Computational Design of Optimal Bimetallic Nanoparticles: Bridging Stability with Adsorption

by

James Robert Dean

BS in Chemistry, Winthrop University, 2016

Submitted to the Graduate Faculty of the

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

James Robert Dean

It was defended on

December 14, 2020

and approved by

Christopher Wilmer, PhD, Associate Professor, Department of Chemical and Petroleum
Engineering

Götz Veser, PhD, Professor, Department of Chemical and Petroleum Engineering

Jill Millstone, PhD, Associate Professor, Department of Chemistry

Dissertation Director: Giannis Mpourmpakis, PhD, Associate Professor, Department of Chemical
and Petroleum Engineering

Computational Design of Optimal Bimetallic Nanoparticles: Bridging Stability with Adsorption

James Robert Dean, PhD

University of Pittsburgh, 2020

In the 20th century, advancements in computational power and chemical theory revolutionized catalyst discovery. Now in the 21st century, machine learning techniques have started making their way into the toolbox of computational chemistry, accelerating material design and discovery.

In this dissertation, we take a two-pronged approach to advance the current state-of-the-art when it comes to rational catalyst design. We begin by illuminating the fundamental physical properties relevant to the adsorption of small molecules to nanoparticles, arriving at a set of universal adsorption descriptors with the use of Density-Functional Theory calculations and machine learning. We then develop CE Expansion, a new open-source genetic algorithm for the rapid optimization of the bimetallic mixing behavior in nanoparticles of any size, shape, or composition, and use this to create the MetalNanoDB, a database of over 5,400 low-energy nanoparticles. Finally, we bring together these two approaches, and demonstrate a new workflow for the high-throughput screening of potential nanocatalysts for their physical properties, targeting CO_2 adsorption as a proof-of-concept (relevant to mitigating the greenhouse effect).

Overall, this work accelerates catalyst design by developing tools that rapidly and accurately model bimetallic nanocatalysts, and efficiently sieve the tremendously large nanomaterials space for targeted catalytic applications.

Table of Contents

Summary of Publications from PhD

1. Dean, J.; Yang, Y.; Austin, N.; Veser, G.; Mpourmpakis, G., Design of Copper-Based Bimetallic Nanoparticles for Carbon Dioxide Adsorption and Activation. *ChemSusChem* **2018**, 11 (7), 1169-1178.

2. Dean, J.; Taylor, M. G.; Mpourmpakis, G., Unfolding Adsorption on Metal Nanoparticles: Connecting Stability with Catalysis. *Science Advances* **2019**, 5 (9), eaax5101.

3. Dean, J.[*]; Cowan, M. J.[*]; Estes, J.; Ramadan, M.; Mpourmpakis G., Rapid Prediction of Bimetallic Mixing Behavior at the Nanoscale. *ACS Nano* **2020**, 14 (7) 8171-8180.

4. Dean, J.; Mpourmpakis, G., High-Throughput Screening of Bimetallic Nanoparticles for CO_2 Adsorption. *In Preparation.*

5. Tan, K.; Dixit, M.; Dean, J.; Mpourmpakis, G. Predicting Metal-Support Interactions in Oxide-Supported Single-Atom Catalysts. *Industrial & Engineering Chemistry Research* **2019**, 58 (44), 20236-20246.

6. Abdelgaid, M.; Dean, J.; Mpourmpakis, G. Improving Alkane Dehydrogenation Activity on γ-Al_2O_3 through Ga Doping. *Catalysis Science and Technology* **2020**, 10, 7194-7202.

---

[*] Both authors contributed equally.

7. <u>Dean, J.</u>; Yang, Y.; Veser, G.; Mpourmpakis, G., Is CuZrO$_3$ a Sandwich? Computational Insights into the Structure and Phases of Copper Zirconate. *In Preparation*.

Works 1-4 are highlighted in this dissertation.

Preface

There are so many friends, family, and organizations to thank for making this four-year endeavor possible. First, I want to thank my advisor, Prof. Giannis Mpourmpakis. Without Yanni's skill as an advisor and mentor, I wouldn't have grown nearly as much as a scientist. Neither would I have had the opportunity to perform this incredibly interesting research if it weren't for his continued support and advice.

I also want to thank my committee members for being a part of my academic journey. Dr. Chris Wilmer, for his expertise in molecular simulation, and for teaching the course that got me started with computer graphics and scientific visualization. Dr. Götz Veser, for his continued experimental collaboration with us, and his course in reactor engineering that helped me transition from my chemistry background into chemical engineering principles. And Dr. Jill Millstone, who has had an inspiring career in NP research, who also provides experimental expertise on NP synthesis and characterization.

The Mpourmpakis Lab (CANELa) especially has been home to the most talented scientists that I've met so far in my career. To Natalie, Pavlo, Michael T., and Xi – thank you all for the training and mentorship that helped me start my PhD successfully. And to those who arrived at the lab with and after me, for all of the collaboration, exchange of ideas, and conversations that kept me excited to visit the office every day: Zihao, Michael C., Robin, Anantha, Emily, Raffaele, Evan, Sungil, Mona, and Maya, thank you. To our post-docs, Mudit, Rosalba, Dylan, and Cristian, whose hard work raised the bar for everyone – thank you. And finally, many thanks to all the bright undergraduates who have been a part of CANELa during my time here: Ethan, Jonathan, Eric, Mahmoud, Mitch, Ashley, Macy, and Peter.

This academic journey wouldn't have been possible without the help of the many organizations and individuals I had the opportunity to work with over the years, but especially Yahui in Prof. Veser's group for the many hours in the lab synthesizing and characterizing our predicted catalysts, in addition to Evangelos, Prof. Chrysanthis, and the entire ADMT lab for helping make the MetalNanoDB a reality.

I also wish to extend my gratitude to the University of Pittsburgh's Center for Research and Simulation (CRC), for making this work possible by providing computational resources. And finally, I wish to thank the National Science Foundation for providing computational resources via the Extreme Science and Engineering Discovery Environment (XSEDE), and for funding my studies at Pitt under grant CMMI-1634880.

To Dalton, Adam, Joe, and the rest of my friends in Pittsburgh, for being there for me when things were at their most difficult, and for making the last four years as enjoyable as they were.

And finally, to my parents, Ken and Ann Dean, my sister, Kimberley, and the rest of the Dean family for always being there for me, constantly inspiring me to grow as a person.

To everyone who made this dissertation possible, from the bottom of my heart: thank you, I couldn't have done this without you!

1.0 Introduction

The rational design of new catalysts has fascinated scientists for the better part of the last century. This was motivated in part by the extremely labor-intense, oftentimes serendipitous process of catalyst discovery in the early 20th century[1, 2], which was exemplified by the history surrounding the development of the Haber-Bosch process. The first steps toward the high-pressure catalytic synthesis of ammonia were taken by Walther Nernst[3, 4] and improved catalysts were later developed by Fritz Haber, leading to his 1918 Nobel Prize in Chemistry[5]. Following this effort, further work was performed by Carl Bosch and Alwin Mittasch[4], which focused on the refinement of ammonia synthesis into an industrially-viable process. In all, over 2,500 catalysts would be tested across 6,500 experiments[6], and the innovations this work yielded in high-pressure reaction engineering would result in Bosch being awarded the 1931 Nobel Prize in Chemistry[5].

In the mid-20th century, advancements in theoretical chemistry, such as DFT[7-9] and HF[10] coincided with the onset of a revolution in computer hardware famously described by Moore's Law[11]. The attention of chemists immediately went to features relevant to catalysis: in 1975, one of the first applications of DFT was an investigation by Kohn – whose Kohn-Sham equations[12] formed the foundations of DFT and led to his 1998 Nobel Prize in Chemistry – regarding chemisorption to metal surfaces[13]. Using DFT, chemists could now investigate reactions and catalysts without needing experiment, and thus it found applications in a wide range of areas including transition-metal complexes[14, 15], the rationalization of catalyst behavior[16, 17], the study of zeolite structure and reactivity[18, 19], and the prediction of reaction pathways[20]. As a result of these advancements, catalysts and reactions could be screened *in-silico* before resource-intensive, time-consuming experiments were conducted. Thus, this allowed limited resources to be targeted

toward systems which had a high probability of working – certainly a long way from Haber's characterization of over 2,500 potential catalysts by hand.

And now, in the 21st century, machine learning has taken the stage as a young and rapidly-growing field[21]. Already, this area has made a significant impact on theoretical chemistry, just as the 20th century's advancements in hardware and theory accelerated experimental chemistry[22]. A major focus has been in the development of models which avoid computationally-expensive methods such as DFT, particularly for the identification of new catalysts. In adsorption alone, the last decade has brought with it a flurry of impactful work. In 2010, a new model connecting the BE of CO to the structure of Au NPs was developed [23], and later on the GCN model connected the BE of several adsorbates to the geometry of surface binding sites [4]. In 2018, Roling et al advanced this idea, incorporating further energetic terms describing a surface site's stability into a predictive model of adsorption [2]. Finally in 2020, we released a new universal model of adsorption to NPs of any size or shape, which was much more accurate than the other linear models that had come before it[24].

1.1 Machine Learning Meets Computational Chemistry

Any discussion of machine learning's impact on computational chemistry would be remis without a word on the impact neural networks have had. In 1997, neural networks were first applied to estimate electronic correlation in atoms and diatomic systems[25], in addition to representing the relatively simple potential energy surfaces of small systems[26]. Later, in 2007, a neural network was used to predict DFT energy as a function of atomic position in bulk Si[27]. The key impact of this work is that it also delivered so-called symmetry functions, a new representation of atomic

structure which now allowed neural network-based atomic potentials to be transferrable to other systems. It is also worth pointing out that both of these networks were trained years before AlexNet (a deep CNN which displayed exceptional accuracy in image classification) put the power of modern neural network techniques front-and-center in the public imagination in 2012[28].

In 2017, ANI-1 was released. This method was a ground-breaking neural network-based approach to computational chemistry, delivering energetics with the accuracy of DFT at the computational expense of a forcefield[29]. As a result of this success, neural network based potentials have become relatively popular to describe chemical interaction – in 2019 alone, the ANN DFT functional was introduced[30], the EANN was developed for molecular dynamics[31], a CNN was trained to predict binding to metal surfaces [32], and a deep neural network was trained to directly predict a molecule's wavefunction[33]. And even now, unabated by 2020's numerous challenges, machine learning continues to take the world of computational chemistry by storm: the wavefunction predictions of Schütt et al[33] have been improved and packaged as a new model called SchNOrb[34], the EANN was successfully applied to the prediction of adsorption energetics[35], and a new GCNN-based approach for the prediction of NMR chemical shifts was developed[36]. Yet despite this continued attention, there is still much fruitful research to be had. Key questions currently are feature selection, finding effective molecular representation schemes, and the development of large, high-quality training sets[26].

This need for large datasets has not been un-noticed, and has led to many different public databases including the OQMD[37, 38], Materials Project[39], MetalNanoDB[24], Catalysis Hub[40], and the NIST CCCBDB[41]. This availability of data has been a critical step in the development of chemistry-focused machine learning, so much that it is a major thrust of the Materials Genome Initiative, a US-backed effort to develop the next generation of chemical research techniques[42].

In addition to neural networks, a wealth of other machine learning techniques has been applied to chemical problems. In the late 1960's, DENDRAL, the first expert system – an algorithm designed to mimic a human expert's decision-making process – to be developed for chemical problems successfully predicted several chemical structures from mass spectrometry data[43]. Despite their age and complexity compared to other machine learning techniques, expert systems still find occasional use in chemistry problems[44], and one was even proposed in 2014 for the prediction of supramolecular structures[45].

Random forests are also quite popular in addressing the complex problems found in chemical modeling. Although conventionally viewed as a classification algorithm, they have enjoyed great success in both regression and classification roles. Random forests have been used to detect food adulterants based on spectroscopic data[46], to predict sites in DNA which can undergo citrullination[47], to model the chemical composition of the atmosphere[48], and even to predict fundamental molecular properties such as solubility[49]. Success has particularly been found by the medicinal chemistry sector, who have employed random forests to aid drug discovery by predicting QSARs[50, 51] and other key pharmaceutical properties[52-55].

1.2 Ab-Initio Structure Prediction

The *ab-initio* prediction of chemical structure is a challenging and constantly-evolving field. In this area, much work in the latter half of the 20th century went into crystal structure prediction[37, 38, 56-61], a longstanding goal of materials science[62]. One particularly effective algorithm in this area has been USPEX [59], which has yielded a plethora of discoveries including new high-pressure crystals of $NaCl$[63] and Al_2O_3[64], the prediction of a high-pressure, high-

temperature methane phase diagram to rationalize phenomena observed on Uranus and Neptune[65], and evidence for a new crystal structure of $(Mg,Fe)SiO_3$ within the conditions present in the Earth's crust[66, 67]. USPEX is a GA typically parameterized by DFT calculations (but can interface with less-expensive methods), and optimizes a crystal structure's cell dimensions and atomic positions.

A significant amount of recent attention has also gone into the prediction of NP structure. This has resulted in a wide range of new techniques entering the computational chemistry toolbox in recent years, including those which globally optimize the NP structure[68, 69], those which predict the trend in segregation behavior[70-74], and those which predict chemical-ordering[75] – that is, the distribution of the elements within a NP.

Many of these approaches utilize GA-based techniques for their structural optimization. Moreover, in general GAs are powerful optimization tools which lend themselves well into the types of high-dimensional optimization problems found in chemistry[76]. This has resulted in their widespread adoption in a variety of chemical problems including mass spectrometry[77], spectroscopy[78, 79], computational geometry optimization[80], drug design[81], and in chemical engineering problems such as kinetics[82], reactor engineering[83-85], process control[86, 87], and transport[88, 89].

Because GAs have performed so well for so many other problems, as part of this thesis we developed our own GA for NP chemical order optimization, CE Expansion[75]. This algorithm (which is explained in more detail in Section 4.0, Rapid Prediction of Bimetallic Mixing Behavior at the Nanoscale) is parameterized by the BCM of NP stability[90], which is a highly accurate description of NP CE. Leveraging this tool, even a typical desktop computer can rapidly screen through millions of structures to optimize a NP's chemical ordering. Moreover, we have even shown that the predicted chemical ordering is exceedingly close to the one that is expected in

experiments – and thus has advanced the state of the art by providing a tool for the rapid generation of improved model NP systems for further computational study.

1.3   Tooling for Rational Catalyst Design

The advancements we have outlined so far all share a common thread: abstraction. Specifically, abstraction to a simpler (or at least, faster to solve) problem. In the 20th century, theoretical chemistry abstracted lab work onto the computer, which allowed the limited resources of an experimental lab to be better targeted. In the 21st century, machine learning has begun to do the same for computational chemistry: screening chemicals in a computationally-efficient manner, to better utilize the limited computational resources available.

These improved techniques have resulted in significant attention being given to the so-called rational design of catalysts. This design paradigm stresses the importance of understanding the fundamental physical properties of catalysts, and their relationship with catalytic activity[91]. Given this understanding, the goal is to tune a catalyst's physical properties such that it is driven to a maximum catalytic activity.

The objective of this dissertation is to provide an improved set of tools for the rational design of novel catalysts. Due to its longstanding[92] relevance as a known greenhouse gas, which has led to no shortage of research in the literature[93-108], for this purpose we focus on CO_2. Part of our justification is that CO_2 conversion is a common subject of study in the field of rational catalyst design[109-111], and thus our contributions have the potential for a wide impact. Moreover, CO_2 is a molecule the properties of which make it very amenable to computational study. Namely, its small size makes it tractable to investigate via DFT, yet it is still complex-enough to bind to a potential

catalyst via multiple atoms. This complexity in its structure also leads to a convenient way to visually infer whether it has received electron density: upon receiving electron density, its bond angle decreases[112, 113].

We begin by building up an understanding of properties relevant to CO_2 adsorption in Section 2.0 (Designing Copper-Based Bimetallic Nanoparticles for CO_2 Activation). Specifically, we identify the importance of properties enabling a transfer of charge to CO_2. This leads to our development of the ΔIP descriptor, which relates the IP of a binding site atom and its neighbors. Following this in Section 3.0 (Unfolding Adsorption on Metal Nanoparticles: Connecting Stability with Catalysis), we develop a universal model of small-molecule adsorption to NPs and surfaces of any size, shape, or composition. Moreover, we note a clear enhancement in performance compared to other contemporary predictive models of adsorption. Also in this chapter, we develop two key descriptors: CE_{local} and MADs. These descriptors, which relate to the stability of the binding site and its intrinsic tendency to bind with the adsorbate, are specifically what enable this new model to perform so well.

Next in Section 4.0 (Rapid Prediction of Bimetallic Mixing Behavior at the Nanoscale), we create CE Expansion, a GA which is able to rapidly optimize any arbitrary NP chemical ordering. After demonstrating its efficiency at quickly identifying low-energy NP chemical orderings, and even making direct comparison to an experimentally synthesized $Fe_{6569}Pt_{16627}$ NP whose exact structure is known with atomic precision[114], we use CE Expansion to optimize an initial set of 18 NP structures. These NPs come from all monometallic and bimetallic systems of Cu, Ag, Au, Pd, and Pt. To facilitate the systematic, mass screening of all possible binding sites on these NPs, we begin Section 5.0 (High-Throughput Screening of Bimetallic Nanoparticles for CO_2 Adsorption) by developing AutoAdsorbatePlacement. This code consists of a hashing strategy

for the automated detection of unique binding sites (based on CE_{local}), as well as the automated placement of molecules in pre-adsorbed (and in the case of CO_2, pre-activated) positions. Overall, this results in a database of over 660 unique CO_2 BEs, which we intend to make public to facilitate further work on this subject.

Finally, combining this dataset with the ΔIP, CE_{local}, and MADs descriptors for adsorption we identified in Sections 2.0 and 3.0, we train several machine learning models for the rapid prediction of CO_2 adsorption. We then use the best model, a random forest, to screen over 46,000 binding sites, as a proof-of-concept of the workflow developed in this study. Finally, our random forest is used to visualize and investigate a hypothetical chemical space defined by our three descriptors.

In the course of this dissertation, we sought to accelerate materials discovery by producing a universal set of tools relevant to materials screening. We have accomplished this goal by I) identifying tabulated or easy-to-calculate universal adsorption descriptors in Sections 2.0, 3.0, and 5.0, II) developing and releasing the open-source GA, CE Expansion, leading to the MetalNanoDB, a publicly-accessible database of 5,454 unique, low-energy NP structures in Section 4.0, III) creating AutoAdsorbatePlacement, an algorithm for the automated, systematic identification of potential binding sites and the placement of adsorbates in Section 5.0, and finally IV) tying all of the prior work together to rapidly screen over 46,000 binding sites and explore a hypothetical chemical space of NP catalysts in Section 5.0. In short, as a result of this dissertation's many contributions to nanocatalyst modeling, we have demonstrated steps towards advancing the current state-of-the-art in rational catalyst design.

2.0 Designing Copper-Based Bimetallic Nanoparticles for CO₂ Activation

The content of this chapter is adapted from Dean, J.; Yang, Y.; Austin, N.; Veser, G.; Mpourmpakis, G. "Design of Copper-Based Bimetallic Nanoparticles for Carbon Dioxide Adsorption and Activation." ChemSusChem 2018, 11 (7) 1169-1179.

We begin this work by studying the activation of CO_2, a well-known greenhouse gas[115]. Doped-Cu NPs for this purpose, as although monometallic Cu poorly activates CO_2[116], prior work has shown that doping Cu NPs with heterometals can facilitate the activation of CO_2[101, 108]. Thus, in this section we investigate the effects of doping Cu NPs with a variety of d-block metals, in order to better-understand the fundamental materials properties relevant to CO_2 adsorption and activation.

