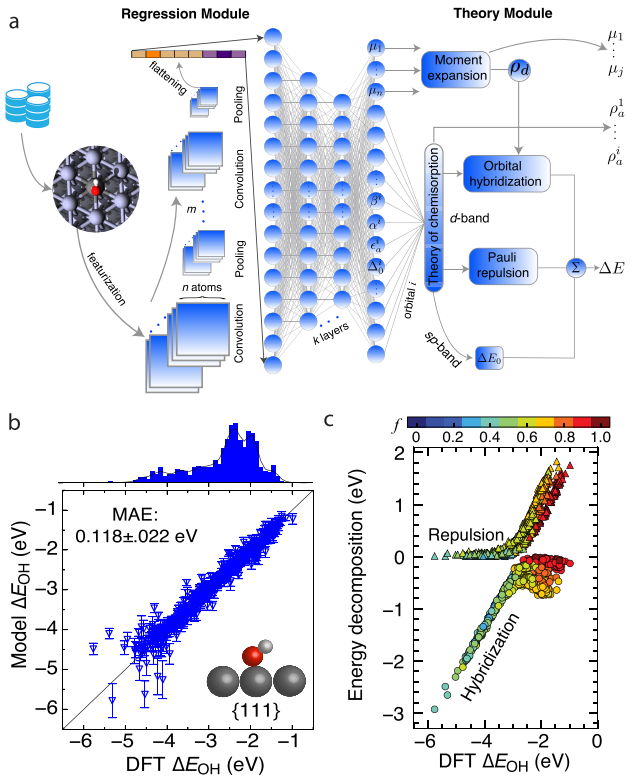


Figure 4. (a) A theory-infused neural network (TinNet) architecture in predicting surface reactivity at *d*-metal sites. (b) TinNet models of *OH adsorption on {111}-terminated metal and alloy surfaces with error bars from 10-fold cross-validation. (c) Orbital hybridization and Pauli repulsion contributions from the metal *d*-states to *OH adsorption energies on all 10-fold test sets deconvoluted by the *post hoc* attribution analysis of TinNet models. Reproduced with permission from ref 12, licensed under a Creative Commons License CC BY 4.0.

predicting physical properties of interest, such as surface reactivity. While built upon deep learning algorithms, specifically graph neural networks, TinNet respects physical principles from the domain in its architecture design, e.g., the well-established *d*-band theory of chemisorption[17] for reactivity prediction. This approach enables TinNet models to encode physical aspects of electronic interactions with a graph representation, successfully marrying the strengths of both theoretical and data-driven realms. By incorporating scientific knowledge into data-driven ML methods, TinNet achieves prediction performance on par with purely regression-based ML methods (as shown in Figure 4b for *OH adsorption on metal surfaces), especially for out-of-sample systems with unseen structural and electronic features. Importantly, TinNet overcomes the explainability limitations of black-box data-driven models and offers physical interpretations of the nature of chemical bonding, enabling the translation of machine-

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

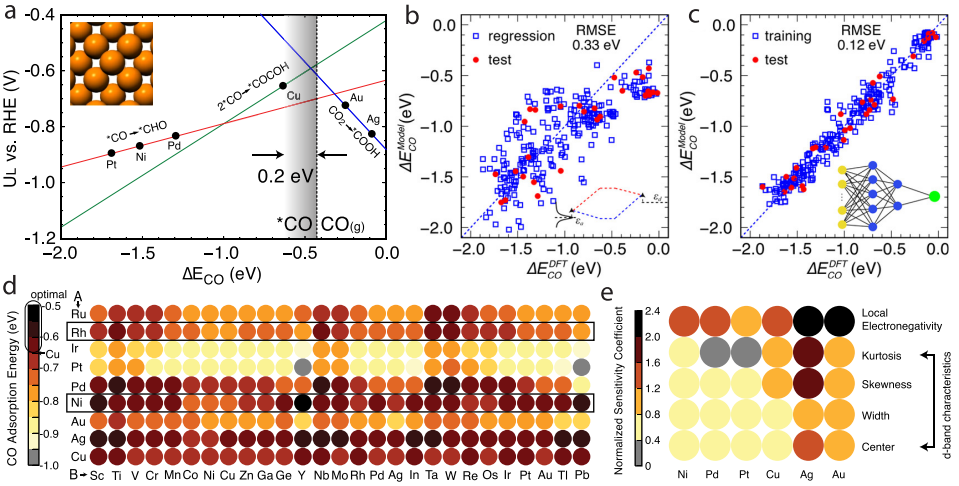


Figure 5. (a) Predicted limiting potentials for key elementary steps of the electroreduction of CO_2 to C_1 and C_2 species as a function of the CO adsorption energy. DFT-calculated CO adsorption energies on a set of idealized bimetallic surfaces versus prediction from (b) the two-level interaction model and (c) the ML model, where the insets are schematics of the methods. (d) Rational screening of CO adsorption energies on second-generation core−shell alloy surfaces (Cu_3B-A@Cu_{ML}) using the developed neural network model. (e) Normalized sensitivity coefficient for feature importance obtained by analyzing the network response to feature perturbations. Reproduced with permission from refs 13 and 18. Copyright 2017 Elsevier and Copyright 2015 American Chemistry Society, respectively.

learned insights into design strategies. As exemplified in Figure 4c, TinNet reveals that *d*-band hybridization governs the adsorption strength of *OH at early transition metal sites, whereas Pauli repulsion becomes more pronounced for late and noble metal surfaces. By integrating scientific knowledge into learning from data, TinNet enhances our fundamental understanding of the nature of chemical bonding and propels forward the field of data-driven catalytic materials design.

3. ACCELERATING CATALYTIC MATERIALS DESIGN

In the realm of catalysis, the linear adsorption-energy scaling and Brønsted−Evans−Polanyi (BEP) relationships have been widely utilized to estimate energetics of elementary reaction steps,[27,28] reducing high-dimensional activity/selectivity maps to only one or two catalytic descriptors, such as adsorption energies of key species at site motifs. These linear correlations emerge from the fundamental interactions of active sites with adsorbed species, including reaction intermediates and transition states. The descriptor-based design approach has become a common practice in heterogeneous catalysis pertinent to sustainable carbon, nitrogen, and oxygen cycles. In the following discussion, we will explore how interpretable ML expedites this design workflow and ushered in a data science paradigm of catalytic materials design.

3.1. Carbon Cycle

3.1.1. Carbon Dioxide Reduction. The electrochemical reduction of carbon dioxide (CO_2) on metal electrodes using renewable electricity has emerged as a viable approach to mitigate greenhouse emissions and propel a circular economy.[29] Notably, copper (Cu) demonstrates high activity and selectivity for converting CO_2 into hydrocarbons and oxygenates, a characteristic attributed to its near optimal surface reactivity in accordance with the Sabatier principle.[30] Theoretical studies[18,31] focusing on Cu(100)-mediated CO_2 electroreduction have suggested the governing role of *CO dimerization in defining onset potentials in C_2 pathways, while a concerted proton−electron transfer to *CO being critical in C_1 pathways (Figure 5a). Inspired by these mechanistic insights, we deliberately selected the *CO adsorption energy as a reactivity descriptor capable of characterizing alloy electrocatalysts for the selective conversion of CO_2 to C_2 species, e.g., ethylene (C_2H_4). To this end, quantum-chemical DFT calculations have been used to study free energy landscapes of relevant steps in C_1 and C_2 pathways on metal (100) surfaces under realistic electrochemical CO_2 reduction conditions, which typically exhibit a high coverage of *CO.[18] Given the computational limitations of exploring a wide chemical space for alloy catalyst design using DFT, a chemisorption model augmented by a shallow artificial neural network (ANN) consisting of only two hidden layers has been established. The approach leverages a small data set of *ab initio* adsorption energies (~250) and electronic fingerprints of idealized bimetallic surfaces to swiftly and accurately predict the surface reactivity of metal sites in an alloying environment. The ML model has demonstrated superior capability in predicting *CO adsorption energies on multimetallic surfaces compared with traditional two-level interaction models (Figure 5b and c), and facilitates high-throughput catalyst screening.[13,18] For a specific type of alloy nanostructure (Cu_3B-A@Cu_{ML}), the model highlighted several {100}-terminated multimetallic alloys that showed promise for efficient and selective CO_2 electrochemical reduction to C_2 species (Figure 5d). With sensitivity analysis methods, we revealed important roles of local electronegativity and higher-order moments in CO chemisorption, particularly on coinage metals (Figure 5e).

3.1.2. Methanol Oxidation. The electrochemical oxidation of methanol is an integral process in direct methanol fuel cells, a sustainable alternative to traditional energy conversion devices such as batteries and H_2 fuel cells. Commercialization of such fuel cells is hampered by the high overpotential of methanol oxidation at the anode, despite using state-of-the-art catalysts. Two reaction pathways,[32] the direct mechanism where methanol is oxidized entirely to carbon dioxide without involving the *CO intermediate, and the indirect mechanism proceeding through the *CO intermediate, were identified in determining the activity trends of metal surfaces. The direct mechanism experiences low activity due to *CO poisoning,

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

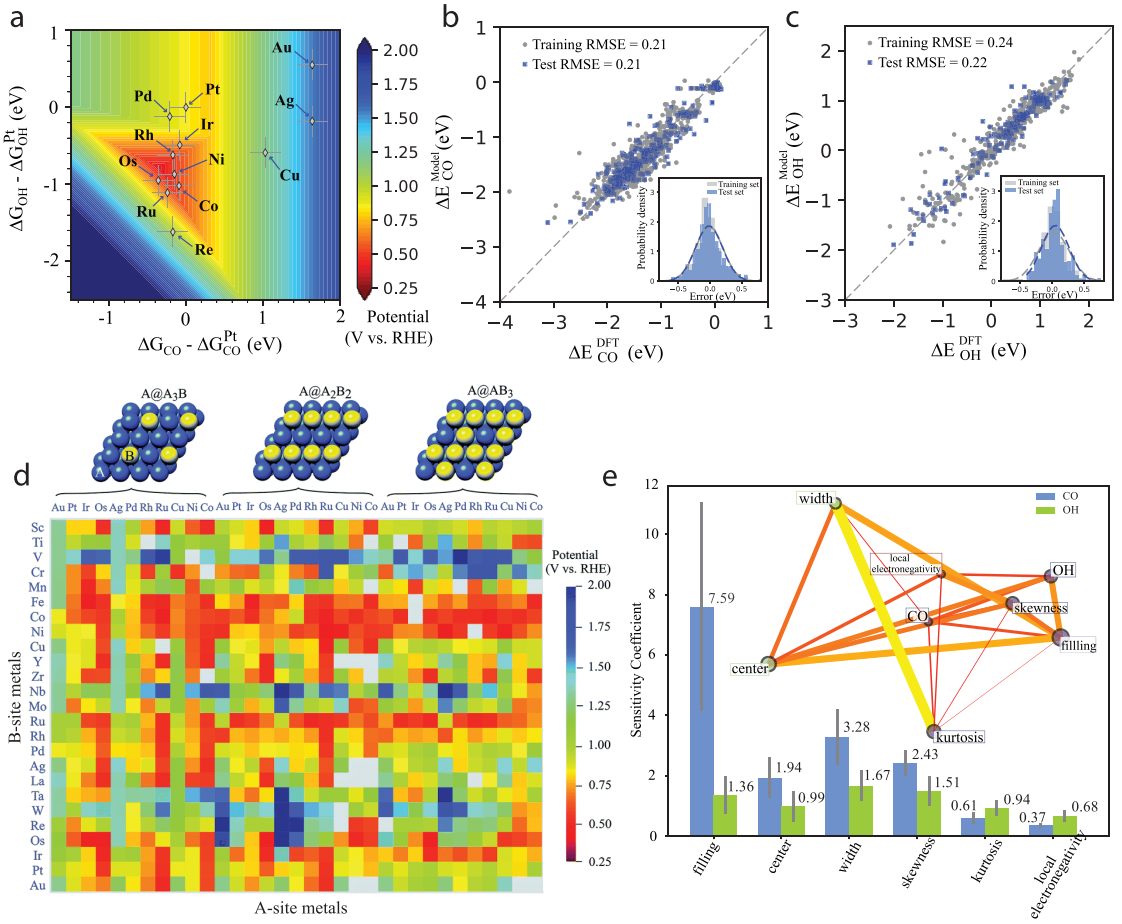


Figure 6. (a) Theoretical limiting potential (V vs RHE) of methanol electrooxidation as a function of *CO and *OH free formation energies relative to Pt(111). DFT-calculated (b) *CO and (c) *OH adsorption energies on a set of idealized bimetallic surfaces vs prediction from ML models. Insets show probability density of prediction errors for the training and test of the neural network models. (d) Model-predicted theoretical limiting potentials of near-surface alloy nanostructures toward methanol electrooxidation based on ML-predicted *CO and *OH free formation energies. The geometric structure of a subset of bimetallic catalysts is shown on the top with varying ratios of A and B on the top layer. (e) Feature importance scores for *CO and *OH models based on the sensitivity analysis of neural network models. The inset shows the clustering and linear dependence of input features and target properties. Reproduced with permission from ref 33. Copyright 2017 Royal Society of Chemistry.

particularly on platinum (Pt) catalysts, limiting the number of active sites available for catalytic turnovers. Conversely, the indirect mechanism embraces improved activity on metal surfaces if *CO removal is not a limiting factor. By leveraging linear scaling relations of *CO and *OH adsorption energies with relevant C- and O-bonded intermediates, a theoretical limiting potential map was developed (Figure 6a) with the top region of the activity volcano mostly occupied by precious metals, e.g., Ru, Rh, and Ir. To find electrocatalysts with earth-abundant metals, ML techniques have been adopted for high-throughput screening[33] (Figure 6b,c). ML models based on artificial neural networks, trained on a relatively small data set of *ab initio* adsorption energies on metal alloy surfaces (~1000 site samples), accurately predict *CO and *OH adsorption energies on novel bimetallic alloy nanostructures (Figure 6d), allowing rapid screening of a broad chemical space and suggesting catalysts with reduced overpotentials, such as Ru@Ru_3Pt and Ru@Ru_2Pt_2 surface alloys. These precious metal alloys exhibit enhanced methanol electrooxidation activity with relatively small theoretical limiting potentials, supporting experimental evidence that Pt/Ru alloys boost the electrocatalytic activity of methanol oxidation by aiding in water dissociation and facilitating the oxidation of *CO species.[34]

Moreover, earth-abundant metal alloys with Fe, Co, and Ni 3*d* metals were shown to be promising from the activity map. Hence, the incorporation of ML algorithms into the descriptor-based design approach potentially fast-tracks the discovery of efficient methanol oxidation catalysts by capturing complex, nonlinear adsorbate−substrate interactions. Sensitivity analysis (Figure 6e) further demonstrated that *CO and *OH adsorption energies are governed by distinct sets of features (center/width for *CO, local electronegativity for *OH), suggesting different bonding mechanisms of the two species on metal surfaces.