2.1 Computational Methodology

All calculations were performed via the Quickstep algorithm[117] as implemented in CP2K[118] using the PBE[119] functional in conjunction with the GTH[120] pseudopotentials and a 500 Ry cutoff, and the DZVP basis sets by VandeVondele and Hutter[121]. In order to account for dispersion interactions, the DFT-D3 correction by Grimme[122] was used. The geometry of the studied NPs was optimized in non-periodic $30 \times 30 \times 30$ Å³ cells until forces lower than $0.02 \frac{eV}{Å}$ were achieved, with a criterion of 10^{-7} for SCF convergence. Vibrational analysis calculations were performed to verify that optimized structures were at an energetic minimum. Charges were calculated using the Bader analysis[123].

We accounted for an icosahedral Cu_{55} NP, where we doped a single surface site at a time with a hetero-metal M. Due to the NP's symmetry, only two unique surface doping sites exist: a CN 6 site, and a CN 8 site (Figure 2.1).
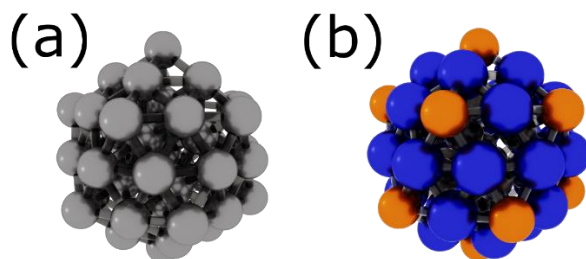
Figure 2.1: (a) Geometry of an icosahedral 55-atom NP with (b) unique surface atoms highlighted as orange for CN=6 and blue for CN=8.

For every species $Cu_{54}M$ with M=Au, Mn, Mo, Ni, Pd, Rh, Ru, Sc, V, Zn, and Zr, two NPs were investigated: one with a dopant in the CN6 and one with the dopant in the CN8 position. The dopants (M) were randomly selected to span over different rows and columns of transition metals, in order to capture the presence (or lack) of any periodic trends in their CO_2 adsorption response. The segregation energies of each NP were calculated using Equation 2.1. In this case, a negative segregation energy is indicative of a dopant preferentially occupying a surface site.

$$E_{seg} = E_{Cu_{54}M\,(surface)} - E_{Cu_{54}M\,(core)} \qquad 2.1$$

Additionally, we calculated the BE of CO_2 on the catalysts under different adsorption configurations by comparing the total electronic energy of the $Cu_{54}M$-CO_2 complex with the energies of the infinitely separated $Cu_{54}M$ and CO_2 systems using Equation 2.2.

$$E_{bind} = E_{Cu_{54}M-CO_2} - E_{Cu_{54}M} - E_{CO_2} \qquad 2.2$$

In order to study chemisorption, two initial configurations were chosen for study: CO_2 with the carbon closest and the oxygens equidistant to the dopant (Figure 2.2 (a)) and CO_2 parallel to the surface of the NP with oxygen closest to the dopant (Figure 2.2 (b)). In both cases, CO_2 was initially located 2Å away from the NP. Physisorption was studied with two initial configurations of CO_2: CO_2 being approximately 4.5Å away from the NP, interacting with its carbon atom (Figure 2.2 (c)) and with one of the oxygen atoms (Figure 2.2 (d)).
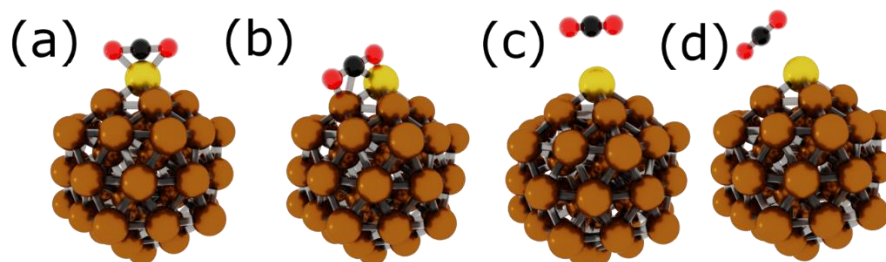
Figure 2.2: Various (initial) adsorption configurations tested on the doped $Cu_{54}M$ NPs, where M represents the chosen dopant located in this case in the CN=6 position. (a) and (b) represent two chemisorption configurations, with CO_2 2Å away from the NP, interacting with carbon and oxygen atoms, respectively, on the dopant site. (c) and (d) represent two physisorption configurations with CO_2 being 4.5Å away from the NP and interacting with its carbon and oxygen atom, respectively.

2.2 Experimental Methodology

All experimental work in this study was performed by Yahui Yang in the lab of Prof. Veser as part of a collaboration with the computational lab of Prof. Mpourmpakis.

2.2.1 Catalyst Preparation

Based on the computational results (see 'Results and Discussion' below), CuZr was chosen for experimental validation of the computational predictions. Synthesis of a mixed CuZr catalyst system proved challenging due to the tendency of this system to phase segregate, and the final catalyst for the present studies was synthesized by a simple wet-impregnation approach despite the comparatively low degree of control over the resulting material (which was hence thoroughly characterized before further study, see below).

Briefly, for a typical bimetallic catalyst with 30wt% combined metal loading (equimolar amount of Cu and Zr), 195 mg of $Cu(NO_3)_2 \cdot 2.5H_2O$ (> 99.99%, Sigma Aldrich) and 196 mg $ZrCl_4$ (> 99.5%, Alfa Aesar) were dissolved in ethanol (1 mL, 200 proof). 300 mg commercial SiC (> 99%, Sigma Aldrich) powder as support material was added to the solution. After magnetic stirring at room temperature for 2 hours, the mixture was dried overnight at 363 K in a vacuum oven. The resulting solid cake was grinded and then calcined in air (0.2 SLM) at 773 K for 3 hours. The weight loading of the resulting catalyst was verified via EDS which showed 11.21 wt% Cu and 19.65 wt% Zr. As a reference, Cu/SiC (15.45 wt% Cu) and ZrO_2/SiC (17.16 wt% Zr) were also synthesized by the same approach.

2.2.2 Sample Characterization

The catalyst was characterized via XRD, XPS, HRTEM and STEM after reduction in H_2 at 523 K for 1 hour prior to any characterization.

XRD Analysis: The sample crystal structures were characterized by Bruker D8 X-ray Diffraction system at 40 kV and 40 mA with Cu Kα radiation (λ = 1.5418 Å). The XRD patterns

were recorded from 20° to 80° (2θ) with a scan speed of 0.3 second/step. The phases were identified using the JCPDS database.

XPS Analysis: XPS was performed using a Thermo ESCALAB 250Xi. Survey and high resolution spectra were collected with a pass energy of 200 eV and 50 eV, respectively. All XPS spectra were measured with a 650 µm X-ray spot size.

Electron Microscopy: HRTEM and STEM characterization were performed using a JEOL JEM-2100F equipped with Oxford X-Max 80TLE EDS detector operating at 200 kV. The measurement of d-spacing from HRTEM images was done manually using ImageJ software.

2.2.3 CO$_2$ Temperature Programmed Desorption

CO$_2$ TPD experiments were performed at ambient pressure using a packed-bed quartz reactor connected to a mass spectrometer (Pfeiffer Vacuum Quadstar). 300 mg catalyst was used for each CO$_2$ TPD experiment. Prior to the adsorption of CO$_2$, the catalysts were reduced in H$_2$ at 523 K for 1 hour and then heated in a He stream at 773 K for 1 hour to clean the surface from any residual adsorbed gases. After cooling to room temperature, the catalyst was saturated with CO$_2$ (15 SCCM of 20vol% CO$_2$ in He), and then flushed with He (25 SCCM) to remove any physisorbed CO$_2$. The TPD experiment was then started with a heating ramp of 20 K/min under He flow (25 SCCM), and the desorbed CO$_2$ was monitored by mass spectrometry.

2.3 Results and Discussion

2.3.1 Identification of Physical Properties Affecting CO$_2$ Binding Strength

As we are interested in studying the effect of metal dopants in Cu NPs on CO$_2$ adsorption, we primarily focus on the dopant's preference to occupy a surface site on the NP. Therefore, potential bimetallic NPs were chosen on the basis of having negative segregation energy on either CN=6 or CN=8. The calculated segregation energies are presented in Figure 2.3. Our calculated dopant segregation energies are qualitatively consistent with those tabulated by Ruban et al[124] on periodic metal surfaces, which indicated favorable segregation for Cu alloys containing Zr, Au, and Pd, as well as favorable antisegregation for Cu alloys containing Ni, Mn, Rh, Ru, and V. In the case of Mo, Ruban and coworkers indicated a moderate antisegregation behavior, which we indeed identify as a borderline case, with a negative segregation energy observed at a CN 6 site, and a positive (i.e. antisegregation) at a CN 8 site. In accordance with this (segregation energy) selection criterion (results in Figure 2.3), we chose Au, Mo, Pd, Sc, Zn, and Zr as potential dopants to investigate the CO$_2$ adsorption. In addition, the monometallic Cu$_{55}$ case was studied, to provide a baseline for which the CO$_2$ adsorption on the different metal dopants of the Cu-based NP could be compared.
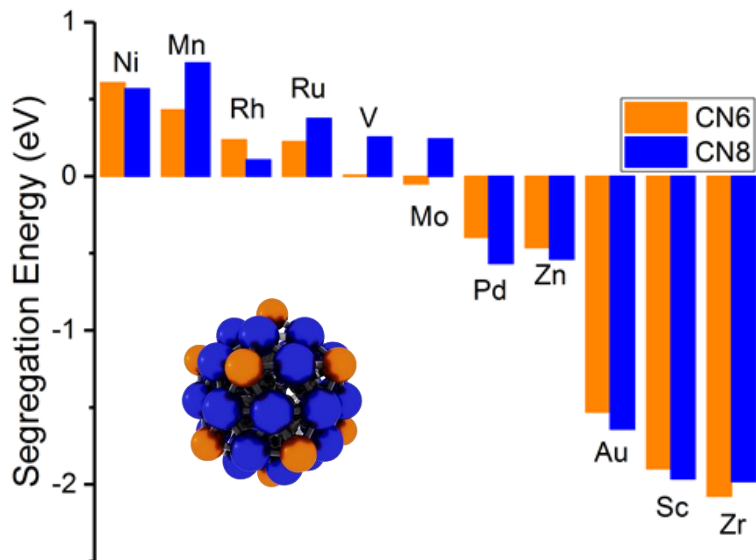
Figure 2.3: Segregation energies for candidate M dopants on Cu₅₄M NPs (M = Au, Mn, Mo, Ni, Pd, Rh, Ru, Sc, V, Zn and Zr), calculated using Equation 2.1. The inset picture demonstrates the different surface coordination of a Ih 55-atom NP colored in consistency with the chart's legend (CN=6: orange and CN=8: blue).

After selecting the dopants with favorable surface segregation, we performed a series of calculations to determine whether CO_2 preferentially physisorbs or chemisorbs to the adsorption site on the NP. The most favorable adsorption configuration is shown in Table 2.1. In the case of favorable physisorption, BEs ranging from -0.10 to -0.17 eV were observed. In the case of favorable chemisorption, BEs ranging from -1.26 to -2.00 eV were observed. In all cases of chemisorption, CO_2 adsorbs more favorably in a bent configuration with one oxygen pointed away from the surface.

Our physisorption energies are comparable to those of Ko et al[125], who indicated that in the case of physisorption on periodic metal surfaces the interactions between any of their studied metal surfaces and CO_2 were relatively weak, with BEs on the order of -0.3 eV regardless of the metal surface. The quantitative difference between Ko's calculated energies and ours (~0.15eV) can be attributed to an intrinsic difference between their systems and our own: their systems were periodic

metal surfaces, and our systems are metal NPs. Because physisorption is dominated by relatively weak dispersion forces, we can therefore expect changes in the adsorption behavior based on differences in the CN and morphology between these two systems.

Table 2.1: CO_2 BEs for several $Cu_{54}M$ NPs (M=Au, Cu, Mo, Pd, Sc, Zn, or Zr), calculated using Equation 2.2. In the case of Cu_{55}, the "dopant CN" refers to the CN of the monometallic binding sites as shown in Figure 2.1.

Species	Dopant CN	BE (eV)
Cu_{55}	6	-0.15
$Cu_{54}Au$	6	-0.17
$Cu_{54}Mo$	6	-1.26
$Cu_{54}Pd$	6	-0.09
$Cu_{54}Sc$	6	-1.43
$Cu_{54}Zn$	6	-0.14
$Cu_{54}Zr$	6	-1.62
Cu_{55}	8	-0.10
$Cu_{54}Au$	8	-0.17
$Cu_{54}Mo$	8	-1.71
$Cu_{54}Pd$	8	-0.11
$Cu_{54}Sc$	8	-1.55
$Cu_{54}Zn$	8	-0.16
$Cu_{54}Zr$	8	-2.00

Since charge transfer and CO_2 BE are directly related[126], we calculated the charge transferred to CO_2 when adsorbed on the NPs (Figure 2.4). We observe stronger binding when more charge is transferred to CO_2, in agreement with previous computational studies[101]. The results of Figure 2.4 show a very clear separation between the two regimes, physisorption and chemisorption. Physisorption is characterized by weak binding and little to no charge transfer to CO_2, whereas, chemisorption occurs when electron density close to one electron is transferred to CO_2.
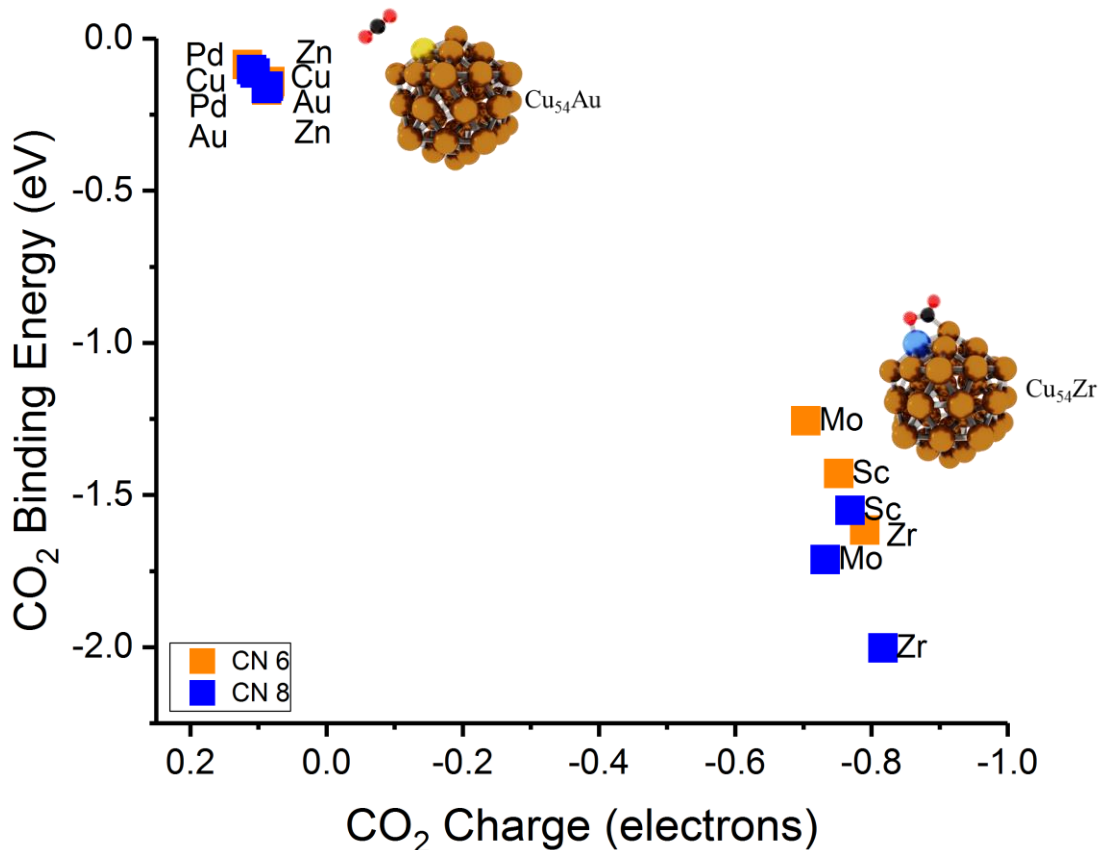
Figure 2.4: BE of CO_2 on $Cu_{54}M$ vs. charge transfer to CO_2. The inset pictures demonstrate the two adsorption regimes: physisorption (example of $Cu_{54}Au$) and chemisorption (example of $Cu_{54}Zr$).

Alongside the more favorable binding resulting from negative charge being transferred to CO_2, we also observe a decrease in bond angle from 180° to approximately 120° as shown in Figure 2.5. This change in bond angle can be explained by referring to a Walsh diagram of CO_2[127], which plots the relative energies of CO_2's molecular orbitals versus the O-C-O bond angle. When CO_2 receives an electron, the geometry of the molecule changes to minimize the energy of the CO_2^- HOMO. This CO_2 bending has been predicted to be at an angle of approximately 120° on the Walsh diagram, in agreement with our calculations.
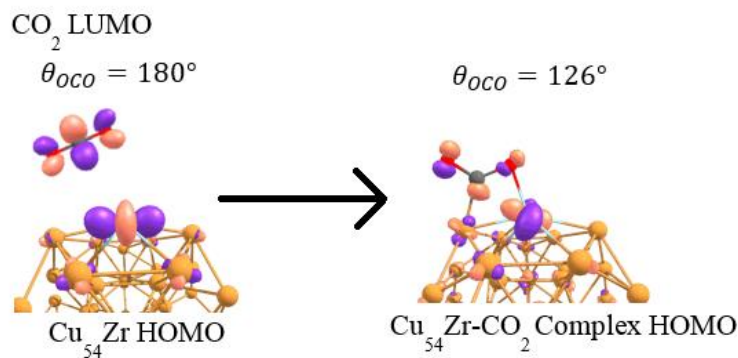
Figure 2.5: Observed change in the CO₂ bond angle (180° → 126°) as a result of chemisorption to Cu₅₄M (in this case M=Zr), as well as illustrations of the frontier molecular orbitals of all involved species.