3.2. Nitrogen Cycle

3.2.1. Nitrate Reduction. Electrochemical nitrate (NO_3^-) reduction, a vital process in mitigating nitrate pollution and restoring the global nitrogen cycle, has recently received lots of interest, particularly on the development of efficient electrocatalysts toward ammonia (NH_3) production.[35−38] Although copper (Cu) shows promise as a catalyst for NO_3^- reduction in alkaline media, it requires high overpotentials for significant current densities. Metal alloy catalysts, e.g., CuNi,[39] often exhibit a performance trade-off, arguably due to the linear energy-scaling relations. Ordered intermetallics possess well-defined structures and compositions, offering excellent electro-

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

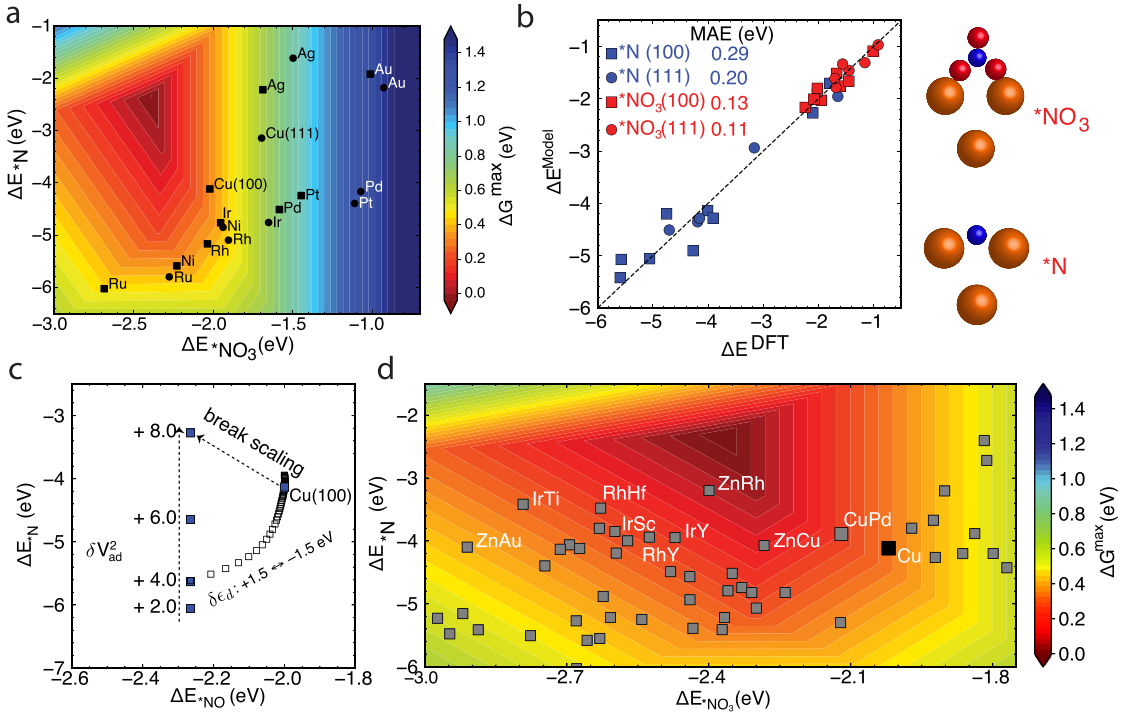


Figure 7. (a) Activity volcano plot of NO_3RR to NH_3 using adsorption energies of the $*NO_3$ and *N as reactivity descriptors. Circle and square symbols represent (111) and (100) metal surfaces, respectively. (b) Bayesian models of chemisorption (*Bayeschem*) for $*NO_3$ and *N on (100)- and (111)-terminated metal surfaces. (c) *Bayeschem*-predicted adsorption energies of $*NO_3$ and *N on Cu(100) upon perturbation of the electronic features of adsorption sites, illustrating the beyond-scaling behavior due to site-specific Pauli repulsion for *N adsorption. (d) DFT-calculated adsorption energies of $*NO_3$ and *N on (100)-terminated B2 intermetallics close to the activity volcano top. Cu(100) and a few interesting systems, e.g., CuPd, are highlighted. Reproduced with permission from ref 35, licensed under a Creative Commons License CC BY 4.0.

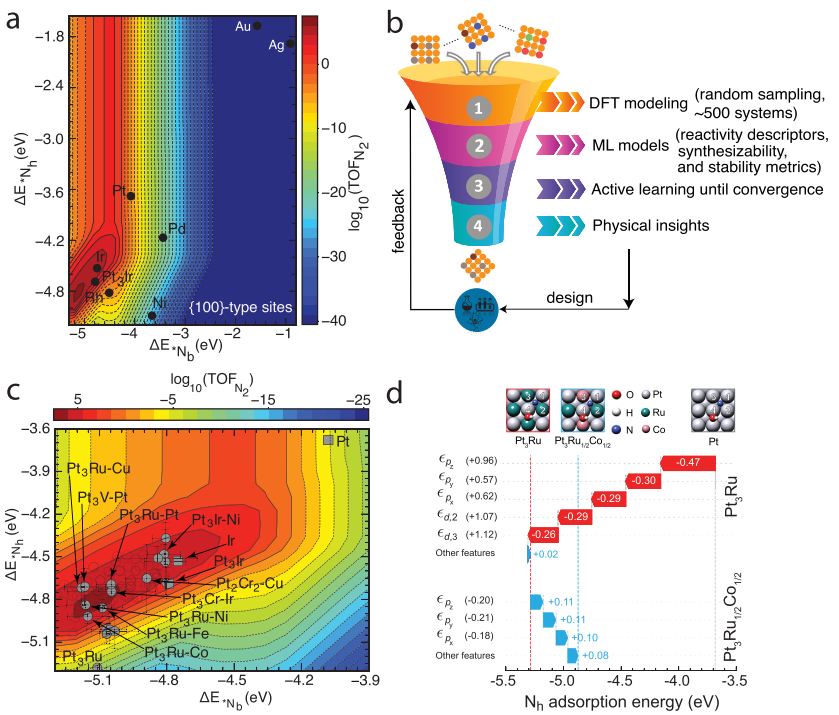


Figure 8. (a) Activity volcano plot of the NH_3 electrooxidation as a function of the bridge and hollow *N adsorption energies on {100}-type site motifs. A few transition and noble metals along with Pt_3Ir are marked on the map. (b) Active learning workflow for accelerating catalytic materials discovery. (c) NH_3 electrooxidation activity map at 0.3 V vs RHE with solid markers showing promising ternary Pt alloy electrocatalysts predicted from the workflow. (d) Local interpretation of $*N_h$ adsorption on $Pt_3Ru(100)$ relative to Pt(100) and $Pt_3Ru_{1/2}Co_{1/2}(100)$ to $Pt_3Ru(100)$. Reproduced with permission from ref 49, licensed under a Creative Commons License CC BY 4.0.

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

catalytic stability and promising activity. Grand-canonical DFT calculations (GC-DFT) were employed to gain mechanistic insights into NO_3^- reduction on metal surfaces in alkaline media using *N and $*NO_3$ adsorption energies as two reactivity descriptors (Figure 7a). This analysis, coupled with the Bayesian models of chemisorption (*Bayeschem*[11]) optimized with adsorption properties of *N and $*NO_3$ on a handful of pristine metal surfaces (Figure 7b), identified a mechanism to break the linear adsorption-energy scaling by leveraging site-specific Pauli repulsion interactions. The findings highlight the unique properties of {100}-oriented surface sites in ordered intermetallic B2 structures, where subsurface metal *d*-orbitals overlap significantly with the hollow *N *p*-orbitals, while bridge-bidentate $*NO_3$ remains largely unaffected (Figure 7c). This mechanism was confirmed by DFT calculations showing that a range of B2 intermetallics with {100}-termination deviate from linear energy scaling relations (Figure 7d). This prompted a successful synthesis of B2 intermetallic CuPd nanocubes, which exhibit improved catalytic performance for NO_3^- reduction to NH_3.[35] This study offers valuable mechanistic insights into NO_3^- reduction to NH_3 on ordered intermetallic catalysts for overcoming adsorption-energy scaling limitations. It underscores how combining quantum chemistry and ML approaches can provide a theory-guided, data-driven catalytic materials design strategy toward sustainability.

3.2.2. Ammonia Oxidation. Electrochemical ammonia oxidation is a crucial reaction in the field of electrocatalysis, as it plays a significant role in the development of sustainable energy technologies.[40,41] The conversion of ammonia (NH_3) to dinitrogen (N_2) through electrochemical oxidation offers a promising pathway for energy generation and storage, particularly with the booming clean hydrogen economy. However, the high overpotentials and the use of expensive catalysts, such as platinum (Pt) and its bimetallics with iridium (Ir), have hindered the widespread implementation of this technology. In recent years, extensive research has been conducted to understand the reaction pathways involved in NH_3 oxidation[42,43] and to design efficient catalysts that can overcome the limitations of Pt−Ir systems.[44,45] Quantum-chemical calculations and experimental studies on Pt single-crystal electrodes have revealed that the reaction is structure-sensitive, with specific site motifs such as {100}-type sites being predominantly active.[46] The reaction likely proceeds through a mechanism[43,47] involving the dehydrogenation of NH_3 to $*NH_x$ (x = 1 or 2), followed by their dimerization to $*N_2H_y$ and further dehydrogenation to N_2. Complete dehydrogenation of NH_3 to *N can possibly lead to surface deactivation. However, the nature of poisoning species is under debate. A traditional view is that the *N species binds too strongly to the hollow sites. However, it is also likely due to the strongly bonded *NO formed by *N with surface *OH at high operating potentials, as evidenced by *in situ* IR spectroscopy.[48] We employed GC-DFT calculations to obtain scaling relations for reaction energetics that were then used in a microkinetic model while considering the surface coverage effects of the key $*NH_2$ species. The approach was demonstrated using two reactivity descriptors, *N at the bridge site and *N at the hollow site of {100}-terminated alloy surfaces (Figure 8a). To go beyond the Pt−Ir catalysts, we explored the design space of ternary Pt alloy nanostructures using a series of ML techniques. TinNet has been applied to predict the reactivity descriptor values of metal sites in ternary alloys for NH_3 electrooxidation.[49] The TinNet models of *N adsorption energies were iteratively developed using an initial data set of ∼800 {100}-terminated bimetallic alloy surfaces and ∼200 {100}-terminated Pt-based ternary alloy surfaces in an active learning scheme. The 10-fold cross-validated mean absolute errors (MAEs) for *N adsorption energies are within 0.1 eV. The inclusion of adsorbate−adsorbate interactions, e.g., *OH, in developing ML models was showcased in this data set. Ideally, the adsorbate−adsorbate interactions can be taken into account in a more generic fashion by developing training data with lateral interactions beyond *OH, for example, by using the formation energy of a given adsorbate configuration as the target property.[50] Figure 8b shows an ML workflow by predicting not only the surface reactivity but also the stability and synthesizability metrics in the screening of desired alloy compositions. Several systems of Pt_3Ru with 3*d* metals (e.g., Fe, Co, and Ni) partially substituting Ru show promising activity, while being Ir-free (Figure 8c). Our experiments have validated the predicted materials with improved activity compared to pure Pt, Pt_3Ru, and Pt_3Ir. The TinNet implementation allocates hidden neurons for the first and second moments of the *d*-states distribution of site atoms for an *n*-atom site ensemble. The *d*-states distribution of the adsorption site is then represented by a superposition of individual *d*-dos constructs, such as semielliptic functions. Other neurons representing interaction parameters of the adsorbate frontier orbitals with the metal *sp*- and *d*-states have the same dimensions and physical meanings for adsorption sites of different atom ensembles. The domain knowledge that can be gained from data comes naturally with the TinNet algorithm, suggesting an important role of adsorbate resonance energies that are influenced by metal *sp*-states upon site perturbation (Figure 8d). This insight cannot be achieved with purely data-driven ML methods. The study highlights the importance of the frontier molecular orbital theory, electronic structure methods, and deep learning algorithms in developing interpretable ML models of chemical bonding. Infusing theory into ML fueled by *ab initio* adsorption properties will eventually lead us to better understand the fundamentals of linear energy relationships and devise strategies to overcome such constraints in catalysis.

3.3. Oxygen Cycle

3.3.1. Oxygen Reduction. The oxygen reduction reaction (ORR) is a crucial process in fuel cell technologies, in which O_2 is reduced to water at the cathode. However, the sluggish kinetics of ORR on traditional catalysts, such as platinum (Pt) nanoparticles, limits the efficiency and economic viability of fuel cells. This has led to extensive research efforts to develop alternative catalysts with enhanced ORR activity and reduced cost. Quantum chemistry has played a significant role in understanding the reaction mechanisms of ORR, particularly on transition and noble metals.[1] One commonly observed mechanism of O_2 reduction is through the associative pathway, where the formation of *OOH from O_2 activation and the removal of *OH to release active sites are potentially rate-limiting steps. Understanding these mechanisms is crucial for designing catalysts with improved ORR activity. Pt monolayer alloy catalysts have emerged as promising alternatives to pure Pt. By combining Pt with other transition metals in a Pt_{ML} nanostructure, Pt surface sites can exhibit enhanced catalytic activity and stability. The use of predictive models or ML algorithms has significantly accelerated the exploration of this

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

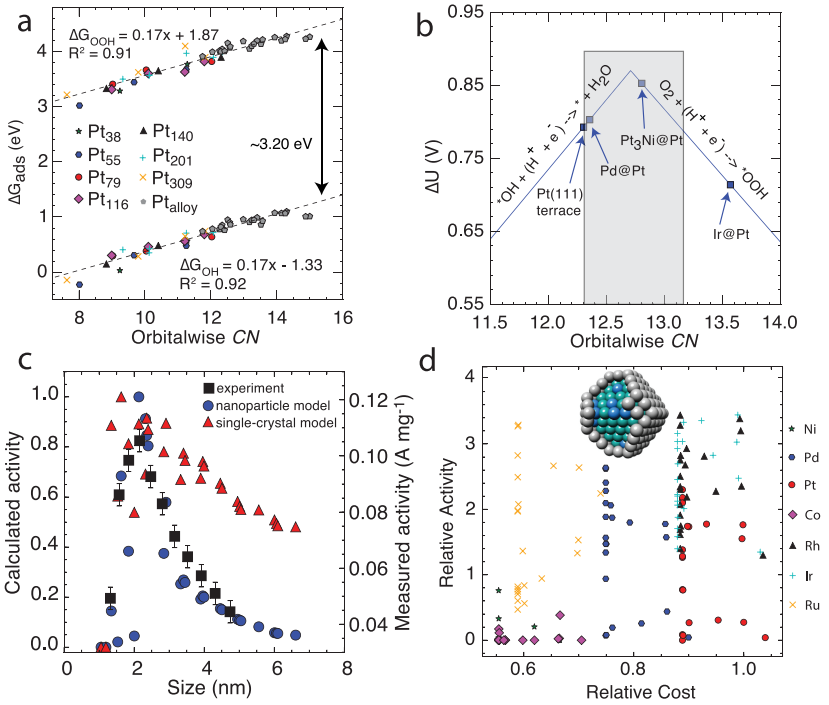


Figure 9. (a) *OH and *OOH free formation energies and (b) theoretical ORR limiting potential at surface sites of Pt nanoparticles described by the s-orbital based coordination number CN^s. (c) Model predicted vs measured mass activity of Pt nanoparticles for ORR. (d) Rapid screening of Pt monolayer core−shell nanoalloys A_3B@Pt_{ML} using the orbitalwise CN^s as a descriptor. A 314-atom nanoparticle with a 3:1 A/B ratio at the core was used (see the inset structural model). Reproduced with permission from ref 53. Copyright 2018 American Chemistry Society.