As CO_2 gains charge in the case of favorable chemisorption, it is therefore relevant to examine the charges localized on the NP, with a focus specifically on the binding (dopant) site. It was found that only binding sites possessing a positive charge on the bare NP (absence of CO_2) were able to chemisorb CO_2 (Figure 2.6 (a)). As the IP, electronegativity, and EA of an atom are all related to one another[128, 129], the net charge of the dopant prior to interaction with an adsorbate can be qualitatively estimated by comparing the atomic IPs[130] of copper and the dopant atom, via Equation 2.3:

$$\Delta IP = IP_{Dopant} - IP_{Cu} \qquad \qquad 2.3$$

This ΔIP parameter has been plotted against the most favorable BE in Figure 2.6 (b). One can easily notice that only dopants with IPs lower than that of Cu were able to chemisorb CO_2. This is consistent with the presence of a positive charge on dopants binding CO_2 the most strongly (Figure 2.6 (a)). The charge transfer on the dopant and the ΔIP parameter are related as shown in Figure 2.6 (c) (solid points). In addition, Figure 2.6 (c) demonstrates that due to the higher electronegativity and IP relative to its dopant, the Cu atoms pull electron density following removal

of charge by CO_2, resulting in the development of an even more positive charge on the dopant atom (opened symbols).
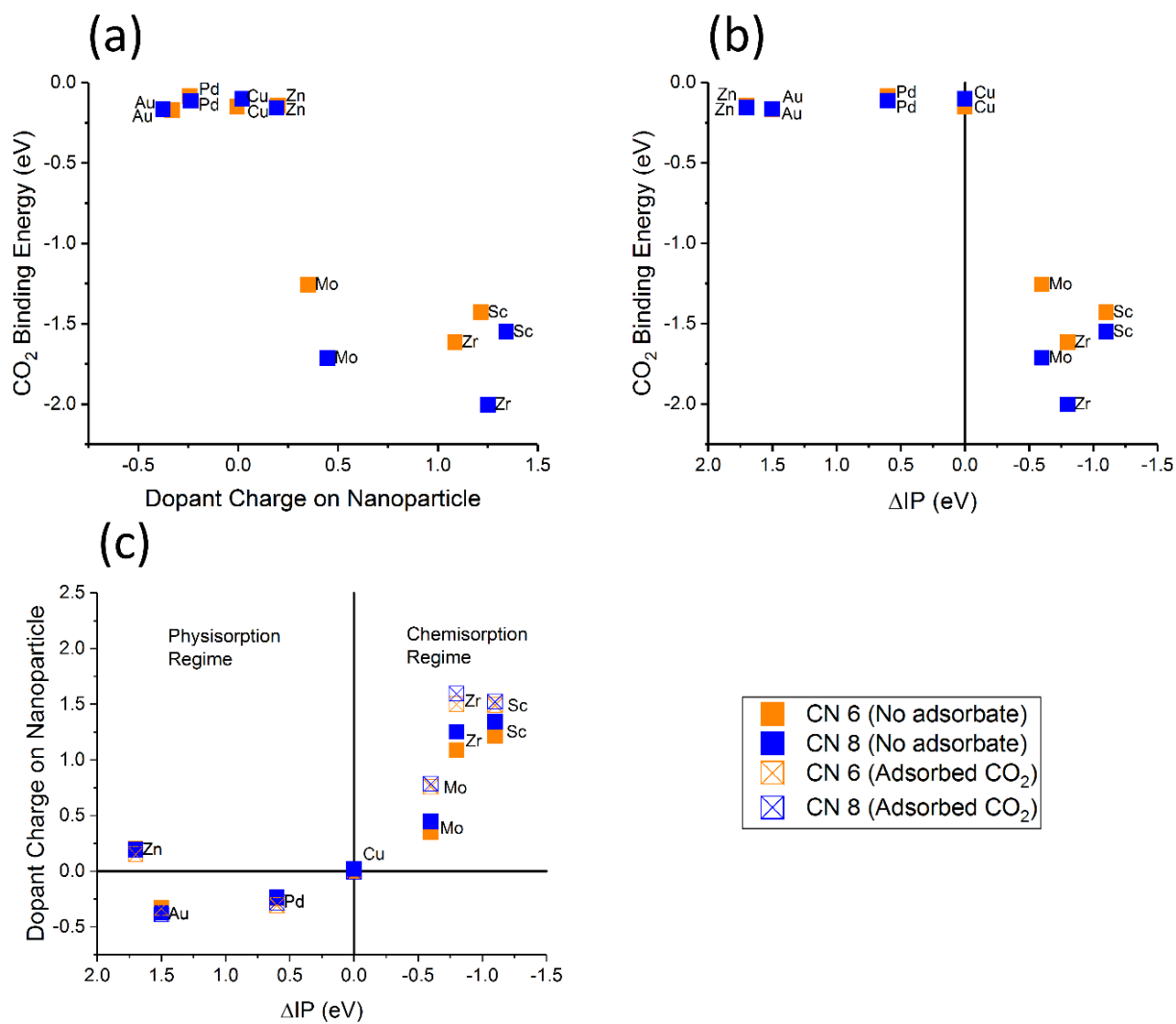
Figure 2.6: (a) BE of CO_2 on $Cu_{54}M$ vs. the charge of M. (b) BE of CO_2 on $Cu_{54}M$ vs. ΔIP of Cu and M, calculated by Equation 2.3. (c) Charge of M versus ΔIP of Cu and M before (solid) and after CO_2 adsorption (opened symbols). In all cases, M = Au, Cu, Mo, Pd, Sc, Zn, or Zr.

Although Figure 2.6 (c) shows that Zn exhibits a positive charge, it cannot activate CO_2. In order to explain this apparent discrepancy, we examine the local d_c of the dopant atoms, since it is known that the local *dc* of a catalyst can directly affect species adsorption[131-133]. Figure 2.7 presents the local d_c of various dopant atoms as well as the LUMO of neutral CO_2, and the HOMO of the CO_2^- anion. The two regimes, chemisorption and physisorption, readily differentiate themselves on this graph: all species which favorably chemisorb CO_2 possess a local d_c higher in energy than that of the CO_2 LUMO. Species which have a local d_c higher in energy than CO_2 are able to facilitate the transfer of electrons to CO_2 because the higher-energy electrons of the local d_c are able to "fall" into the lower-energy CO_2 LUMO (Figure 2.5). As the local d_c of Zn is significantly lower than the CO_2 LUMO, the electron transfer cannot occur and CO_2 cannot be activated.
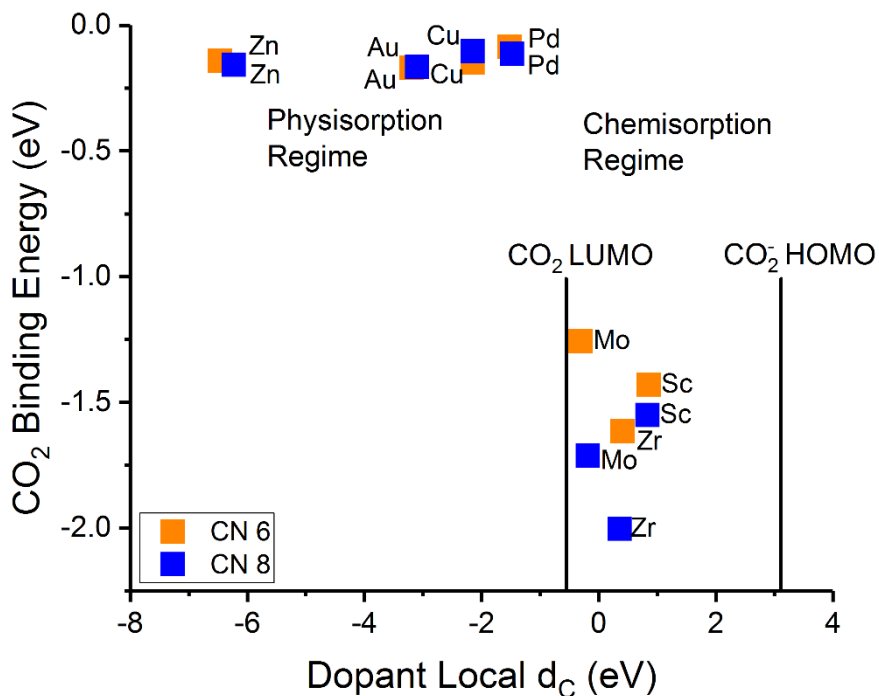
Figure 2.7: BE of CO_2 on $Cu_{54}M$ vs. M local dc. The vertical lines indicate the respective energies of the LUMO of CO_2 and the HOMO of a CO_2^- anion, where the molecules are in the gas phase and without any NPs. M = Au, Cu, Mo, Pd, Sc, Zn, or Zr.

Taking all these observations together, in order for CO_2 chemisorption to occur, the adsorption site on the NP must have a local d_c higher in energy than that of the CO_2 LUMO. This allows for electron transfer to occur to CO_2. In addition, the dopant atom should possess a positive charge which becomes more positive upon CO_2 chemisorption. In other words, the dopant should have the tendency to give electrons and become positively charged. This positive charge can be qualitatively estimated by comparing the IPs of Cu and the dopant atom. This charge transfer also serves to enhance the binding with CO_2, as the partially-negative oxygen atoms gain a stronger attraction to the partially-positive dopant atom.

2.3.2 Further Investigations into Doping Cu Nanoparticles with Zr

Of the studied dopant structures, $Cu_{54}Zr$ exhibited the strongest CO_2 BE, which is consistent with the descriptors we have outlined: it has a local dc higher in energy than CO LUMO, develops the highest charge (~0.8 electrons) on CO2 and has the second-highest most positive charge and negative ΔIP value. This observation of the relatively good ability of Zr to activate CO2 on Cu NPs is consistent with the findings of Fischer et al[94], who observed that the presence of ZrO2 promotes the rate and selectivity of methanol production on Cu/SiO2 catalysts. In addition, Fischer, Hong and Liu[134] suggested that the function of ZrO2 sites in the methanol synthesis reaction is to dissociatively chemisorb CO2. It should be noticed that previous computational findings by Austin et al[101] regarding CO2 adsorption on Zr-decorated Cu NPs showed a favorable dissociative chemisorption of CO2 on the Zr site. In Table B.1 of Appendix B.2 we present calculations of CO2 adsorption on Zr-doped Cu NPs of different sizes and we observe strong CO2 chemisorption on each of the NPs. This further supports that it is the presence of the Zr dopant that results to strong CO2 adsorption and activation and not a NP size effect.

Taking into account the ability of Zr to strongly chemisorb and activate CO_2 when used as a dopant on Cu_{55} NPs (**Figure** 2**.**4 – Figure 2.7), and the aforementioned body of experimental and theoretical work supporting a potential oxidation of this site, we decided to address the CO_2 adsorption behavior on $Cu_{54}Zr$ NPs, where the dopant site has been oxidized from the CO_2 environment. To this end, we constructed $Cu_{54}ZrO_2$ and $Cu_{54}ZrO_4$ NPs with the Zr atom on either a CN=6 or a CN=8 site prior to oxidation. The lowest-energy oxidation configurations (shown below in Figure 2.8) were selected for further CO_2 adsorption calculations.
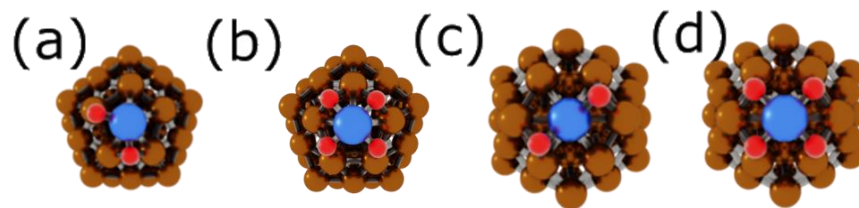
Figure 2.8: Surface oxidation configurations for studied Cu$_{54}$Zr NPs, with Zr atom being oxidized with either 2 or 4 oxygen atoms: (a) Cu$_{54}$ZrO$_2$, with Zr on a CN=6 site prior to oxidation. One oxygen bridges a Zr-Cu bond, and the other is on a hollow-site position between Cu and Zr. (b) Cu$_{54}$ZrO$_4$, with Zr on a CN=6 site prior to oxidation. All four oxygen atoms occupy hollow-site positions between Cu and Zr. (c) Cu$_{54}$ZrO$_2$, with Zr on a CN=8 site prior to oxidation. Both oxygens are on hollow-sites between Cu and Zr. (d) Cu$_{54}$ZrO$_4$, with Zr on a CN=8 site prior to oxidation. All four oxygen atoms are on hollow sites between Cu and Zr.

Upon constructing oxidized Zr surface-sites on the Cu NPs, we investigated the adsorption of CO$_2$ onto these sites. The adsorption configurations shown in Figure 2.9 were found to be the lowest in energy, and the adsorption energy values are listed in Table 2.2.
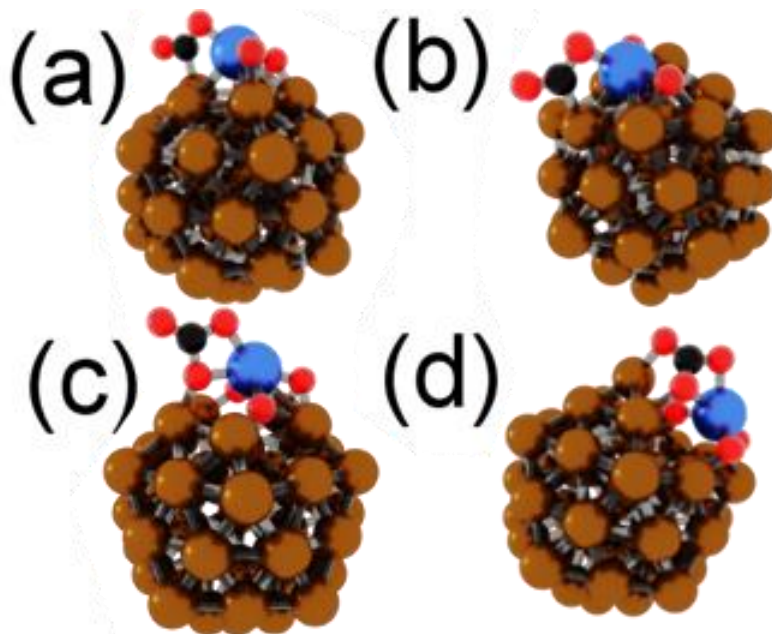
Figure 2.9: Lowest-energy adsorption configurations of CO_2 onto $Cu_{54}ZrO_2$ with Zr occupying a CN=6 (a) and CN=8 (b) adsorption site prior to oxidation and $Cu_{54}ZrO_4$, with Zr occupying a CN=6 (c) and CN=8 (d), prior to oxidation.

Notably, we observe favorable chemisorption on all oxidized $Cu_{54}Zr$ systems studied. The BE is decreased upon the addition of two oxygen atoms (Figure 2.9 (a-b)), whereupon CO_2 adsorbs on the oxidized surface Zr (and metallic Cu), and not the surface oxygen atoms. Interesting enough is the observation that in the case of four oxygen atoms (a highly oxidized Zr site, Figure 2.9 (c-d)), CO_2 binds very strongly, reacting with a surface O^{2-} anion to form carbonate. A similar adsorption configuration of CO_2 on the $Cu_{54}ZrO_2$ system forming carbonate was investigated, but was found to be higher in energy than the adsorption configurations shown in Figure 2.9 (a-b). In all four cases in Figure 2.9, NEB calculations using physisorbed CO_2 as a starting point and the respective adsorbed configuration as an ending point reveal a barrierless chemisorption of CO_2.

Table 2.2: Lowest-energy adsorption configurations of CO_2 onto $Cu_{54}ZrO_2$ with Zr occupying a CN=6 (a) and CN=8 (b) adsorption site prior to oxidation and $Cu_{54}ZrO_4$, with Zr occupying a CN=6 (c) and CN=8 (d), prior to oxidation.

Species	CO_2 BE (eV)	Zr Local d_c (eV)	Zr Charge (e⁻)
$Cu_{54}Zr$ (Zr on CN=6)	-1.62	0.4	1.1
$Cu_{54}ZrO_2$ (Figure 2.9 a)	-1.18	0.9	1.9
$Cu_{54}ZrO_4$ (Figure 2.9 c)	-1.67	1.1	2.3
$Cu_{54}Zr$ (Zr on CN=8)	-2.00	0.4	1.3
$Cu_{54}ZrO_2$ (Figure 2.9 b)	-0.94	0.8	1.8
$Cu_{54}ZrO_4$ (Figure 2.9 d)	-1.37	1.2	2.3

The reason CO_2 can still bind the Zr atoms despite their higher degree of oxidation is the local d_C of Zr in the NPs: The local d_c of the Zr in $Cu_{54}ZrO_X$ (X=0, 2, 4) generally increases (Table 2.2) as the Zr is further oxidized. As a result, these values are still close and higher in energy than the CO_2 LUMO (Figure 2.7), which makes the adsorption and activation of CO_2 continually favorable. Overall, our results demonstrate that although Zr is a very oxophilic metal and binds CO_2 strongly, oxidation of this site does not dramatically alter its ability to adsorb and activate CO_2. This is an important observation since it shows that surface Zr sites on doped Cu NPs can adsorb and activate CO_2 no matter their degree of oxidation.

2.3.3 CuZr/SiC Synthesis and Characterization

To qualitatively verify the computational observations that the presence of surface Zr on Cu NPs results in strong adsorption of CO_2, we synthesized CuZr bimetallic catalysts and evaluated CO_2 adsorption via TPD experiments. The catalysts were prepared using a simple wet impregnation approach, and were thoroughly characterized first to confirm the presence of a mixed

CuZr phase and Zr sites on the surface. After evaluation of various support materials (used to minimize sintering of NPs during experimentation), SiC was chosen as support structure due to the fact that it shows no ability to adsorb CO_2 (unlike common catalyst supports, such as SiO_2 and Al_2O_3, see Figure B.1), and hence allows unambiguous attribution of any observed CO_2 adsorption to the supported CuZr NPs.

STEM-EDS was used to obtain localized elemental information. Figure 2.10 shows a typical TEM image (left) and STEM-EDS spectra of select spots of CuZr/SiC (right). Two different particle morphologies are observed: A small number of larger chunky particles with diameters between 50-100 nm (dark), and agglomerates of smaller NPs with diameters of ~10 nm. The large particles were identified as pure Cu in EDS, while the smaller particles are composed of both Cu and Zr. EDS mapping shows a fairly homogeneous distribution of Cu and Zr in the selected region, indicating the presence of small, bimetallic NPs. The difficulty in forming a uniform catalyst phase likely reflects the computationally observed high segregation energies of Cu and Zr (Figure 2.3).
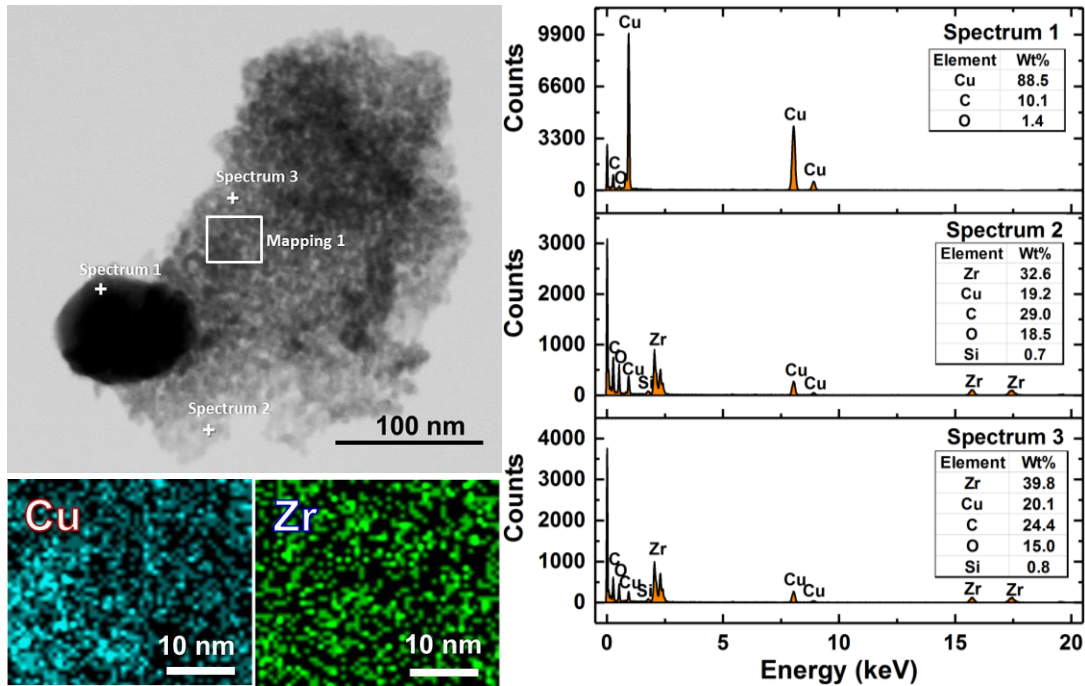
Figure 2.10: STEM-EDS mapping of selected region and STEM-EDS spectra of selected spots of CuZr/SiC. Insert tables list the elemental composition of the selected spots.