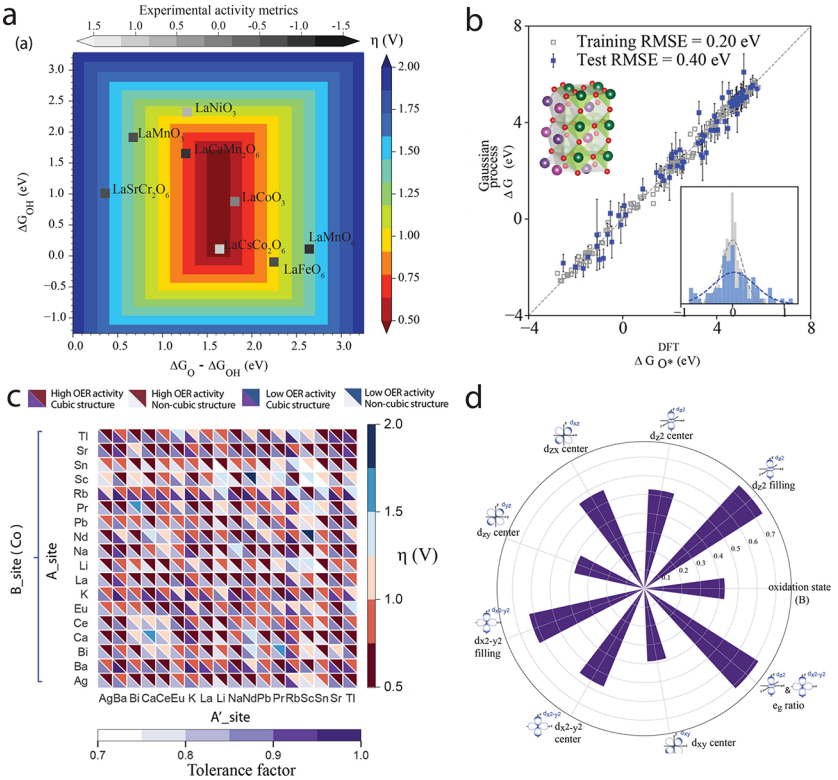


Figure 10. (a) Heat map of theoretical overpotential (V vs RHE) of oxygen evolution using the *OH and *O−*OH free formation energies as activity descriptors. Experimental measured activity relative to $LaCoO_3$ was color coded.[58] (b) Parity plot for DFT-calculated vs Gaussian process regression model prediction of adsorption free energies (*O taken as an example) on perovskite surfaces. (c) Heat map visualization of the OER activity and stability of double perovskites as a function of the mixed A-site metals with Co at the B-site in terms of the color-coded OER overpotential and tolerance factor. (d) Polar distribution plot of feature importance by *post hoc* attribution analysis. Reproduced with permission from ref 58. Copyright 2020 American Chemistry Society.

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

immense design space for new catalyst compositions. One important concept in the design of metal-based nanocatalysts is the coordination number (CN) of surface sites.[21−23] The regular CN and its generalized variant[22] represent a physically intuitive descriptor that quantifies the number of neighboring atoms around a specific site. The orbitalwise CN^{α} has shown improved linear correlations with surface *OH and *OOH binding energies on Pt nanoparticles and alloy surfaces (Figure 9a). By using the s-orbital based CN^{s}, an activity volcano of ORR on metal-based catalysts was developed (Figure 9b), capturing activity trends of known materials, e.g., Pt_3Ni@Pt.[51] By considering the heterogeneity of surface sites, the orbitalwise CN^{s} descriptor captured experimentally observed size effects of the ORR activity of Pt nanoparticles, outperforming the traditional single-crystal model (Figure 9c). The CN^{s} model has been used to rapidly screen through the A_3B@Pt_{ML} alloy nanoparticles of a broad range of A and B elements, and identifies improved catalysts with reduced cost (Figure 9d). These basic understanding of structure−reactivity relationships can be further integrated with deep learning algorithms[52] to significantly accelerate the development of more efficient, highly complex, and cost-effective fuel cell electrocatalysts.

3.3.2. Water Oxidation. Water oxidation is a bottleneck reaction in various energy conversion processes, such as water splitting for hydrogen production in electrolyzers and H_2O_2 production in electrochemical devices.[54−56] The high overpotential and sluggish kinetics of water oxidation have hindered the development of those technologies. To address this challenge, researchers have extensively studied the activity trends of metal oxide catalysts toward water oxidation.[57] The theoretical limiting potential has emerged as a powerful metric to understand the activity trends of catalysts. The volcano plot maps the limiting potential of water oxidation onto reactivity descriptors, such as the free formation energy of oxygen intermediates, i.e., *O, *OH, or their combinations (Figure 10a). It has been observed that the top region of the active volcano is occupied by scarce and expensive noble metal catalysts, such as RuO_2 and IrO_2, consistent with experimental observations. Perovskites with an ABO_3-type structural configuration have received great attention attributed to their superior catalytic properties, with a few La-based perovskites showing promising activity from both theory and experiment (Figure 10a). To design perovskite catalysts with improved performance, ML models using Gaussian Processes (GP) have been employed (Figure 10b) for materials screening. These models utilize a set of B-terminated ABO_3 surfaces to unravel underlying structure−reactivity relationships and predict the adsorption energies of oxygen intermediates on diverse surfaces. We developed the electronic descriptors based on the metal *d*-states distribution, e.g., band moments of the electronic density of states. By training the GP model with a small data set of ∼350 samples in an adaptive learning fashion, it becomes possible to rapidly screen and identify surfaces that can synergistically activate water oxidation without being poisoned by surface intermediates. By combining the insights from mechanistic studies with the predictive power of ML models, we can efficiently explore the materials subspace ($ACoO_3$ with mixed A metals) and identify promising perovskites for water oxidation (Figure 10c). With Kullback−Leibler divergence originated from the context of information theory, important physical insights have been drawn showing the critical role of the e_g *d*-orbitals (occupancy ratio) of the B-site metal,[55] in governing the interaction of surface intermediates with perovskite surfaces (Figure 10d).

4. CONCLUSIONS AND PERSPECTIVE

The framework in this Account represents one of the collective efforts within the scientific community in harnessing the power of data and domain knowledge for materials discovery. Equipping the descriptor-based design strategy with interpretable ML algorithms has dramatically accelerated the screening of heterogeneous catalysts across an immense chemical space. By engineering the numerical representation of active sites using physically transparent features such as the distribution statistics of valence electronic states, coordination numbers, intrinsic properties of constituent elements, and high-level feature representations extracted by convolution operations from chemical graphs of atoms and bonds, ML models trained with *ab initio* adsorption data can capture complex, nonlinear adsorbate−substrate interactions with high accuracy, while being interpretable for physicochemical insights. This approach has shown great promise in materials design for catalyzing sustainable carbon, nitrogen, and oxygen cycles with renewable electricity as the energy input. By virtue of the interpretability of the above-mentioned ML models irrespective of their underlying complexities, *post hoc* attribution analysis drastically advances our fundamental knowledge of chemical bonding at solid surfaces beyond the conventional wisdom.

Interpretable ML has emerged as a promising approach to address the lack of transparency and the explainability of deep learning algorithms. However, there are several challenges that need to be tackled to realize its full potential in catalytic materials design. One of the main challenges is the development of new strategies to effectively incorporate scientific knowledge and domain expertise into the learning process, thus breaking the accuracy-interpretability trade-off.[16] The theory-infused neural network (TinNet) exhibits emergent behaviors in interpretability, whereas it is fundamentally limited in generalizability of the underlying neural network architectures, especially with out-of-distribution samples. Imposing constraints or scientific rules through other means, e.g., physics-inspired graph kernels,[15] symbolic regression with logical reasoning,[59] might provide improved accuracy while attaining interpretability.

Another challenge is on the theory side. While metals and their alloys can be well described in DFT with GGA-type functionals and their surface properties are relatively well understood, metal compounds such as metal oxides are not described at a similar level of maturity partly because of strongly correlated electrons and their complex electronic structure. It requires a better understanding of the underlying physical interactions that govern the behavior of complex material systems. The emergence of comprehensive databases, e.g., OC22,[60] holds significant promise for tackling this challenge. Given vast amounts of data, we can pretrain ML models, which essentially means we can have a foundational understanding of the system. This foundational knowledge can be incredibly broad and generic, capturing the basic trends and relationships in the data. Then, transfer learning, a technique where a pretrained model is further trained on a new data set, can be utilized. The beauty of this approach is that it leverages the vast amount of data available in databases to capture broad trends and then uses a limited number of new calculations to fine-tune the model for specific scenarios. In this regard, it is crucial to develop an accurate set of data beyond GGA-levels

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

that can be used for benchmarking the ML framework and, more importantly, advancing our theoretical knowledge of molecule−substrate interactions at highly complex solid materials.

Finally, it is of importance to further develop efficient and reliable interpretation methods. Many currently employed interpretation techniques assume feature independence, which is not necessarily valid for physical systems. Attribution analysis of specific features might create hypothetical systems that are not accessible in real-world scenarios. Undeniably, the future of interpretable ML is shining bright, and it has the potential, if continually progressed, to revolutionize scientific materials research toward sustainability in the years to come.

AUTHOR INFORMATION

Corresponding Author

Hongliang Xin − *Department of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-9344-1697; Email: hxin@vt.edu

Authors

Tianyou Mou − *Department of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0002-7389-6712

Hemanth Somarajan Pillai − *Department of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0003-3131-7396

Shih-Han Wang − *Department of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0003-4418-2080

Yang Huang − *Department of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/accountsmr.3c00131

Notes

The authors declare no competing financial interest.

Biographies

Hongliang Xin received his B.S. degree from Tianjin University in 2002, M.S. degree from Tsinghua University in 2005, and Ph.D. degree from the University of Michigan in 2011, all in Chemical Engineering. After completing postdoctoral research at Michigan and Stanford/SLAC, he joined the Department of Chemical Engineering at Virginia Tech in 2014 and was promoted to Associate Professor in 2020. His research group focuses on the development of an interpretable machine learning framework for advancing theory of heterogeneous catalysis and catalytic materials discovery.

Tianyou Mou received his B.S. degree from Beijing University of Chemical Technology and Rutgers University in 2018, and Ph.D. degree from Virginia Tech in 2023, both in Chemical Engineering. He is currently a postdoc at the Brookhaven National Laboratory. His research interest is to understand interfacial kinetics of heterogeneous catalytic reactions using multiscale simulations and accelerate catalytic materials discovery with machine learning.

Hemanth Somarajan Pillai received his B.S. degree from Syracuse university in 2017 and PhD from Virginia Tech in 2022, both in Chemical Engineering. He is currently a postdoc at the Fritz Haber Institute working on coupling kinetics and mass transfer models to provide insights into the relationship between the morphology of an electrocatalyst and its catalytic performance.

Shih-Han Wang received his B.S. degree from National Taiwan University of Science and Technology in 2013, M.S. degree from National Taiwan University in 2015, and has been pursuing his Ph.D. degree at Virginia Tech since 2017, all in Chemical Engineering. His research focus is to develop theory-infused neural networks for predicting chemical reactivity and electronic properties of heterogeneous catalysts.

Yang Huang received his B.E. degree in Materials Physics from University of Science and Technology Beijing in 2015, M.S. degree in Materials Science and Engineering from University of California San Diego in 2017, and is currently pursuing his Ph.D. degree in Chemical Engineering at Virginia Tech. His research focus is understanding mechanisms of catalytic processes with quantum chemistry and integrating fundamental principles into deep learning for advancing catalysis theory and catalytic materials design.

ACKNOWLEDGMENTS

H.X., S.-H.W., and H.S.P. acknowledge funding support from the NSF CAREER program (CBET-1845531). H.X. and T.M. acknowledge funding support from the NSF Chemical Catalysis program (CHE-2102363). H.X. and Y.H. acknowledge the partial support from the NSF CBET Catalysis program (CBET-2245402) and the US Department of Energy, Office of Basic Energy Sciences under contract no. DE-SC0023323. T.M. and S.-H.W. thank the NSF Non-Academic Research Internships for Graduate Students (INTERN) program for supporting their internships in the Brookhaven National Lab (2022) and the National Institute of Standards and Technology (2023), respectively.

REFERENCES

(1) Kulkarni, A.; Siahrostami, S.; Patel, A.; Nørskov, J. K. Understanding Catalytic Activity Trends in the Oxygen Reduction Reaction. *Chem. Rev.* **2018**, *118* (5), 2302−2312.

(2) Shih, A. J.; Monteiro, M. C. O.; Dattila, F.; Pavesi, D.; Philips, M.; da Silva, A. H. M.; Vos, R. E.; Ojha, K.; Park, S.; van der Heijden, O.; Marcandalli, G.; Goyal, A.; Villalba, M.; Chen, X.; Gunasooriya, G. T. K. K.; McCrum, I.; Mom, R.; López, N.; Koper, M. T. M. Water Electrolysis. *Nature Reviews Methods Primers* **2022**, *2* (1), 1−19.

(3) Chen, B. W. J.; Xu, L.; Mavrikakis, M. Computational Methods in Heterogeneous Catalysis. *Chem. Rev.* **2021**, *121* (2), 1007−1048.

(4) Nørskov, J. K.; Abild-Pedersen, F.; Studt, F.; Bligaard, T. Density Functional Theory in Surface Chemistry and Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (3), 937−943.

(5) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. Density Functionals for Surface Science: Exchange-Correlation Model Development with Bayesian Error Estimation. *Phys. Rev. B Condens. Matter* **2012**, *85* (23), 235149.

(6) Mou, T.; Pillai, H. S.; Wang, S.; Wan, M.; Han, X.; Schweitzer, N. M.; Che, F.; Xin, H. Bridging the Complexity Gap in Computational Heterogeneous Catalysis with Machine Learning. *Nature Catalysis* **2023**, *6* (2), 122−136.

(7) Kitchin, J. R. Machine Learning in Catalysis. *Nature Catalysis* **2018**, *1* (4), 230−232.

(8) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable Machine Learning for Knowledge Generation in Heterogeneous Catalysis. *Nature Catalysis* **2022**, *5* (3), 175−184.