To further identify the structure of the CuZr phase, high resolution TEM was used. The presence of clear lattice fringes indicates good crystallinity of the NPs (Figure 2.11) and measurement of the lattice parameter allows identification of the NPs as pure ZrO_2 (0.296 nm and 0.297 nm for monoclinic and tetragonal ZrO_2 (111)[135, 136]; see Figure 2.11(a) and Figure 2.11 (b), respectively) and as $CuZrO_3$ mixed oxide particles (0.314 nm, for $CuZrO_3$ (211)[137]; Figure 2.11 (a)).

28
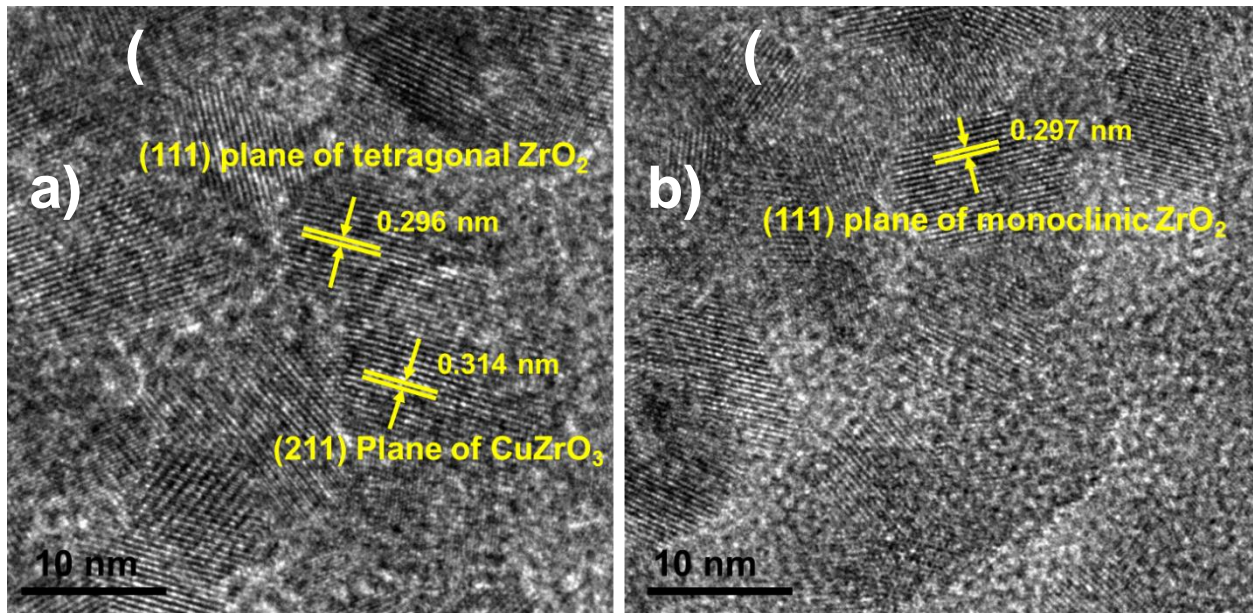
Figure 2.11: HRTEM images of CuZr/SiC with measured lattice spacing, confirming the presence of CuZrO$_3$, m-ZrO$_2$ and t-ZrO$_2$ nanocrystallites.

The presence of both pure Cu and ZrO$_2$ phases as well as a mixed CuZr oxide phase is further confirmed via XRD (see Figure 2.12). For comparison, the XRD pattern of ZrO$_2$/SiC is also shown in this figure, in which monoclinic ZrO$_2$ (m-ZrO$_2$, $2\theta = 28.14^\circ$, 31.47°, 49.26°, 50.08°, JCPDS number ICSD 00-037-1484) is detected as the dominant phase. A noticeable shift in the ZrO$_2$ peak at ~28° from 28.14° for pure ZrO$_2$/SiC to 28.21° indicates the presence of a CuZrO$_3$ phase in the CuZr/SiC sample (JCPDS number ICSD 00-043-0953).
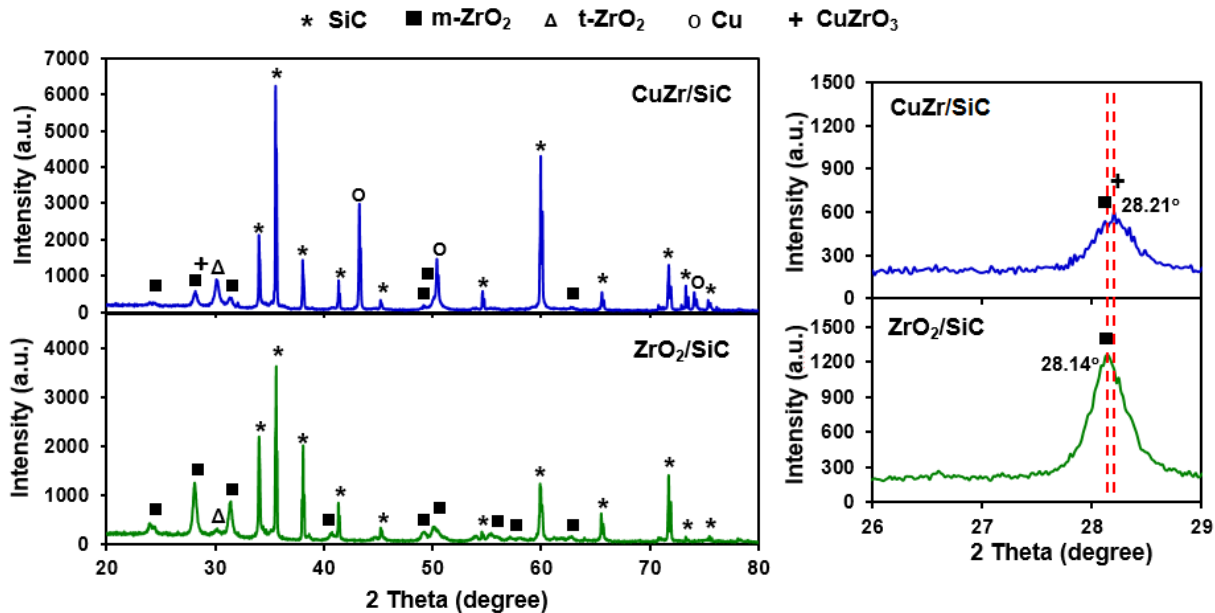
Figure 2.12: XRD Patterns of CuZr/SiC and ZrO₂/SiC. The zoom region shows peak shift from 28.14°, to 28.21°, indicating the presence of CuZrO₃ phase.

The catalyst characterization hence shows presence of both phase-separated pure Cu and ZrO$_2$ particles, as expected from the tendency of Cu and Zr to phase segregate, and also confirms the presence of a bimetallic CuZrO$_3$ phase. It is furthermore noteworthy that Cu in the mixed oxide NPs apparently cannot be reduced to its metallic form at the experimental conditions.
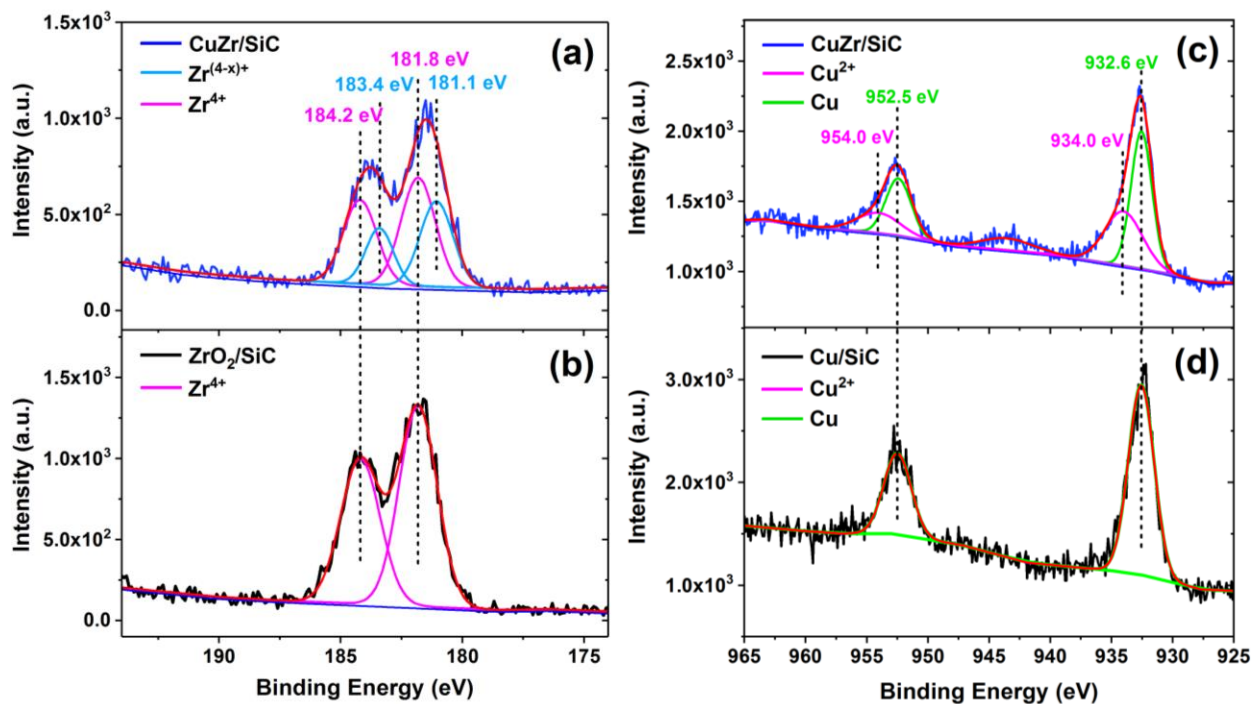
Figure 2.13: High-resolution Zr 3d XPS spectra of (a) CuZr/SiC and (b) ZrO₂/SiC. Two sets of BE peaks were observed, confirming the presence of Zr⁴⁺ and Zr⁽⁴⁻ˣ⁾⁺ species on sample surface. High-resolution Cu 2p XPS spectra of (c) CuZr/SiC, (d) Cu/SiC. The emergence of Cu²⁺ peaks indicates the presence of CuZrO₃ species.

Next, the oxidation states of Cu and Zr in the CuZr/SiC samples were analyzed via XPS after pre-reduction in H_2 at 523K. As shown in Figure 2.13 (a), the high resolution Zr 3d XPS spectrum of CuZr/SiC shows two peaks at 181.8 eV and 184.2 eV, corresponding to $3d_{5/2}$ and $3d_{3/2}$ of Zr^{4+}, respectively, in close agreement with the Zr 3d peaks of the pure ZrO_2/SiC sample (Figure 2.13 (b)). Peak deconvolution reveals another set of peaks in the Zr 3d XPS spectrum of CuZr/SiC at slightly lower energies (181.1 eV and 183.4 eV, respectively), indicating the presence of partially reduced $Zr^{(4-x)+}$ species [138]. However, no metallic Zr (~178 eV) was detected, confirming that all available Zr sites on the sample surface are present in an oxidized state. This is consistent with our computational results (see Table 2.2), which indicate that Zr exists in an oxidized state in

the bimetallic NPs due to its high oxophilicity, with Zr being positively charged (see Zr charge range of +1.1 to +2.3 in Table 2.2).

As to the Cu oxidation state, the Cu 3p spectrum (Figure 2.13 (c)) shows two peaks at 932.6 eV and 952.5 eV which agree closely with the Cu 3p the peaks in the pure Cu/SiC sample (Figure 2.13 (d)), confirming the presence of metallic Cu^0. However, a second set of peaks at 934.0 eV and 954.0 eV indicate the simultaneous presence of some Cu^{2+}, which is consistent with the observation of $CuZrO_3$ nanocrystallites in XRD and HRTEM.
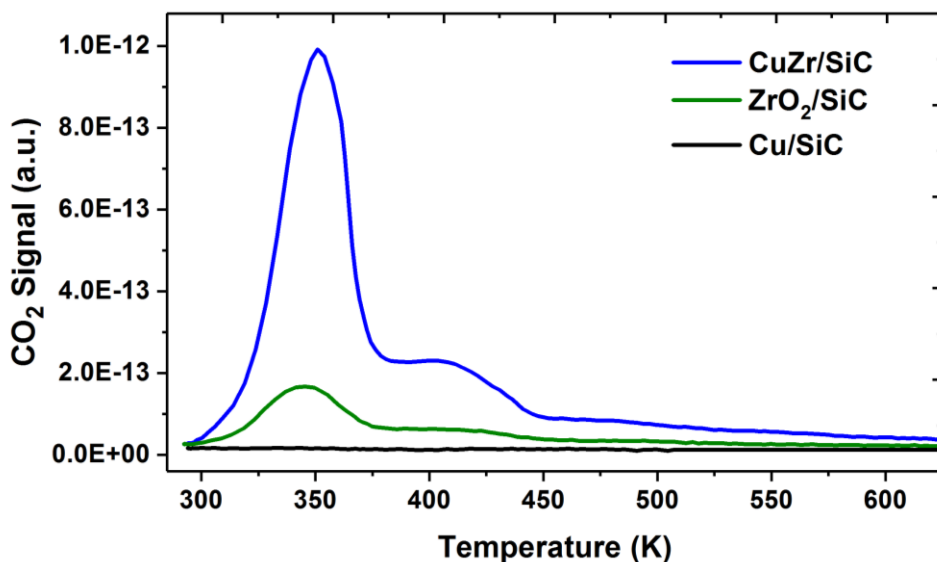
Figure 2.14: CO_2 TPD profiles of CuZr/SiC, Cu/SiC and ZrO_2/SiC.

Finally, we performed CO_2 TPD experiments to analyze the CO_2 BE on the CuZr/SiC catalysts.

Figure 2.14 shows CO_2 desorption profiles obtained from CuZr/SiC along with the two reference materials Cu/SiC and ZrO_2/SiC. The absence of any CO_2 desorption from the pure Cu

sample implies little to no CO_2 adsorption on this sample, i.e. very weak interactions between CO_2 and Cu. This has been well studied in the past and our finding is furthermore consistent with the computational observation that CO_2 can only physisorb on metallic Cu [139, 140]. In contrast, both the bimetallic CuZr and the pure ZrO_2 samples show desorption peaks. Comparing these two samples, we observe a significant increase in CO_2 adsorption capacity for the bimetallic, reflecting the smaller particle size in this sample (14.7 nm for the bimetallic CuZr sample vs 23.9 nm for the ZrO_2 sample, as determined from XRD).

Deconvolution of the CO_2 TPD spectra (Figure 2.15) allows further analysis of adsorption sites and quantification of the respective BEs (using a standard Redhead analysis[141]). The results are summarized in Table 2.3. In addition to two desorption peaks contributed by pure ZrO_2 at 343.59 K and 398.72 K (which is present in both samples), additional nearby peaks at 356.63 K and 424.93 K for the CuZr/SiC sample indicate the presence of stronger CO_2 binding sites on the mixed oxide phase. Previous CO_2 TPD studies on ZrO_2 materials have shown the formation of bicarbonate and carbonate species upon CO_2 adsorption due to the interaction of CO_2 with surface hydroxyl groups and with Zr^{4+}-O^{2-} pairs, respectively.[142-144] Bell and coworkers [143] investigated CO_2 TPD by infrared spectroscopy and found that bicarbonate species decompose more readily at lower desorption temperature followed by carbonate species decomposition at higher temperature. In both cases, CO_2 is the only product formed. We hence assign site 1 and 2 to the decomposition of bicarbonate species, while site 3 and 4 to the decomposition of carbonate species. To further support this assignment, we performed CO_2 adsorption calculations on a ZrO_2 ($\bar{1}11$) surface in the presence and absence of surface hydroxyls (see Figure B.2 – Figure B.3). Our calculations suggest a barrierless adsorption of CO_2 as bicarbonate species (on surface hydroxyls present from dissociated water on the ZrO_2 surface) with a desorption energy of 0.57 eV (Figure B.4) and a CO_2

adsorption to carbonate with an energetic barrier of 0.13 eV (from physisorbed to chemisorbed) and an overall desorption barrier of 0.81 eV, respectively (Figure B.5). TPD experiments on ZrO_2/SiC sample suggest that CO_2 desorption energy on site 1 and site 3 are 0.99 eV and 1.15 eV, respectively, which is qualitatively consistent with the computational observations, i.e. with CO_2 being desorbed from bicarbonate species at significantly lower energies than from carbonate.
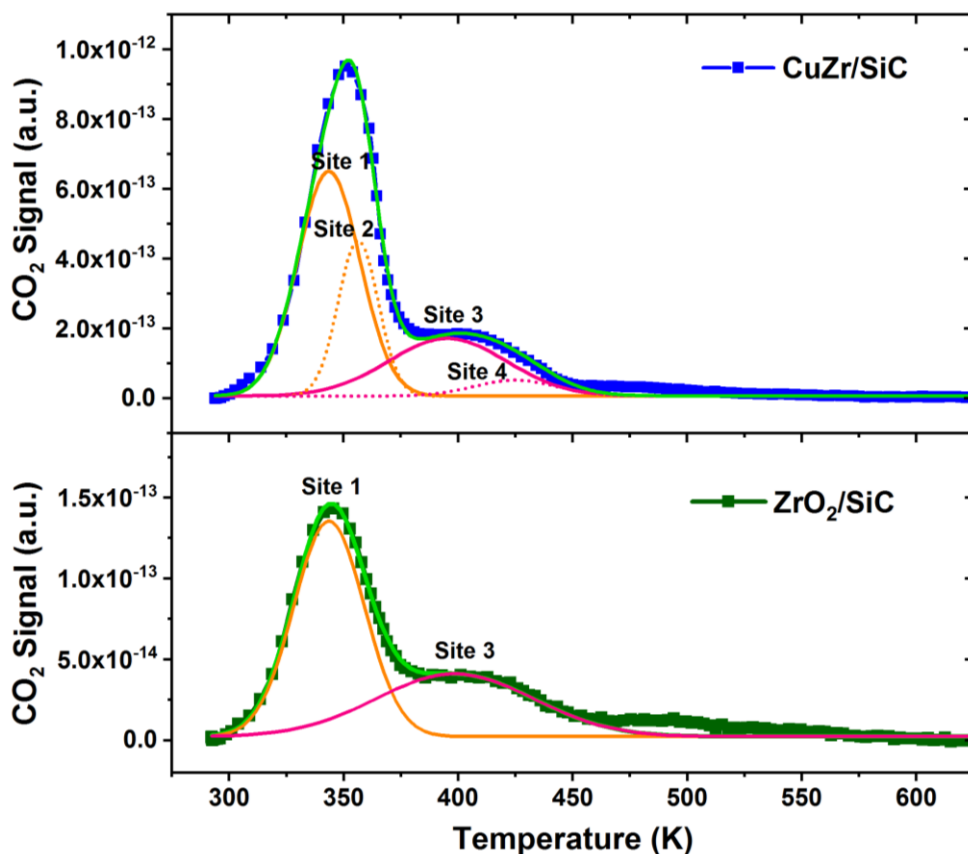
Figure 2.15. Deconvoluted CO_2 TPD spectra over CuZr/SiC and ZrO_2/SiC. The light green lines are the cumulative fit curves for deconvoluted peaks.