(9) Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. Towards the Computational Design of Solid Catalysts. *Nat. Chem.* **2009**, *1* (1), 37−46.
(10) Omidvar, N.; Pillai, H. S.; Wang, S.-H.; Mou, T.; Wang, S.; Athawale, A.; Achenie, L. E. K.; Xin, H. Interpretable Machine Learning of Chemical Bonding at Solid Surfaces. *J. Phys. Chem. Lett.* **2021**, *12* (46), 11476−11487.
(11) Wang, S.; Pillai, H. S.; Xin, H. Bayesian Learning of Chemisorption for Bridging the Complexity of Electronic Descriptors. *Nat. Commun.* **2020**, *11* (1), 6132.
(12) Wang, S.-H.; Pillai, H. S.; Wang, S.; Achenie, L. E. K.; Xin, H. Infusing Theory into Deep Learning for Interpretable Reactivity Prediction. *Nat. Commun.* **2021**, *12* (1), 5288.
(13) Li, Z.; Ma, X.; Xin, H. Feature Engineering of Machine-Learning Chemisorption Models for Catalyst Design. *Catal. Today* **2017**, *280*, 232−238.
(14) Andersen, M.; Reuter, K. Adsorption Enthalpies for Catalysis Modeling through Machine-Learned Descriptors. *Acc. Chem. Res.* **2021**, *54*, 2741.
(15) Xu, W.; Reuter, K.; Andersen, M. Predicting Binding Motifs of Complex Adsorbates Using Machine Learning with a Physics-Inspired Graph Representation. *Nature Computational Science* **2022**, *2* (7), 443−450.
(16) Omidvar, N.; Xin, H. Algorithm-Derived Feature Representations for Explainable AI in Catalysis. *Trends in Chemistry* **2021**, *3*, 990.
(17) Hammer, B.; Nørskov, J. K. Electronic Factors Determining the Reactivity of Metal Surfaces. *Surf. Sci.* **1995**, *343* (3), 211−220.
(18) Ma, X.; Li, Z.; Achenie, L. E. K.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO2 Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **2015**, *6* (18), 3528−3533.
(19) Xin, H.; Holewinski, A.; Linic, S. Predictive Structure-Reactivity Models for Rapid Screening of Pt-Based Multimetallic Electrocatalysts for the Oxygen Reduction Reaction. *ACS Catal.* **2012**, *2* (1), 12−16.
(20) Xin, H.; Linic, S. Communications: Exceptions to the *d*-Band Model of Chemisorption on Metal Surfaces: The Dominant Role of Repulsion between Adsorbate States and Metal D-States. *J. Chem. Phys.* **2010**, *132* (22), 221101.
(21) Lopez, N.; Janssens, T. V. W.; Clausen, B. S.; Xu, Y.; Mavrikakis, M.; Bligaard, T.; Nørskov, J. K. On the Origin of the Catalytic Activity of Gold Nanoparticles for Low-Temperature CO Oxidation. *J. Catal.* **2004**, *223* (1), 232−235.
(22) Calle-Vallejo, F.; Martínez, J. I.; García-Lastra, J. M.; Sautet, P.; Loffreda, D. Fast Prediction of Adsorption Properties for Platinum Nanocatalysts with Generalized Coordination Numbers. *Angew. Chem., Int. Ed.* **2014**, *53* (32), 8316−8319.
(23) Ma, X.; Xin, H. Orbitalwise Coordination Number for Predicting Adsorption Properties of Metal Nanocatalysts. *Phys. Rev. Lett.* **2017**, *118* (3), 036101.
(24) Gamerman, D.; Lopes, H. F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed..; Chapman and Hall/CRC, 2006.
(25) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120* (14), 145301.
(26) Back, S.; Yoon, J.; Tian, N.; Zhong, W.; Tran, K.; Ulissi, Z. W. Convolutional Neural Network of Atomic Surface Structures To Predict Binding Energies for High-Throughput Screening of Catalysts. *J. Phys. Chem. Lett.* **2019**, *10* (15), 4401−4408.
(27) Abild-Pedersen, F.; Greeley, J.; Studt, F.; Rossmeisl, J.; Munter, T. R.; Moses, P. G.; Skúlason, E.; Bligaard, T.; Nørskov, J. K. Scaling Properties of Adsorption Energies for Hydrogen-Containing Molecules on Transition-Metal Surfaces. *Phys. Rev. Lett.* **2007**, *99* (1), 016105.
(28) Bligaard, T.; Nørskov, J. K.; Dahl, S.; Matthiesen, J.; Christensen, C. H.; Sehested, J. The Brønsted-Evans-Polanyi Relation and the Volcano Curve in Heterogeneous Catalysis. *J. Catal.* **2004**, *224* (1), 206−217.
(29) Stephens, I. E. L.; Chan, K.; Bagger, A.; Boettcher, S. W.; Bonin, J.; Boutin, E.; Buckley, A.; Buonsanti, R.; Cave, E.; Chang, X.; Chee, S. W.; da Silva, A. H. M.; De Luna, P.; Einsle, O.; Endrődi, B.; Escribano, M. E.; de Araujo, J. V. F.; Figueiredo, M. C.; Hahn, C.; Hansen, K. U.; Haussener, S.; Hunegnaw, S.; Huo, Z.; Hwang, Y. J.; Janáky, C.; Jayathilake, B. S.; Jiao, F.; Jovanov, Z. P.; Karimi, P.; Koper, M. T. M.; Kuhl, K.; Lee, W. H.; Liang, Z.; Liu, X.; Ma, S.; Ma, M.; Oh, H.-S.; Robert, M.; Cuenya, B. R.; Rossmeisl, J.; Roy, C.; Ryan, M. P.; Sargent, E. H.; Sebastián-Pascual, P.; Seger, B.; Steier, L.; Strasser, P.; Varela, A. S.; Vos, R. E.; Wang, X.; Xu, B.; Yadegari, H.; Zhou, Y. 2022 Roadmap on Low Temperature Electrochemical CO2 Reduction. *J. Phys. Energy* **2022**, *4*, 042003.
(30) Peterson, A. A.; Abild-Pedersen, F.; Studt, F.; Rossmeisl, J.; Nørskov, J. K. How Copper Catalyzes the Electroreduction of Carbon Dioxide into Hydrocarbon Fuels. *Energy Environ. Sci.* **2010**, *3* (9), 1311−1315.
(31) Calle-Vallejo, F.; Koper, M. T. M. Theoretical Considerations on the Electroreduction of CO to C 2 Species on Cu(100) Electrodes. *Angew. Chem., Int. Ed.* **2013**, *52* (28), 7282−7285.
(32) Ferrin, P.; Nilekar, A. U.; Greeley, J.; Mavrikakis, M.; Rossmeisl, J. Reactivity Descriptors for Direct Methanol Fuel Cell Anode Catalysts. *Surf. Sci.* **2008**, *602* (21), 3424−3431.
(33) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-Throughput Screening of Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A Mater. Energy Sustain.* **2017**, *5* (46), 24131−24138.
(34) Dinh, H. N.; Ren, X.; Garzon, F. H.; Piotr Zelenay; Gottesfeld, S. Electrocatalysis in Direct Methanol Fuel Cells: In-Situ Probing of PtRu Anode Catalyst Surfaces. *J. Electroanal. Chem.* **2000**, *491* (1), 222−233.
(35) Gao, Q.; Pillai, H. S.; Huang, Y.; Liu, S.; Mu, Q.; Han, X.; Yan, Z.; Zhou, H.; He, Q.; Xin, H.; Zhu, H. Breaking Adsorption-Energy Scaling Limitations of Electrocatalytic Nitrate Reduction on Intermetallic CuPd Nanocubes by Machine-Learned Insights. *Nat. Commun.* **2022**, *13* (1), 2338.
(36) Liu, J.-X.; Richards, D.; Singh, N.; Goldsmith, B. R. Activity and Selectivity Trends in Electrocatalytic Nitrate Reduction on Transition Metals. *ACS Catal.* **2019**, *9* (8), 7052−7064.
(37) Gao, Q.; Yao, B.; Pillai, H. S.; Zang, W.; Han, X.; Liu, Y.; Yu, S.-W.; Yan, Z.; Min, B.; Zhang, S.; Zhou, H.; Ma, L.; Xin, H.; He, Q.; Zhu, H. Synthesis of Core/shell Nanocrystals with Ordered Intermetallic Single-Atom Alloy Layers for Nitrate Electroreduction to Ammonia. *Nature Synthesis* **2023**, *2*, 458.
(38) Han, S.; Li, H.; Li, T.; Chen, F.; Yang, R.; Yu, Y.; Zhang, B. Ultralow Overpotential Nitrate Reduction to Ammonia via a Three-Step Relay Mechanism. *Nature Catalysis* **2023**, *6*, 402.
(39) Wang, Y.; Xu, A.; Wang, Z.; Huang, L.; Li, J.; Li, F.; Wicks, J.; Luo, M.; Nam, D.-H.; Tan, C.-S.; Ding, Y.; Wu, J.; Lum, Y.; Dinh, C.-T.; Sinton, D.; Zheng, G.; Sargent, E. H. Enhanced Nitrate-to-Ammonia Activity on Copper-Nickel Alloys via Tuning of Intermediate Adsorption. *J. Am. Chem. Soc.* **2020**, *142* (12), 5702−5708.
(40) Lyu, Z.-H.; Fu, J.; Tang, T.; Zhang, J.; Hu, J.-S. Design of Ammonia Oxidation Electrocatalysts for Efficient Direct Ammonia Fuel Cells. *EnergyChem.* **2023**, *5* (3), 100093.
(41) Tian, Y.; Mao, Z.; Wang, L.; Liang, J. Green Chemistry: Advanced Electrocatalysts and System Design for Ammonia Oxidation. *Small Struct.* **2023**, *4*, 2200266.
(42) Yang, K.; Liu, J.; Yang, B. Electrocatalytic Oxidation of Ammonia on Pt: Mechanistic Insights into the Formation of N2 in Alkaline Media. *J. Catal.* **2022**, *405*, 626−633.
(43) Pillai, H. S.; Xin, H. New Insights into Electrochemical Ammonia Oxidation on Pt(100) from First Principles. *Ind. Eng. Chem. Res.* **2019**, *58* (25), 10819−10828.
(44) Li, Y.; Li, X.; Pillai, H. S.; Lattimer, J.; Mohd Adli, N.; Karakalos, S.; Chen, M.; Guo, L.; Xu, H.; Yang, J.; Su, D.; Xin, H.; Wu, G. Ternary PtIrNi Catalysts for Efficient Electrochemical Ammonia Oxidation. *ACS Catal.* **2020**, *10* (7), 3945−3957.
(45) Li, Y.; Pillai, H. S.; Wang, T.; Hwang, S.; Zhao, Y.; Qiao, Z.; Mu, Q.; Karakalos, S.; Chen, M.; Yang, J.; Su, D.; Xin, H.; Yan, Y.; Wu, G. High-Performance Ammonia Oxidation Catalysts for Anion-