Table 2.3. Summary of CO₂ adsorption sites on CuZr/SiC and ZrO₂/SiC, including desorption temperature (T_{des}, K) and the corresponding desorption energy (E_{des}, eV) calculated using Redhead equation.

Sample	Site 1		Site 2		Site 3		Site 4	
	T_{des} (K)	E_{des}(eV)	T_{des} (K)	E_{des}(eV)	T_{des} (K)	E_{des}(eV)	T_{des} (K)	E_{des}(eV)
ZrO_2/SiC	343.59	0.99	/	/	398.72	1.15	/	/
CuZr/SiC	343.59	0.99	356.63	1.03	398.72	1.15	424.93	1.23

Finally, comparing the calculated CO_2 desorption energy from the carbonate species formed on ZrO_2 (0.81eV) with the ones formed on the surface of the $Cu_{54}ZrO_x$ with x=2, 4 (1.67 and 1.37 eV in Figure 2.9 (c) and (d) and Table 2.2), we observe that CO_2 binds stronger on the mixed CuZr oxide than on ZrO_2. This is again in qualitative agreement with the TPD desorption energies which show a shift towards higher energies (stronger adsorption) on the mixed oxide (Site 4 in Table 2.3) compared to pure ZrO_2 (Site 3 in Table 2.3) We acknowledge that although the CO_2 adsorption trends between our calculations and the TPD experiments are in qualitative agreement, there is not a quantitative agreement on the adsorption strength. This is likely due to the differences between the (highly heterogeneous) actual sites on the synthesized catalyst surface compared to the simulated ones in the computational framework. Nevertheless, our experiments qualitatively confirm all major computational results, i.e.: i) Cu NPs do not adsorb CO_2, ii) Cu and Zr metals have a strong tendency to segregate, iii) Zr has high propensity to oxidize, and iv) even when the Zr sites are oxidized (forming a mixed CuZr oxide), they are still able to effectively adsorb CO_2 with an adsorption energy that exceeds that of pure zirconia.

2.4 Conclusions

A DFT study was carried out on various singly-doped bimetallic icosahedral $Cu_{54}M$ (M= Au, Cu, Mn, Mo, Ni, Pd, Rh, Ru, Sc, V, Zn, or Zr) NPs in order to identify potential Cu-based bimetallic catalysts that adsorb and activate CO_2 and unravel the underlying physical properties governing CO_2 activation. Dopants were initially screened based upon their preference to occupy surface sites on the NP. We identified Au, Cu, Mo, Pd, Sc, Zn, and Zr as dopants that prefer to reside on the surface of Cu NPs and we further calculated the CO_2 adsorption on these surface sites. Based on the adsorption calculations, we have developed a set of criteria which are necessary for the favorable chemisorption of CO_2 onto $Cu_{54}M$ NPs: (i) the local d_c of the dopant must be higher and close in energy relative to the LUMO of CO_2, and (ii) the dopant atom at the binding site must have an IP and electronegativity lower than that of Cu.

Based on these criteria, we identified $Cu_{54}Zr$ as a catalyst that binds strongly and activates the CO_2 molecule. Due to the highly exothermic adsorption of CO_2 on Zr-doped Cu NPs relative to the other dopants and the oxophilicity of Zr, we further investigated the CO_2 adsorption on $Cu_{54}ZrO_X$ (x = 2, 4) NPs, where the surface Zr has been oxidized. Chemisorption of CO_2 onto these oxidized sites was observed in every case, with a slight decrease in the CO_2 BE compared to $Cu_{54}Zr$, which was attributed to the shift in the local d_C of the oxidized metal.

These computational observations were qualitatively confirmed through experiments using a mixed CuZr catalyst. The catalysts showed the presence of both pure (i.e. phase segregated) Cu and ZrO_2 phases (in agreement with the computational prediction that Cu and Zr have a high segregation energy) as well as a mixed $CuZrO_3$ phase. While pure Cu was not able to adsorb any CO_2, the mixed CuZr oxide phase showed strong CO_2 adsorption, exceeding that of the pure ZrO_2 phase.

Overall, our computational and experimental results suggest that Zr-doped Cu NPs can strongly adsorb and activate CO_2, even if the surface Zr sites become oxidized. The resulting bimetallic system could hence be a promising material for CO_2 utilization via hydrogenation. Further studies to confirm this prediction are currently on-going.

3.0 Unfolding Adsorption on Metal Nanoparticles: Connecting Stability with Catalysis

The content of this chapter is adapted from Dean, J.; Taylor, M. G.; Mpourmpakis, G. "Unfolding Adsorption on Metal Nanoparticles: Connecting Stability with Catalysis." Science Advances 2019, 5 (9) eaax5101.

To supplement our work in identifying materials properties relevant to CO_2 adsorption, we now proceed into the development of more generalizable models of adsorption. In this section, we develop a set of universal adsorption descriptors (CE_{local}, MADs, and IPEA), and demonstrate their effectiveness by creating linear models of adsorption capable of describing the adsorption of CH_3, CO, and OH to a variety of metal systems, including both nanoparticles and slabs of Cu, Ag, and Au, slabs of Ni, Pd, Pt, Rh, and Ir, and even a bimetallic AgCu NP.

3.1 Computational Methods

3.1.1 Nanoparticle Adsorption Calculations

All adsorbate-NP BEs were calculated with DFT, using Quickstep [117] as implemented in CP2K [118]. Exchange-correlation was accounted using the PBE functional. The DZVP basis set [121] was used with the GTH pseudopotentials [145] at a 500 Ry cutoff. SCF cycles were performed with a convergence criterion of 10^{-7} Ha. Geometry relaxations were converged to forces below 0.02 eV/Å. Visual depictions of the NPs with labeled CNs can be found in the Appendix C.6, Figure C.4 (A-E), and depictions of the top adsorption configurations studied can be found in

Figure C.4 (F-I). DFT-calculated BEs for each NP-adsorbate pair we investigated can be found in Table C.6.

3.1.2 Single-Metal Adsorption Calculations

To calculate MADs we used the molecular structures of the complexes between a single metal and the adsorbate as illustrated in Figure 3.2. The energy for these complexes were evaluated using the PBE functional with the def2-SV(P) basis set [146] and the RI/MARIJ approximations as implemented in the Turbomole package [147, 148]. Each structure was relaxed using a quasi-Newton-Raphson method for all multiplicities at or lower than septet (to account for spin state of each complex) and the lowest-energy multiplicity was used to calculate the MADs in Equation 3.1 as:

$$MADS = E_{bind,M-Ads} = E_{M-Ads\ Complex} - E_M - E_{Ads} \qquad 3.1$$

Where E_X represents the electronic energy of species X, M is a metal, and Ads is the adsorbate. Visual depictions of the gas-phase metal-adsorbate binding description can be found in Figure C.4 (J-M). Calculated values for MADs can be found in Table C.6.

3.2 Results and Discussion

3.2.1 Feature Identification

We expect several factors to contribute to the adsorption interaction between a site on a monometallic NP and an adsorbate, based upon prior literature [23, 149]. We hypothesize that the most important factors are 1) the stability of the adsorbate in the gas phase [150], 2) the tendency of the

metal and adsorbate to interact and form chemical bonds, 3) the stability of the bare NP [149], and 4) the stability of the binding site [23]. We use stability here to express the thermodynamic strength of the bonds that are on NPs of different sizes/shapes/composition, or of adsorbates at different metals and sites on the NPs. As a result, we refer to energy differences (thermodynamics) arising from different bond formations (e.g. metal-metal, metal-adsorbate). The rationale behind our hypothesis can be found in the energetic terms used in the calculation of a BE in DFT, which is the difference between the electronic energy of the adsorbate-metal complex (E_{Ads-M}) and that of the adsorbate (E_{Ads}) and the metal NP or surface (E_M). For instance, E_{Ads} is related to the stability of the adsorbate, E_{NP} to the stability of the NP, and E_{Ads-M} to the strength of adsorbate metal interactions and the local energy landscape of the adsorption site, i.e. the stability of the site. Using these physical descriptors, we represent the BE generally as:

$$E_{binding,true} = \hat{f}(\text{Stab}_{Ads}, \text{Stab}_{NP}, \text{Int}_{Ads-M}, \text{Stab}_{Site}) + \epsilon \qquad 3.2$$

Where $E_{binding, true}$ refers to the true BE of the adsorbate to a specific site on a NP. In this work, we take the true binding interaction to be the DFT BE (Equation 3.1). Stab_{Ads}, Stab_{NP}, and Stab_{site} refer to the stability descriptors for the adsorbate, NP, and adsorption site respectively. Finally, Int_{Ads-M} refers to a descriptor for the strength of adsorbate-metal interactions. Since it is unreasonable to expect a simple linear model to capture all the physics of adsorption, we acknowledge this limitation with an irreducible error term, ϵ. In this work, we consider our chosen physical descriptors (Stab_{ads}, Stab_{NP}, Int_{Ads-M}, and Stab_{site}) to all be in units of energy, thus forming a first-order polynomial relation with the BE ($E_{binding, true}$). As a consequence, we do not investigate cross-terms between these descriptors.

3.2.1.1 Defining a Local Cohesive Energy

Intimately related to the stability of the metal and the strength of metal-metal bonds is the CE of the bulk metal, which is defined as the amount of energy required for the atoms in the system to achieve infinite separation. CE is captured in NPs by the BCM of Yan et al [151], which asserts CE as the sum of every metal-metal bond energy in the NP. In this work, we apply the same concept to describe the stability of binding sites, with the justification that chemically unsaturated sites (i.e. less stable due to having fewer metal-metal bonds) tend to bind adsorbates stronger. To this end, we introduce the local CE descriptor, CE_{local}, defined by applying the BCM only to the metal atom participating directly in the binding interaction. In other words, we take the summation of BC-model-approximated bond energies connecting with the metal atom in the binding site, which is defined as:

$$
\begin{aligned}
Stab_{site} \approx CE_{local} &= \sum_{i=1}^{CN} E_{bond\ A-B_i} \\
&= \sum_{i=1}^{CN} \left(\frac{\gamma_A CE_{bulk,A}}{CN_A} \sqrt{\frac{CN_A}{CN_{bulk,A}}} + \frac{\gamma_{B_i} CE_{bulk,B_i}}{CN_{B_i}} \sqrt{\frac{CN_{B_i}}{CN_{bulk,B_i}}} \right)
\end{aligned}
$$

3.3

In Equation 3.3, A-B_i refer to the atoms in the neighborhood of the binding site. Atom A refers to the metal atom directly participating in the top-site adsorption. Atom B_i is one of the metal atoms directly bound to atom A. The summation index i refers to the bonds between atom A and atoms B_i, and ranges between 1 and the CN of atom A. As an example, an adsorbate bonding to a site with CN 6 would have 6 bonds, meaning the summation would have 6 terms (bonds A-B_1, A-B_2, …, A-B_6). Each bond energy E_{bond} can be approximated with the BCM, which we directly substitute inside the summation parentheses in Equation 3.3. The other terms in the equation can be described as follows: $CE_{bulk\ X}$ is the bulk CE of atom X, CN_X is the CN of atom X, $CN_{bulk\ X}$ is

the bulk CN of atom X, and γ_X is the BCM gamma coefficient of atom X [151]. This definition of Stab$_{site}$ (Equation 3.3) allows us to capture not only the CN and metal identity of the atom directly bound to the adsorbate, but also the electronic effects that come from the geometry of the local site, such as stronger bonds resulting from under-coordinated atoms adjacent to the binding atom. This formulation of CE$_{local}$ additionally differentiates between atoms of the same CN as well as different metal types.

Using this description of the local geometry, we can focus on its ability to describe the binding of a single adsorbate-metal pair. In Figure 3.1 A, we plot the DFT-calculated top-site BE of CO to a 172-atom Au cube, and a 147-atom Au cuboctahedron/icosahedron. We note a strongly inverse relationship between CE$_{local}$ and BE. In addition, we visualize both the BE (Figure 3.1 A – Figure 3.1 D) and CE$_{local}$ (Figure 3.1 E – Figure 3.1 G) that the different surface sites on the NPs exhibit. Overall, Figure 3.1 demonstrates that the strongest adsorption sites are the ones exhibiting the weakest local cohesion.
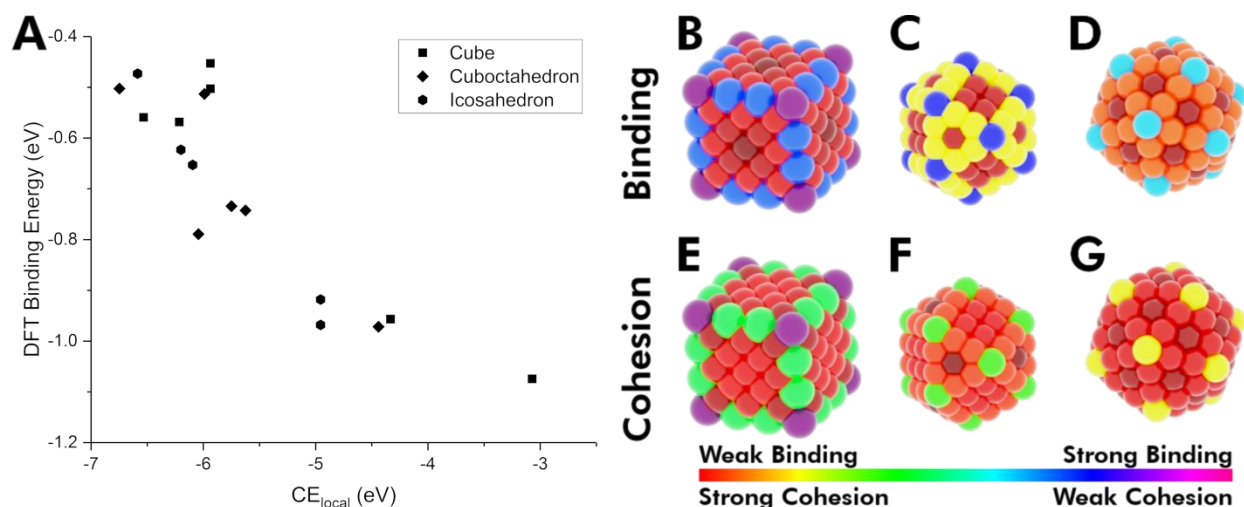
Figure 3.1: The BE of CO on various sites of Au NPs as a function of CElocal: 172-atom cube (rectangles), 147-atom icosahedron (hexagons) and 147-atom cuboctahedron (rhombs). Heat map of different sites on the NPs with respect to their BE of CO (B)-(D) and to their CElocal (E)-(G) . The color scheme follows the range: strongest CO binding / weakest CElocal are colored in violet, and the weakest binding / strongest CElocal are colored in red.

3.2.1.2 Additional Adsorption Descriptors

Other descriptors chosen for the NP binding were i) the CE of the entire NP, ii) the metal atom binding to the adsorbate and iii) a term involving the IP and EA of the adsorbate. The total NP CE was chosen because we expect the stability (or lack thereof) of a given NP to play cumulatively a role in its ability to bind a molecule [152]. The gas phase BE between a single metal atom and the adsorbate was chosen because we expect to provide a good (and fast to calculate) descriptor of the tendency of the adsorbate to bind a specific metal. Finally, the negative average of the IP and EA, which we call "IPEA" was specifically chosen since it has been shown to be a first-order finite difference approximation of the adsorbate's chemical potential within HSAB theory [153]. In addition to this, it is also the negative of Mulliken's definition of electronegativity

43

[153]. The values of each of these descriptors can be found in Table C.1 – Table C.3, along with more detailed justifications. It is worth noting that the adsorbate-metal atom BE is the only DFT calculation required to parameterize the model (once it is trained), which is computationally inexpensive to perform. Furthermore, all other quantities (CE_{local}, CE_{NP}, and IPEA) can be rapidly determined through a combination of simple algebra and tabulated physical properties.

3.2.2 A Universal Model of Adsorption

Following our choice of chemical descriptors, we conducted linear regression utilizing the Caret package [154] as implemented in R [155], performing OLS regression in conjunction with 10-fold CV for Equation 3.4 (which is the first-order polynomial equation we define in our discussion of Equation 3.2).

$$E_{bind,model} = a + b * CE_{local} + c * CE_{NP} + d * IPEA + e * MADs \qquad 3.4$$

Where $E_{bind, model}$ is the model's predicted BE, and a/b/c/d/e are constants. CE_{local} is the local site's cohesion. CE_{NP} is the CE of the whole NP. IPEA is the negative average of the IP and EA. MADs is the Metal-ADsorbate BE.

Utilizing this model, we perform an OLS regression using as a training set the top-site adsorption of three different adsorbates (i.e. methyl radical (CH_3), carbon monoxide (CO), and hydroxyl radical (OH)), onto three different metals (Cu, Ag, and Au) with five different NP morphologies (172-atom cube, 55/147-atom icosahedron, 55/147-atom cuboctahedron) as shown in Figure 3.2 A. Although we only focus on adsorption on top sites (Figure C.4 F-I), we note that previous work [156] has shown that the top site BE of various adsorbates (including methyl/hydroxyl

radicals) correlates with other surface site BEs (e.g. bridge and hollow sites) over different metals

(including Cu/Ag/Au). Regression statistics can be found on Table 3.1 (case (i)).

Table 3.1: OLS regression information for (i) 4-descriptor model that includes CElocal, IPEA, MADS and CE$_{NP}$, (ii) 3-descriptor model that excludes CE$_{NP}$, and (iii) equivalent 3-descriptor model utilizing the slab dataset found in the literature. [157]. All cases are trained using datasets where CH$_3$, CO, or OH adsorb to Cu, Ag, or Au.

(i) Trained on NPs, 4-Descriptor Model
RMSE: 0.179 eV, MAE: 0.145 eV, R^2: 0.936,
Maximum Error: 0.619 eV, Remaining DOF: 157

	Coefficient Estimate	Standard Error	p-Value
Intercept	1.51477	0.15876	$<2*10^{-16}$
CE$_{local}$	-0.1450	0.01663	$3.85*10^{-15}$
IPEA	0.33171	0.01280	$<2*10^{-16}$
MADS	0.67858	0.01522	$<2*10^{-16}$
CE$_{NP}$	-0.0002	0.05388	0.998

(ii) Trained on NPs, 3- Descriptor Model
RMSE: 0.179 eV, MAE: 0.144 eV, R^2: 0.933,
Maximum Error: 0.619 eV, Remaining DOF: 158

	Coefficient Estimate	Standard Error	p-Value
Intercept	1.51509	0.12148	$<2*10^{-16}$
CE$_{local}$	-0.14502	0.01410	$<2*10^{-16}$
IPEA	0.33171	0.01274	$<2*10^{-16}$
MADS	0.67857	0.01501	$<2*10^{-16}$

(iii) Trained on Slab dataset, 3-Descriptor Model
RMSE: 0.122 eV, MAE 0.102 eV, R^2: 0.979,
Maximum Error: 0.259 eV, Remaining DOF: 113

	Coefficient Estimate	Standard Error	p-Value
Intercept	1.67677	0.09220	$<2*10^{-16}$
CE$_{local}$	-0.14590	0.01079	$<2*10^{-16}$
IPEA	0.28743	0.01005	$<2*10^{-16}$
MADS	0.79516	0.01187	$<2*10^{-16}$

In the final functional form of the model, the coefficients regressed make excellent physical sense (Table 3.1). A negative sign on the CE_{local} coefficient suggests that as the cohesion of the local site increases (becomes more negative) its binding affinity to the adsorbates decreases. Similarly, a positive sign on IPEA, which is related to the adsorbate's chemical potential, denotes that a less stable adsorbate (higher chemical potential) will have a higher tendency to bind to a metal. Finally, the direct correlation with MADS, the BE between the adsorbate and a single metal atom in the gas phase, is also very intuitive as it describes the intrinsic tendency of the metal to bind the adsorbate with a more negative MADS indicating a stronger binding interaction.