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX

Exchange Membrane Direct Ammonia Fuel Cells. *Energy Environ. Sci.* **2021**, *14* (3), 1449−1460.
(46) Koper, M. T. M. Structure Sensitivity and Nanoscale Effects in Electrocatalysis. *Nanoscale* **2011**, *3* (5), 2054−2073.
(47) Katsounaros, I.; Figueiredo, M. C.; Calle-Vallejo, F.; Li, H.; Gewirth, A. A.; Markovic, N. M.; Koper, M. T. M. On the Mechanism of the Electrochemical Conversion of Ammonia to Dinitrogen on Pt(1 0 0) in Alkaline Environment. *J. Catal.* **2018**, *359*, 82−91.
(48) Matsui, T.; Suzuki, S.; Katayama, Y.; Yamauchi, K.; Okanishi, T.; Muroyama, H.; Eguchi, K. In Situ Attenuated Total Reflection Infrared Spectroscopy on Electrochemical Ammonia Oxidation over Pt Electrode in Alkaline Aqueous Solutions. *Langmuir* **2015**, *31* (42), 11717−11723.
(49) Pillai, H. S.; Li, Y.; Wang, S.-H.; Omidvar, N.; Mu, Q.; Achenie, L. E. K.; Abild-Pedersen, F.; Yang, J.; Wu, G.; Xin, H. Interpretable Design of Ir-Free Trimetallic Electrocatalysts for Ammonia Oxidation with Graph Neural Networks. *Nat. Commun.* **2023**, *14* (1), 792.
(50) Mou, T.; Han, X.; Zhu, H.; Xin, H. Machine Learning of Lateral Adsorbate Interactions in Surface Reaction Kinetics. *Curr. Opin. Chem. Eng.* **2022**, *36*, 100825.
(51) Stamenkovic, V. R.; Fowler, B.; Mun, B. S.; Wang, G.; Ross, P. N.; Lucas, C. A.; Markovic, N. M. Improved Oxygen Reduction Activity on Pt3Ni(111) via Increased Surface Site Availability. *Science* **2007**, *315* (5811), 493−497.
(52) Zhang, X.; Wang, Z.; Lawan, A. M.; Wang, J.; Hsieh, C.-Y.; Duan, C.; Pang, C. H.; Chu, P. K.; Yu, X.-F.; Zhao, H. Data-driven Structural Descriptor for Predicting Platinum-based Alloys as Oxygen Reduction Electrocatalysts. *InfoMat* **2023**, DOI: 10.1002/inf2.12406.
(53) Wang, S.; Omidvar, N.; Marx, E.; Xin, H. Overcoming Site Heterogeneity In Search of Metal Nanocatalysts. *ACS Comb. Sci.* **2018**, *20* (10), 567−572.
(54) Grimaud, A.; May, K. J.; Carlton, C. E.; Lee, Y.-L.; Risch, M.; Hong, W. T.; Zhou, J.; Shao-Horn, Y. Double Perovskites as a Family of Highly Active Catalysts for Oxygen Evolution in Alkaline Solution. *Nat. Commun.* **2013**, *4*, 2439.
(55) Suntivich, J.; May, K. J.; Gasteiger, H. A.; Goodenough, J. B.; Shao-Horn, Y. A Perovskite Oxide Optimized for Oxygen Evolution Catalysis from Molecular Orbital Principles. *Science* **2011**, *334* (6061), 1383−1385.
(56) Seitz, L. C.; Dickens, C. F.; Nishio, K.; Hikita, Y.; Montoya, J.; Doyle, A.; Kirk, C.; Vojvodic, A.; Hwang, H. Y.; Norskov, J. K.; Jaramillo, T. F. A Highly Active and Stable IrOx/SrIrO3 Catalyst for the Oxygen Evolution Reaction. *Science* **2016**, *353* (6303), 1011−1014.
(57) Man, I. C.; Su, H.-Y.; Calle-Vallejo, F.; Hansen, H. A.; Martínez, J. I.; Inoglu, N. G.; Kitchin, J.; Jaramillo, T. F.; Nørskov, J. K.; Rossmeisl, J. Universality in Oxygen Evolution Electrocatalysis on Oxide Surfaces. *ChemCatChem.* **2011**, *3* (7), 1159−1165.
(58) Li, Z.; Achenie, L. E. K.; Xin, H. An Adaptive Machine Learning Strategy for Accelerating Discovery of Perovskite Electrocatalysts. *ACS Catal.* **2020**, *10* (7), 4377−4384.
(59) Cornelio, C.; Dash, S.; Austel, V.; Josephson, T. R.; Goncalves, J.; Clarkson, K. L.; Megiddo, N.; El Khadir, B.; Horesh, L. Combining Data and Theory for Derivable Scientific Discovery with AI-Descartes. *Nat. Commun.* **2023**, *14* (1), 1777.
(60) Tran, R.; Lan, J.; Shuaibi, M.; Wood, B. M.; Goyal, S.; Das, A.; Heras-Domingo, J.; Kolluru, A.; Rizvi, A.; Shoghi, N.; Sriram, A.; Therrien, F.; Abed, J.; Voznyy, O.; Sargent, E. H.; Ulissi, Z.; Zitnick, C. L. The Open Catalyst 2022 (OC22) Dataset and Challenges for Oxide Electrocatalysts. *ACS Catal.* **2023**, *13* (5), 3066−3084.

https://doi.org/10.1021/accountsmr.3c00131
Acc. Mater. Res. XXXX, XXX, XXX−XXX