For this model (case (i) of Table 3.1) we observed a CV R^2 of 0.936 and a CV RMSE of 0.179 eV. Of note here is the high p-value and error associated with the CE of the NP. As a result, we conducted a second regression where we did not include CE_{NP}, but instead we only included the remaining three regressors and an intercept. This model is plotted in Figure 3.2 B, and its regression information can be found in the middle panel of Table 3.1. We see that the model (case (ii) of Table 3.1) fits nearly as well, with an overall model R^2 of 0.933 and a CV RMSE of 0.179 eV. If we additionally plot the BE as a function of NP cohesion (Figure C.1), we can see that, at least in the case of our chosen adsorbate-NP combinations, it does not provide an adequate descriptor for the binding interaction. Because the inclusion of the NP CE offers no improvement over a less-complex model, we decide to ignore this term and use the form outlined in Equation 3.5 for the remainder of our investigation:

$$E_{bind,model} = a + b * CE_{local} + c * IPEA + d * MADs \qquad 3.5$$

The p-value, CV RMSE, and CV R^2 all indicate that our model is generalizable to other metal-adsorbate systems. Additionally, the remaining DOF, calculated as the dimensionality of the dataset minus the dimensionality of the regression equation, for every model in Table 3.1 are large,

indicating overfitting is unlikely. Overall, this shows that our model's predictions correlate well with the DFT-calculated values. This is important, because we ultimately wish to provide a computationally-efficient framework for predicting DFT adsorption energies. The error metrics, RMSE, MAE, and R^2 offer a picture of how likely the model is to fail/succeed compared to DFT. The CV RMSE and MAE have values of 0.179 and 0.144 eV respectively, which indicates that, on average, the model can be expected to be within 0.144-0.179 eV of the true DFT-calculated value. Finally, the good R^2 values of 0.933 and 0.936 indicate that our model is strongly correlated with the actual DFT predictions.

We tested our model's generalizability by splitting our data into training/test sets based upon adsorbate, metal identity, and morphology. In these series of "leave-one-in" tests, we restrict the training set to only a single morphology, adsorbate, or metal, and then test the developed model on the data that was left out. The results of these tests can be found in Table C.4 and Figure C.2. Overall, they provide strong evidence that our model is capturing the underlying physics of the binding interaction: when we train on a single metal or morphology, we still capture the other metals or morphologies with good accuracy. In the case of training on a single adsorbate, we see that although the model fits are worse, they still capture the trends of the binding interactions (e.g. the adsorbates that were not tested are off parity by a constant amount). This is because the model has only a single possible value in the case of the IPEA term. Therefore, it reduces to a model which has only three terms, which are the intercept, CE_{local}, and MADs. However, this indicates that our model is highly robust. Even when the adsorbate descriptor is missing, the model captures trends in the binding interactions. Therefore, our CE_{local} and MADs descriptors may be applicable to a wide array of adsorbates and NPs.
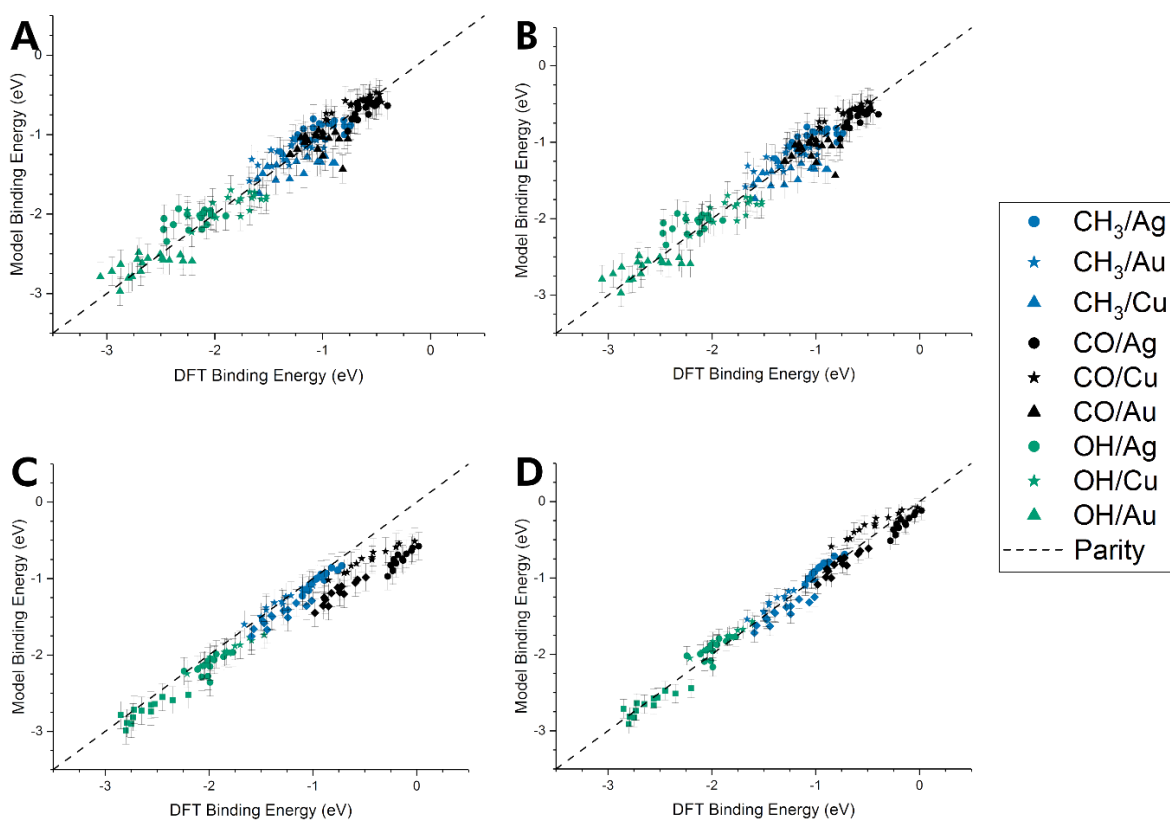
Figure 3.2: Parity plot of the model-predicted BE of adsorbates (OH, CO, and CH3) on various metal systems versus the DFT BE (eV). (A) The model both trained and tested on PBE DFT data for NPs (Au/Ag/Cu, 55-172 atoms) which includes the CENP term. This model corresponds to case (i) of Table 3.1 (B) The model both trained and tested on PBE DFT data for NPs (Au/Ag/Cu, 55-172 atoms) which does not include the CENP term. This model corresponds to case (ii) of Table 3.1. (C) The model trained on PBE DFT data for NPs (Au/Ag/Cu, 55-172 atoms) and tested against RPBE DFT data for top-site adsorptions on metal surfaces (Au/Ag/Cu) from the literature slab dataset. [157] This model corresponds to case (ii) of Table 3.1. (D) The model both trained and tested with RPBE DFT data for top-site adsorptions on metal surfaces (Au/Ag/Cu) from the slab dataset. This model corresponds to case (iii) of Table 3.1. In all cases, error bars are determined from the 10CV RMSE on the training set. Our DFT calculated BE of the different adsorbates on the various sites of the metal NPs are shown in Table C.6.

49

In addition to nonperiodic NP systems, we also investigated periodic (slab) systems. Recently, Roling et al [157] developed several scaling relations for adsorption on metal slabs, and alongside these relations reported a large body of DFT calculations describing adsorption of CH_3, CO, and OH on the surface of several metal slabs of Ag, Au, Cu, Ir, Ni, Pd, Pt, and Rh. Using the scaling relations, they derive a model which describes metal-adsorbate BEs to a high degree of accuracy, but is parameterized by several DFT calculations. In their model, the metal-adsorbate BE is asserted as a function of the DFT-calculated BE of the metal atom to the surface, and of the gas-phase BE of an adsorbate to a metal atom (which we call MADs in this work). Although the gas-phase metal-adsorbate BE needs only to be parameterized once per metal-adsorbate system, calculating the metal's BE to its surface needs to be done once for every potential binding site, which requires a DFT optimization for the slab both with and without the metal atom. In scenarios where adsorption energies across many different sites are to be investigated, the metal-surface BE term becomes a limiting step, and may prove impossible for larger NPs (or slabs) that are infeasible to be calculated with DFT. Hence our model, which does not need to be re-parameterized for every binding site, allows for significantly higher throughput.

Here, we utilize the RPBE-calculated top-site adsorption energies on slabs reported by Roling et al [157], and focus on the Cu, Ag, and Au slabs adsorbing CH_3, CO, and OH. We choose this particular dataset because it is a recent study encompassing a variety of FCC metal slabs from the d-block (including those we investigate from our own NP calculations), using the same adsorbates present in our calculations. It is also a large dataset, which allows us to have a more accurate assessment of our model's generalizability to these systems. We take our model which has been fit to only our PBE DFT NP data, and use their data as a test set. In Figure 3.2 C, we plot the DFT-calculated BEs of the slabs and compare them with the predictions of our model. We see that

although there is some deviation from parity we still capture the overall trends of the binding interaction. This deviation is most-likely because our training set uses the PBE functional with a Gaussian plane wave basis set and was trained entirely on NPs (no slabs were included in the training set). The slab dataset was produced from calculations with the RPBE functional, and a planewave basis set. This results to the important observation that even though the adsorption data are with different functional, basis set, and adsorption environment (a slab, not an NP), we are still able to predict the DFT BE to a high degree of accuracy, based on our developed model. This is further evidence that our model is capturing the underlying physical trends of the systems. We note, however, that RPBE tends to produce more accurate CO adsorption energies than PBE, which over-binds CO relative to experiments [158]. We observe this over-binding behavior in Figure 3.2 C, where the PBE-trained model predicts CO adsorption energies to be more negative relative to the RPBE DFT data. Finally, to prove that our introduced descriptors and developed model is universal and applicable to periodic systems, we plot in Figure 3.2 D the model's behavior when it is both trained and tested on the slab dataset. In Figure 3.2 D we immediately note that the discrepancy with CO moves back to parity. As expected, this indicates that when the model is parameterized using the same functional used to calculate benchmark BEs the results are more accurate. As it can be seen, all the data impressively fall on the parity line. Overall, this result combined with our leave-one-in tests (Table C.4 and Figure C.2) shows that our model only needs to be fed a small subset of the possible morphologies that a NP or slab may take on, and can then predict binding to other morphologies. The implication is that computationally-inexpensive systems can be used to parameterize the model via DFT, and then it can extend to significantly-larger systems which would otherwise be computationally intractable to investigate.

3.2.2.1 Extension to Bimetallics

So far, our analysis has been entirely based on monometallic systems. Although monometallic systems are important to understand, a much larger and challenging chemical space can be found in the realm of bimetallics. For example, although there are 40 d-block metals (including La and Ac), there are $\frac{40^2 - 40}{2} = 780$ possible unique bimetallic alloys, without considering the additional materials space dimensions of NP size, shape, chemical ordering, etc. If we additionally consider chemical ordering in a bimetallic 55-atom NP (a computationally tractable metal NP size with DFT) of a single arbitrary morphology, there are a maximum of 2^{55} different ways both elements could be arranged in the NP. This number is already very large, and further grows when different morphologies are considered. Therefore, fast and accurate descriptions of bimetallic alloys are a necessity if we are to explore even a fraction of these systems.

All evidence so far indicates that our model is physically sound, and therefore we hypothesize that it should be extendable to bimetallic systems. Moreover, because CE_{local} is a direct extension of the BCM as we describe in Equation 3.3 and takes into account the atomic identity of both the metal atom involved directly in the adsorption and every metal atom that has formed bonds with (first neighbors), this descriptor extends naturally to bimetallic systems. Additionally, IPEA has no dependence on binding site and MADs is only dependent on the atom directly bound to the adsorbate and does not require modification to capture top-site binding in bimetallic systems. Therefore, we continue using the identical physics-based descriptors for bimetallic NPs and slabs (as were used with monometallic systems), investigating top-site adsorption energetics. In Figure 3.3, we plot the BE of the CH_3, CO, and the OH as predicted by our model when trained only on

monometallic NPs against the DFT-calculated binding on several sites of the icosahedral 55-atom

NPs ($Cu_{31}Ag_{24}$ and $Cu_{22}Ag_{33}$).
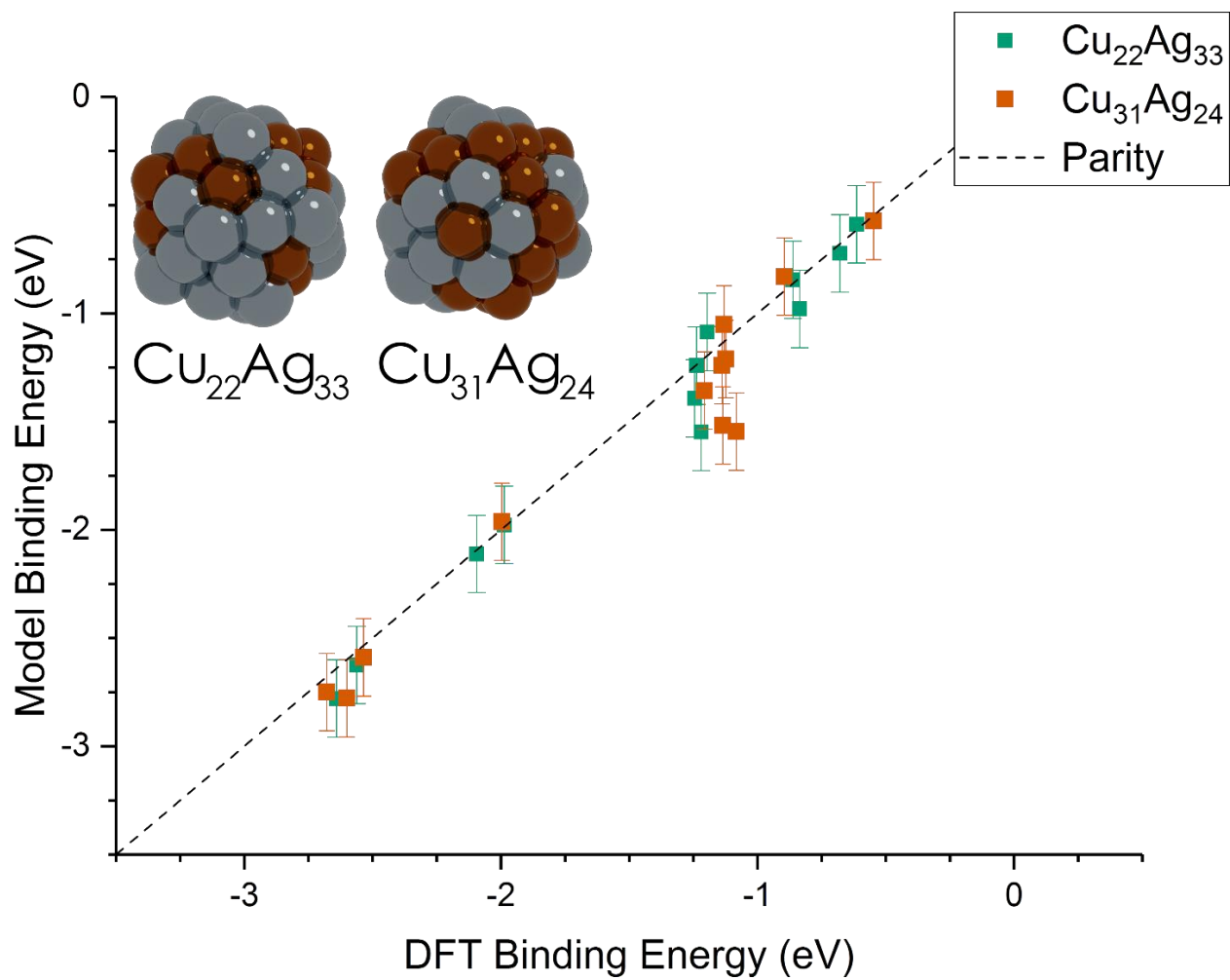
Figure 3.3: Parity plot between our developed model and DFT calculations on icosahedral bimetallic (Cu$_{55-x}$Ag$_x$ x=24,33) NPs. The model is trained on CH3, CO, and OH adsorbing on monometallic Ag, Cu, and Au NPs and is able to capture adsorption on bimetallic NPs. Images of the two NPs are shown as inset, with copper and silver atoms colored brown and grey, respectively.

We can immediately see the model captures trends in adsorption on the bimetallic Cu/Ag NPs very accurately. This further points towards the physical basis of our model, since the model is only trained on monometallic systems and can still accurately capture the binding of the adsorbates on bimetallic NPs (which the model has never seen before). Furthermore, this demonstrates that our model is generalizable and applicable to both monometallic and bimetallic NPs. Additionally, this is also indicative of the good performance of the CE_{local} descriptor. Other descriptors such as the CN, and GCN [159] either do not have obvious extensions to systems of different metals or require extensive modification to capture multimetallic environments [160]. The BCM [151] from which we derive CE_{local}, however, is deliberately formulated to describe bond strength between different metals, and has been shown to perform very well on bimetallic alloys. We note, however, that Cu and Ag have similar electronegativities [161]. Therefore, the tested $Cu_{55-x}Ag_x$ NPs do not show strong charge transfer since our calculations indicate point charges of less than ± 0.1 |e-|. Bimetallic systems that potentially develop significant charge transfer could impact the BE of adsorbates (and affect for example electrocatalytic behavior [162]) and may require an additional descriptors (such as binding site electronegativity) to capture such effects [163].

3.2.2.2 Extension to d7 and d8 Metal Slabs

Having shown that we are able to train on just one d9 metal and capture the adsorption trends of the other d9 metals (Figure C.2), we believe that we should be able to extend our model to other systems. We again utilized the slab dataset [157], in which several d9 (Cu, Ag, and Au), d8 (Ni, Pd, and Pt), and d7 (Rh and Ir) metals were investigated. In Figure 3.4 A we only trained the model on the Roling dataset of CH_3, CO, and OH adsorbed to Cu, Ag, and Au NPs, and show that

it is able to capture the general adsorption trends for the other columns of the periodic tables, although deviations occur on a per-adsorbate and per-metal basis.
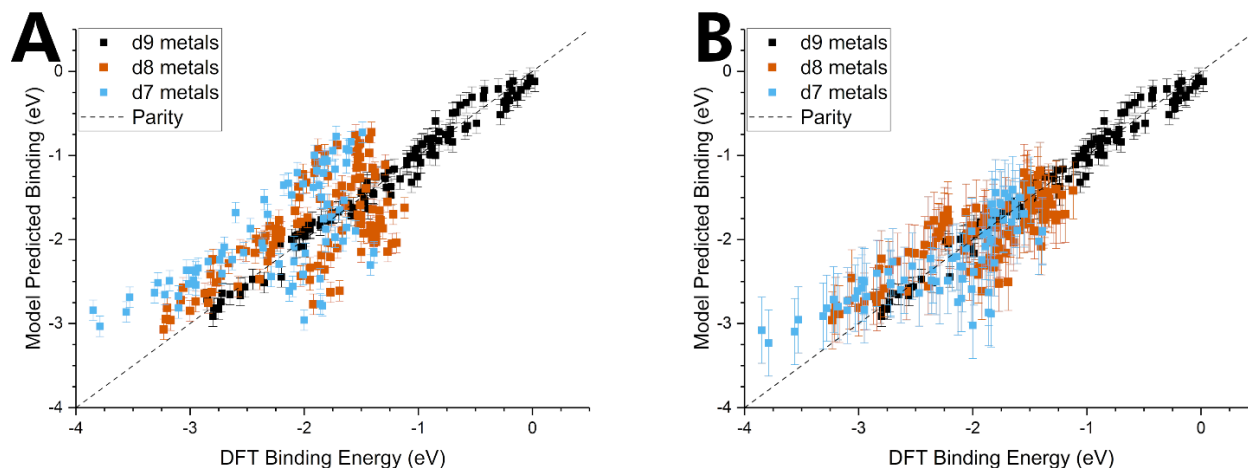
Figure 3.4: Our 3-descriptor model extended to slab dataset [157]. (A) The model trained on the slab dataset [157] on Cu, Ag, and Au surfaces and tested against the Rh, Ir, Ni, Pd, Pt, Cu, Ag, and Au surfaces from the slab dataset. (B) The equivalent model when trained separately for each column of the d-block, still using the slab dataset. Error bars in every case are the 10-fold cross validated RMSE of the training set.

We then expanded our training set to include every periodic system in the slab dataset (CH_3, CO, and OH adsorbed to Rh, Ir, Ni, Pd, Pt, Cu, Ag, or Au slabs), and attempted to fit our 3-descriptor model to it. This resulted in a poor fit: although it captures to some degree of accuracy the different metal-adsorbate pairs, it is unable to capture differences in the binding site of each slab (Figure C.3 A). This suggests that our model is missing some physical descriptor that can accurately differentiate between different columns of the periodic table. Searching for such a descriptor, we then investigated a variety of additional physical descriptors: The CN of the local site, the same first-order approximation of the chemical potential (of the metal), the hardness of

the metal, the d-count of the metal, the covalent radius of the metal, the resistivity of the metal (related to the electronic structure), and the melting point of the metal (related to the strength of bonds). Even allowing the model to overfit by including all 10 of the potential physical descriptors we investigate, we do not see a significant improvement over our original model. We present the summary of our search for an effective descriptor via OLS in Table C.5 and Figure C.3.

We attempted to leverage more complex machine-learning techniques in order to provide additional avenues to improve the adsorption model. Utilizing LASSO (see Appendix C.4), we find the best descriptors to be the CN, the chemical potential of the metal (calculated to the same way as IPEA), the covalent radius of the metal, CE_{local}, IPEA, and MADs. Of these, the covalent radius of the metal and the chemical potential of the metal exhibit low coefficients. Higher importance is given to the other descriptors, which, in the absence of CN (which can be justified by how CE_{local} is a demonstrably better descriptor), reduces to the model we have developed with slightly different coefficients (which is to be expected, considering LASSO is not the same as OLS).

We additionally utilize symbolic regression (see Appendix C.5), which is a highly flexible (but also interpretable) machine-learning technique. Despite this enhanced flexibility, we again converge to the same three descriptors: CE_{local}, IPEA, and MADs. This is a further evidence that our model has a physically-sound basis: the symbolic regression was given free rein to combine any of the investigated set of descriptors using addition, subtraction, multiplication, division, or any combination thereof, yet still quickly predicted the form of our model, albeit with different coefficients (which is again expected, as the coefficients were generated via GA and not formal OLS regression).

Although this provides excellent support for our physics-based chosen descriptors, it also indicates that no one of the descriptors we have investigated so far are able to capture the difference in elements from column-to-column. In the absence of a good tabulated physical descriptor for this – if one even exists – we instead perform fits on a per-column basis (Figure 3.4 B). We observe a general agreement with parity on the d8 and d7 metals, and excellent agreement for the d9 metals.

3.2.2.3 Extension to Rh Nanoparticles and NH3

We also tested our model to an entirely different adsorbate-metal pair: NH_3 and Rh. We see that, when we train on only Ag/Cu/Au with the three adsorbates (OH, CH_3 and CO) the Rh and NH_3 systems are off by a consistent amount as shown in Figure 3.5 A. Thus, although there is a deviation from parity, this deviation appears to be consistent and the binding trend is still captured.

Applying a constant correction to Rh and NH_3 to minimize the RMSE (and applying both corrections simultaneously to the Rh-NH_3 systems), results to bringing the data into parity (Figure 3.5 B). We again see that, although we cannot find a physical descriptor able to capture the differences between columns of the periodic table (we trained on d9 metals with CH_3/CO/OH and tested on the completely different system Rh with NH_3), we still capture the overall binding trend, albeit with an offset from parity.

Finally, we parameterized the model on Rh using all three adsorbates (Figure 3.5 C), and show that although the model has difficulties finding differences in the BEs of each site, it is still able to capture well the trend between different adsorbates. We observe general trends in the different systems when we train on Ag/Au/Cu, but offsets appear by some constant amount from the graph by metal and adsorbate. Further, when attempting to fit a model to all data simultaneously

58

(Figure C.3), we see a stratification by adsorbate-metal pair, where roughly similar energies are being predicted for the same pairs with little difference in CN (and thus the CE_{local}).
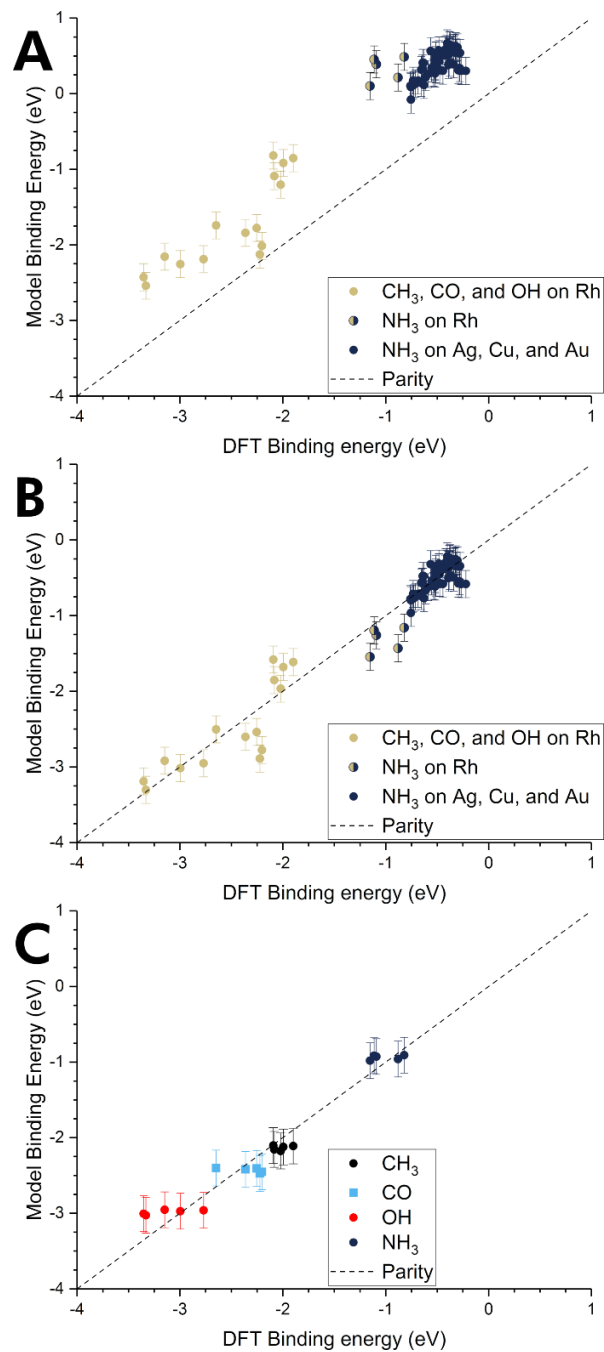
Figure 3.5: Extension of our model to Rh and NH₃. (A) The model parameterized on our Ag, Cu, and Au NPs adsorbing CH₃, CO, and OH, and tested against Rh and NH₃. (B) The equivalent model, with empirical (constant) corrections for Rh and NH₃. In the case of NH₃ bound to Rh, both corrections are simultaneously applied and indicated by two-colored dots. (C) The model trained on CH₃, CO, OH, and NH₃ adsorbing on icosahedral/cuboctahedral Rh[55].

3.2.3 Impact of this Model

In this work we develop a simple, yet powerful physics-based model for capturing trends in the strength of binding interactions between different adsorbates and metal NPs using machine learning techniques. The model introduces three simple descriptors the CE_{local}, IPEA, and MADS that is able to capture adsorption on any site of metal NPs including monometallic and bimetallic systems.

Regarding the generalizability of the model, we are able to train the model on DFT-calculated adsorption results on NPs and capture the adsorption behavior on periodic surfaces. Further, we utilize both LASSO and Symbolic Regression to search a larger space of potential descriptors that may differentiate between columns in the d-block, as well as to validate our model formulation. Both techniques demonstrated the importance of these three descriptors in adsorption. These simple descriptors can effectively model a wide range of binding interactions, including variations on the type of metals and composition, adsorption sites and adsorbates.

Beyond the functional form of this model, we present a new descriptor for the chemistry of the local site, CE_{local}. This descriptor, which represents the stability of the adsorption site, is a localization of the BCM for NP CE^{151}, where only the bonds between the binding-site metal atom and its neighbors are taken into account. The CE_{local} descriptor allows to describe the stability of the local binding site by taking into account not only the CN of its atoms, but also the coordination of its neighbors in addition to the chemical identity of the metal and its neighbors.

Our CE_{local} descriptor addresses shortcomings of other similar descriptors for local site reactivity. CN and GCN models [159] only contain geometric information, neglecting elemental composition of the site. Additionally, these descriptors do not have an obvious extension to multimetallic systems, whereas CE_{local} requires no modification to describe adsorption on

multimetallic binding sites. It is able to describe accurately adsorption on bimetallic NPs even when the model is parameterized exclusively on monometallic systems. Other descriptors, such as the orbitalwise-CN [164] are difficult to calculate and require knowledge or approximations of the electronic structure of the material of interest.

For high-throughput screening, where many different metals need to be investigated, computationally inexpensive and readily accessible descriptors are essential. We highlighted that using data tabulated for most experimental-relevant adsorbates and metals and parameterizing our model with only a single DFT calculation per metal atom-adsorbate pair we allow for rapid high-throughput screening of the adsorbate-NP and adsorbate-surface search space. Impressively, the model captures adsorption trends on bimetallic systems. Although we have not tested ternary systems here, we have no reason to believe that these physical properties will not remain relevant to accurately modeling multi-metallic systems. Further, we note that CE_{local} is constructed to account for practically any combination of metals on the local site. Future work will address if this model is directly extendable to ternary systems.

3.2.4 Comparison to Other Models in the Literature

Beyond the work of Roling et al. [157], highlighted earlier in extending our model to periodic systems, recent work has also focused on developing adsorption models using statistical techniques. The automated screening and modeling approach of Tran and Ulissi (31) describes binding-sites in terms of their atomic number, Pauling electronegativity, CN, and median of adsorption energies between the adsorbate and pure metal. This approach yields accurate adsorption energies showing good agreement with DFT, but uses statistical models with enhanced complexity such as k-nearest-neighbors and variations on the random forest. Enhanced model

complexity typically requires larger datasets to prevent overfitting; this highlights the advantage of using a simpler linear model (see discussion on the "Leave-One-In" tests in Figure C.2).

The work of Andersen et al. [163] similarly utilizes a variety of DFT (e.g. d-band and densities of states) and experimentally-derived descriptors (e.g. IP, EA, and CN), showing that DFT-calculated descriptors tend to provide better adsorption predictions than tabulated experimental descriptors such as the IP and EA. Using SISSO, they present several models, including models which take the form of first-order polynomials (form similar to Equation 3.4). Comparing with the first-order polynomial models (named Φ_0 in their work), they show training-set RMSE ranging from approximately 0.38 eV to 0.16 eV as the number of terms increases, which is similar to what we observe in our multi-d-column linear regressions. Although the DFT-based approach can provide better predictions (due to comparing DFT adsorption energies with DFT-derived properties), we note that in this work we utilize tabulated data for more rapid screening, parameterizing our model with minimal DFT calculations.

Overall, our model generally exhibits CV RMSE of about 0.12 to 0.18 eV. This error is comparable to other models in literature [157, 160, 163] and relatively small compared to the simplicity and applicability of our model. However, we suggest that our model should be primarily used for screening potential catalytic/adsorption materials to retrieve qualitative materials performance trends. For a quantitative property estimation, especially for properties that are very sensitive to error propagation as turnover frequencies in catalysis, these should be calculated by higher-fidelity methods once candidate materials are identified by our model (screening).

Our choice of descriptors has a strong physical basis. The IPEA can be viewed as the tendency of an adsorbate to react with other species, the CE_{local} is the equivalent tendency of the adsorption site to form bonds with adsorbates and the MAD is a tuning of interaction between a

63

metal and an adsorbate. Our model can accurately capture adsorption on both NPs and periodic surfaces, despite not having a descriptor for the stability of the whole surface or NP (i.e. CE_{NP}). This indicates that long-range interactions on a metal surface/NP play a minor role in determining the BE of the adsorbates. It is the stability of the local site on a surface or NP (CE_{local}) that contributes the most to adsorption on this site. We should acknowledge, however, that the adsorbates we investigate in this work are all relatively small, and most of their atoms are either directly participating in the metal-adsorbate bond or are nearest-neighbors to the bonding interaction. We hypothesize that for larger adsorbates, this IPEA descriptor may need to be replaced by a new descriptor characterizing the electronics of only the portion of the adsorbate directly involved in the bonding interaction. This can be justified via well-known effects such as induction and hyperconjugation, which (in the case of σ-bonds) tend to be limited to the range of only one or two bonds. Fortunately, there is a large body of metal-catalyzed reactions such as the Haber-Bosch[165, 166], Fischer-Tropsch [167], Water-Gas Shift Reaction [168], etc. where small-molecules similar to the ones we have investigated comprise most of the reactants and intermediates.

3.3 Conclusions

In summary, we introduce a novel adsorption model that is able to accurately describe the BE of molecules on any site of metal NPs, including alloys. Our model is simple in its form and utilizes data that can be readily accessible (or calculated on-the-fly). With surface adsorption being a critical step in catalysis, we anticipate to be highly applicable as a screening tool for the high-throughput search of potential catalysts. With the rise of large databases and recent advancements in machine learning, such high-throughput searches tend to require cheap, but physically-relevant

descriptors for reactivity of metals. In addition, our model can advance the discovery of nanosensors since it allows for screening of adsorbates and metals at the same time with regards to their interaction strength. Importantly, to the best of our knowledge, this is the first demonstration connecting adsorption properties of metal NPs (including bimetallics of random composition and chemical ordering) with the stability of the adsorption site. As a result, our model can also significantly impact materials optimization, by designing nanostructures that exhibit the desired adsorption response within a stability cutoff.

4.0 Rapid Prediction of Bimetallic Mixing Behavior at the Nanoscale

The content of this chapter is adapted from Dean, J.; Cowan, M. J.; Estes, J.; Ramadan, M.; Mpourmpakis, G. "Rapid Prediction of Bimetallic Mixing Behavior at the Nanoscale." *ACS Nano* 2020. 14 (7) 8171-8180.

Having identified several key descriptors of adsorption, we now turn our attention to the problem of NP stability. Specifically, we investigate the determination of which chemical orderings (how the elements are distributed within a NP irrespective of size, shape, or composition) maximize a NP's stability with respect to CE. The search space this presents is one which can potentially hold a great many permutations, and in the worst case becomes intractable to screen via brute force alone. To address this challenging optimization problem, in this section we develop CE Expansion, an open-source GA which is able to rapidly – and efficiently – optimize the chemical ordering of NPs of any size, shape, or composition.

4.1 Computational Methods

4.1.1 Density-Functional Theory

DFT calculations were performed using CP2K[169] implementing Quickstep.[117] The PBE functional[119] was used (Unrestricted Kohn-Sham approach) in conjunction with a DZVP basis set,[170] GTH pseudopotentials[120] with a 500 Ry cutoff and an SCF convergence within 10^{-8} Ha. A box size of 30x30x30 Å was applied for all systems. In addition, we utilized Fermi-Dirac smearing

with an electronic temperature of 300 K. For vibrational frequency calculations, the Hessian was constructed via CP2K's built-in finite difference method with displacements of 0.02 Bohr. Geometries were optimized until forces were below 4.5×10^{-4} Ha/Bohr. In the case of 13- and 55-atom NPs, the energetic minimum was confirmed via the lack of imaginary modes in the vibrational analysis. In cases where imaginary vibrational frequencies were observed, a tighter force-convergence criterion of 10^{-4} Ha/Bohr was used in conjunction with an SCF convergence criterion of 10^{-9} Ha. If imaginary frequencies remained despite the enhanced convergence criteria, the structure was not considered for further analysis. In all cases, vibrational analysis was performed with an SCF cutoff of 10^{-9} Ha, and 0.02 Bohr displacements for the central differencing scheme.

4.1.2 Thermodynamic Properties

CE is conceptually the average bond energy holding a NP or crystal together. Thermodynamically, this can be represented as the energy gained or lost in the separation of all the atoms in a NP to an infinite distance. This is described in Equation 4.1 as the difference in energy of the cluster (with formula A_xB_y) and each separate atom (with single atoms denoted as A_1 and B_1).

$$CE_{A_xB_y} = E_{A_xB_y} - \left(xE_{A_1} + yE_{B_1}\right) \qquad 4.1$$

EE can be calculated from CE, and provides a measure for-which we can quantify the tendency of two metals to mix in a NP: the more-negative the EE, the more favorable the mixing.

$$EE = CE_{A_xB_y} - \left(\frac{x}{x+y}CE_{A_{x+y}} + \frac{y}{x+y}CE_{B_{x+y}}\right) \qquad 4.2$$

Entropy of mixing ($\Delta S_{\text{mix } A_x B_y}$) is calculated per Equation 4.3.[171] We denote fractional composition via capital chi; for example $X_A = 0.5$ indicates 50% of the atoms in the cluster are of type A. k_B is Boltzmann's constant.

$$\Delta S_{\text{mix } A_x B_y} = (X_A \log X_A + X_B \log X_B) k_B \qquad 4.3$$

In Appendix D.1 (Connecting Excess Energy with Enthalpy of Mixing) we show an approximation relating the EE with the enthalpy of mixing (ΔH_{mix}). This approximation allows us to determine the free energy of mixing (ΔG_{mix}) for any of the $A_x B_y$ systems presented in this work (Equation 4.5).

$$\Delta G_{\text{mix}} = \Delta H_{\text{mix}} - T\Delta S_{\text{mix}} \approx EE - T\Delta S_{\text{mix}} \qquad 4.4$$

Utilizing the free energy of mixing, we calculate the Boltzmann distribution probabilities[172] (p_i) for the presence of the three different shapes studied at a given size, composition, and temperature T (Equation 4.5)

$$p_i = \frac{e^{\frac{\Delta G_{\text{mix}i}}{k_B T}}}{\sum_{j=1}^{3}\left(e^{\frac{\Delta G_{\text{mix}j}}{k_B T}}\right)} \qquad 4.5$$

Finally, bond fractions are calculated for a given structure as the percentage of that type of bond relative to all bonds in the system (e.g. $C_{A\text{-}A}$ represents the count of A-A bonds). The bond fraction F is calculated via Equation 4.6.

$$F_{A-A} = \frac{C_{A-A}}{C_{A-A} + C_{B-B} + C_{A-B}} \qquad\qquad 4.6$$

4.1.3 Genetic Algorithm

To determine the optimal chemical ordering of a bimetallic NP at a given size, shape, and composition, we implemented a GA using Python and C. The GA begins by creating an initial population (generation 0) of 50 bimetallic NPs with random chemical ordering. Each bimetallic NP is represented as a binary array, where zeros correspond to the first atom type in the alphabetized chemical formula (e.g. AgCu would represent Ag as 0, and Cu as 1). As a consequence of these constraints, all arrays have the same length and sum to the same value. To propagate to the next generation, the population is first evaluated for stability using the BCM,[90] given in Equation 4.7.

$$CE_{NP} = \frac{1}{n}\Sigma_1^m \left(\gamma_i \frac{CE_{bulk,i}}{\sqrt{CN_i\ CN_{bulk,i}}} + \gamma_j \frac{CE_{bulk,j}}{\sqrt{CN_j\ CN_{bulk,j}}} \right) \qquad 4.7$$

The BCM calculates the CE of a NP with n atoms by summing contributions from all bonds m in the system. Half bond contributions between atoms i and j are calculated by their respective CN, bulk CN (CN_{bulk}), bulk CE (CE_{bulk}), and a weight factor based on gas phase bond dissociation energies (γ). The model was implemented in C, enabling rapid screening of stability for virtually any size bimetallic NP.

Once CE values were calculated for each NP k, fitness scores f were developed using Equation 4.8, where $|CE_{NP}|_{min}$ is the minimum absolute value of CE within the population.

$$f_k = \left|CE_{NP,k}\right| - \left|CE_{NP}\right|_{min} \tag{4.8}$$

This scales the fitness scores such that the least stable NP in the population has a fitness of 0. We next used the roulette wheel selection algorithm[173] to determine NP as parents for mating. The probability of NP k (p_k) being selected for mating was based on its fitness score relative to the population, as given in Equation 4.9.

$$p_k = \frac{f_k}{\sum_{i=1}^{n} f_i} \tag{4.9}$$

To ensure that concentration was constrained, we developed a pairwise crossover algorithm) to mate two parents into two NP children (Figure 4.1).
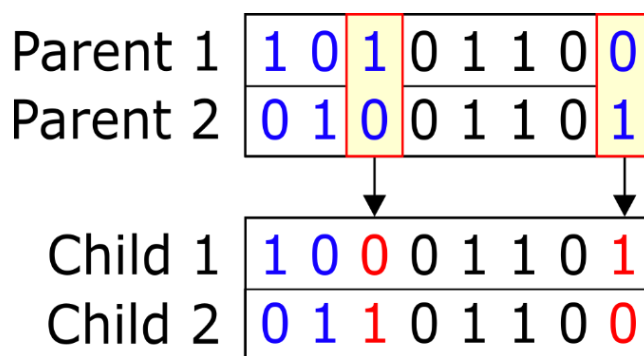
Figure 4.1: Example of the pairwise crossover algorithm implemented in the GA. Identical positions are in black while positions with different atom types are in blue. The selected pair within the parents is highlighted in yellow while the swapped atom types are in red.

The approach first finds all positions that do not match between the parents (blue numbers in the parents of Figure 4.1). Next, a pair of these positions are repeatedly chosen at random until each parent has one of each atom type (0 and 1) selected within the pair (highlighted boxes in Figure 4.1). The atom types over the selected pair are then swapped between parents. The pair

selection and swapping continue until half of the different positions are exchanged. The process results in two new children NPs that exhibit chemical ordering traits from both parents. After using the pairwise crossover algorithm to create 49 children (leaving one position in the population to pass on the current most stable NP), 80% of the new NPs are mutated. Mutations are achieved by randomly swapping 1s and 0s within a NP. The number of swaps is calculated using Equation 4.10, where N_0 and N_1 are the number of atom type 0 and 1, respectively.

$$n_{mutation\ swaps} = 0.02 \times \min(N_0, N_1) \qquad 4.10$$

Employing Equation 4.10 provides a mutation scheme that maintains genetic diversity within the NP population at any size and concentration. After mutation, the fittest NP from the previous population is added to the 49 NPs resulting in a new population – thus propelling the GA to the next generation. The overall algorithm continues until the most fit NP remains the same over 2000 generations. At this point, we apply a Metropolis-Hastings algorithm on the most stable NP for 5000 steps in an effort to find a potentially more stable structure with a similar chemical ordering. The resulting minimum CE structure is returned as our GA-optimized NP. Code which implements the GA is available on GitHub (https://github.com/mpourmpakis/ce_expansion). The linked repository includes all required functions for GA simulations as well as functions to visualize and analyze the results. Also included is a python package to build and interface with a SQL database of GA results.

4.2 Results and Discussion

4.2.1 Benchmarking the Performance of CE Expansion

We began by benchmarking our GA to assess how rapidly it would find low-energy states. We selected AgCu as a model system since it is known to favor a Cu-core Ag-shell configuration.[174] Utilizing our GA, we benchmarked an icosahedral $Ag_{1234}Cu_{1635}$ NP to see if it would capture the expected core-shell behavior, despite being presented with an extremely large search space ($\approx 10^{849}$ possibilities excluding symmetry). Figure 4.2 reveals that the GA converges to the experimentally-expected core-shell NP. Furthermore, the GA was able to screen the 1.45 million NPs used for this benchmark at a rate of 70,000 NPs/min on a single core of a typical desktop computer. Random search, having less overhead compared to the GA, screens NPs at a faster rate ($\approx 400,000$ NPs/min). However, the speed advantage of random search is undermined by the absence of an optimization scheme, and the approach does not efficiently sample the vast configuration space. This results in a failure to i) converge to the core-shell mixing behavior expected by experiment, and ii) find chemical orderings as low in energy as those found by the GA (Figure 4.2).

Consequently, there are clear differences in the surface makeup of the GA-optimized NP and the best NP from random search. Furthermore, although the GA converges to a structure with some Cu atoms on the surface, this does not stem from difficulties in finding the minimum. Rather, this stems from the composition we chose: the subsurface and core atoms are completely filled by Cu, and there is nowhere else to place remaining Cu atoms but on the surface.
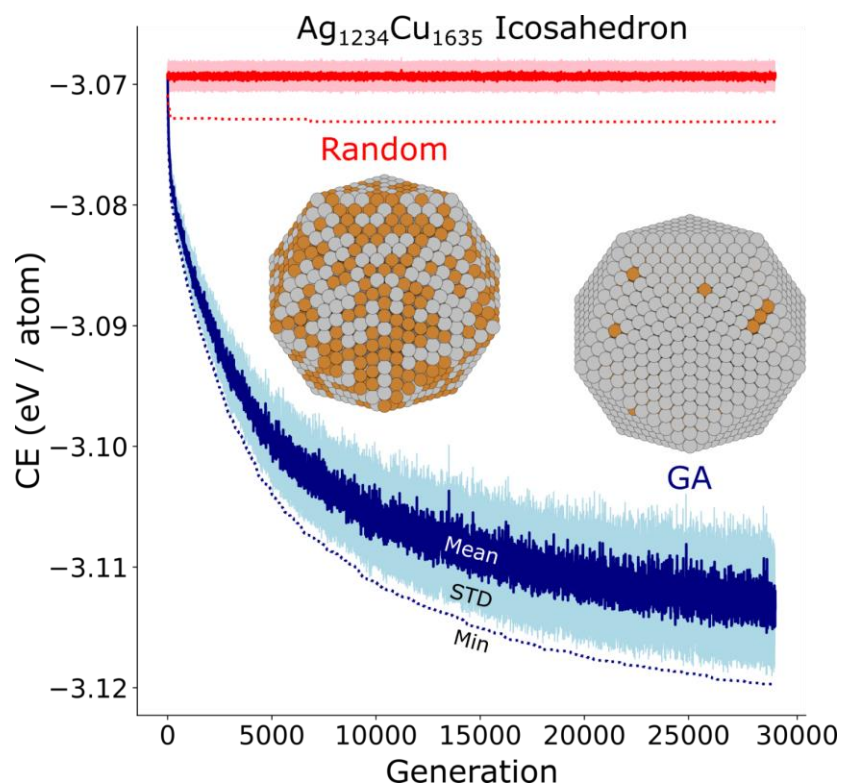
Figure 4.2: Benchmark of GA performance. We compare GA optimization (blue) against random search (red) on an icosahedral Ag$_{1234}$Cu$_{1635}$ NP. Darker solid lines and lighter shaded regions represent the mean and STD of a generation, respectively. Dotted lines indicate the minimum CE found at a given generation. The minimum CE structures are inlaid.

To further gauge the accuracy and scalability of our optimization framework, we next apply our GA to an Fe$_{6569}$Pt$_{16627}$ NP for-which the structure has been experimentally determined.[175] Through the elemental radial distribution functions shown in Figure 4.3, we reveal that the GA-optimized NP (Figure 4.3 a) captures the Fe-rich surface and Pt-rich core exhibited in the experimental NP (Figure 4.3 b). It is worth noting that using percent-composition exaggerates the composition differences in shells with fewer atoms: for example, in the first shell (Radius = 0-2.5 Å) there are 5 atoms. Because one atom in this shell is different, this is reflected by a 20%

difference between this shell's percentage composition and experiment. The differences are less exaggerated in shells with larger number of atoms. For example, in a shell containing 500 atoms, a difference of one atom would be reflected by just 0.2% in composition.

Additionally, we note that the BCM only considers thermodynamic effects. Therefore, slight differences between the GA-optimized and experimental NPs could be attributed to kinetic effects within the large experimental system (e.g. cluster growth, atomic diffusion, mass transfer, etc). More important than the radial distribution is, arguably, the distribution of atoms at the surface, where interfacial phenomena (e.g. adsorption and catalysis) occur. Hence, it is important that the methodology accurately captures the surface composition. We observe that the GA produces an excellent prediction of the true experimental surface. The results are in stark contrast to a random search approach (Figure 4.3 c) where we generated many random chemical orderings (sampling the same number of NPs as the GA) and report the most favorable based on their CE. The random NP search completely fails to capture both the radial distribution as well as the surface composition of the experimental NP. Overall, the results demonstrate the success of the GA to effectively sample a massive configurational space of experimental bimetallic NPs. Importantly, we show that our methodology allows for the fast and accurate screening of NPs which are far beyond the reach of DFT.
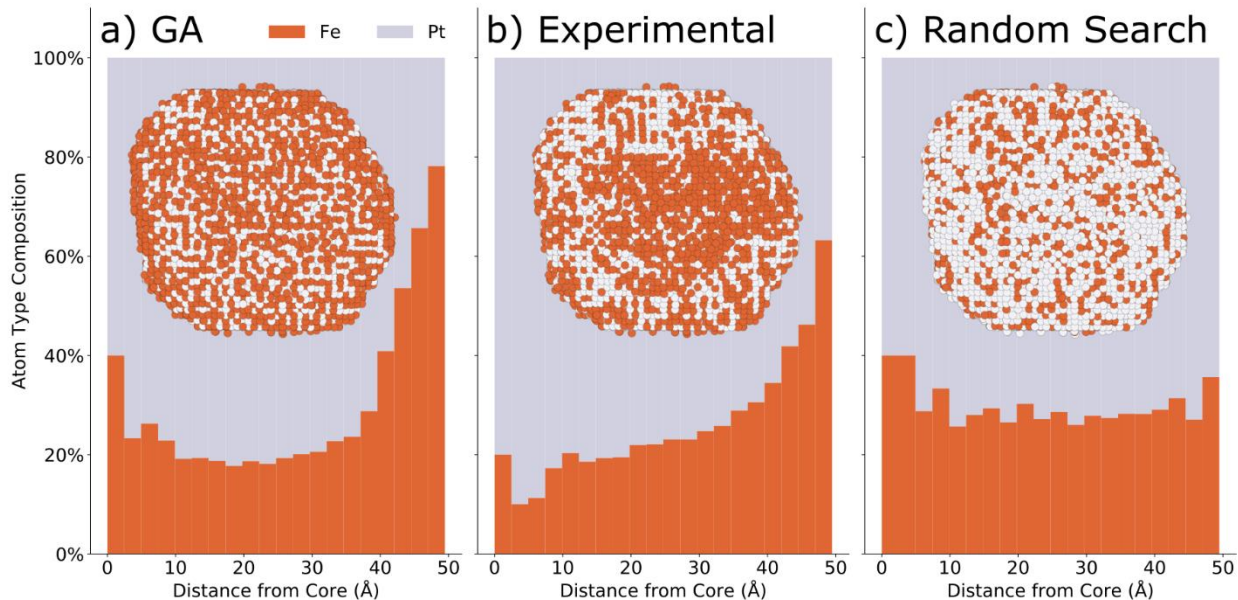
Figure 4.3: Comparison of different methodologies for identifying optimal chemical ordering. Shown are the structures of a) the GA-derived Fe$_{6569}$Pt$_{16627}$ NP to b) the experimental structure reported by Yang et al.[175] and c) the lowest-energy NP identified via random search. Core-centered radial distributions of Fe-Pt composition are represented as bar plots. The NP structures are inlaid.

4.2.2 Comparison between the Bond-Centric Model and Effective Medium Theory

The screening of bimetallic NPs for their chemical ordering is an emerging field. For example, in 2018 Larsen et al.[176] demonstrated the use of mixed integer programming coupled with a semi-empirical EMT[177] for the direct optimization of NP chemical ordering. The authors investigated icosahedral 309-atom AgAu NPs, determining the global ground-state chemical ordering for all possible compositions. These EMT global minima provide a unique dataset to directly compare the BCM and assess its effectiveness in capturing optimal mixing behavior. Thus, we used our GA to optimize the same set of NPs. We note that although the GA-optimized NPs are not guaranteed to be global minima (due to the algorithm's stochastic nature), they are expected

to be relatively close. We next selected a subset of NPs from both datasets and calculated the EE of each system using DFT. The results, shown in Figure 4.4, reveal that for every case our GA-optimized NPs exhibit lower EE (i.e. higher thermodynamic preference to mix) than the global minimum structures from EMT. In addition, we observe differences in the composition predicted to be the minimum in energy. The GA predicts a minimum energy composition close to 50/50 Ag/Au, which shows a strong agreement with DFT calculations. Conversely, EMT deviates from DFT in capturing the correct EE trend and the most favorable composition. To be more confident that the minimum energy composition is close to 50%, we additionally checked the optimized chemical orderings predicted by our GA near this composition and found that the overall EE trend holds (i.e. most favorable mixing near 50/50 Ag/Au). We note that the results only compare the capabilities of the BCM to EMT and do not make any comparison to the two optimization approaches. Nevertheless, our results prove that, unlike EMT, the BCM correctly captures mixing behavior of AgAu NPs.
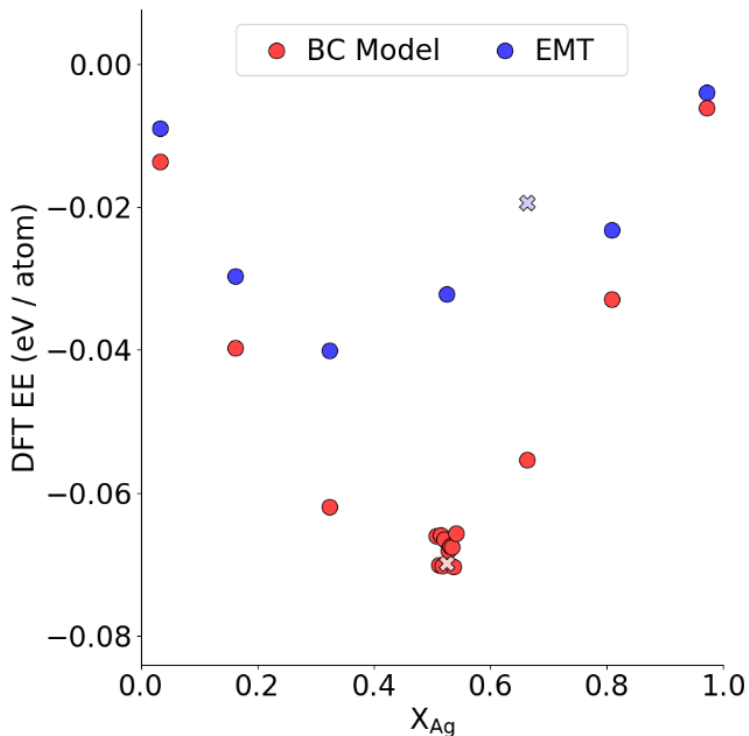
Figure 4.4: Comparison of different methodologies of optimizing chemical ordering. DFT is used to compare the EE of 309-atom AgAu icosahedral NPs with chemical ordering optimized using EMT (blue) and the BCM (red) at different compositions. X indicates the composition predicted to be of minimum energy by the respective method.

4.2.3 High-Throughput Study of Bimetallic Nanoparticles

Based on our successful results with the predicted structure of the $Fe_{6569}Pt_{23196}$ NP, and the comparison with a recent method of screening NPs for their chemical ordering, we investigated a large variety of bimetallic alloy systems. We investigated the alloys AuCu, AgAu, and AgCu, chosen because they have all been shown to be described well by the BCM.[90] Icosahedrons, Cuboctahedrons, and EPBs of up to 3871 atoms ($\approx$4 nm diameter) were investigated. For systems

with up to 309 atoms, all possible compositions were investigated. Above this size, compositions were investigated in 10 % increments as closely as the chemical formula would allow.

We choose these particular morphologies because they generally result in clusters containing so-called "magic numbers" of atoms (13, 55, 147, etc). This is important in the identification of low-energy NP structures, as NPs with magic numbered sizes tend to be of high stability.[178] In addition, we note the synthetic accessibility of these morphologies.[179-181] The EPB structure is particularly relevant to the study of nanowires, which have been demonstrated in Ag to grow via a continued elongation of this morphology.[181] Although irregular morphologies such as the $Fe_{6569}Pt_{16627}$ NP (Figure 4.3) are oftentimes experimentally synthesized, we have demonstrated that our methodology extends well to even these systems as long as a reasonable bond network can be estimated. Hence, our choice of these morphologies should be sufficient to probe the limits of our methodology, as they provide a wide range of possible coordination environments in experimentally-relevant nanostructures. Furthermore, their "regular" shape lends itself well to the programmatic generation of reasonable NP structures at any size.

Overall, we report minimum-energy configurations for 5,454 different structures. We illustrate a small fraction of the range of structures in Figure 4.5, where each NP shown is the most stable chemical ordering and composition at the given size, shape, and metal pair. Considering the number of energy calculations required for the GA to converge, and the scaling behavior of DFT (limited to systems of ~hundreds of metal atoms), this study is computationally infeasible using first principles methods. In addition, because of the number of possible configurations each NP can assume (ignoring symmetry, 2^N per cluster of N atoms), an exhaustive brute force search of our chosen chemical space is impossible even with the computational speedup provided by the BCM. It is only by combining a computationally inexpensive method (the BCM) with a tried-and-

tested optimization technique (GA) that we have the capacity to determine NPs towards global optimum chemical ordering within this tremendous materials space. Based on our results shown in Figure 4.5, we find an agreement between our algorithm's predictions and literature reports. For example, AgAu NPs have been produced with both core-shell and homogeneously-mixed motifs, however the core-shell structure is meta-stable due to the tendency for Ag and Au to alloy. Upon a heating-cooling cycle, the thermodynamic-stable, homogeneous-mixed AgAu structure emerges,[182] which is generally what our model predicts as the most stable chemical ordering for AgAu NPs of all shapes and sizes. In the case of AgCu, a DFT-based global optimization observed Cu-core / Ag-shell NPs to be the most-favorable, which was attributed to the tendency for the two to remain in (and undergo transformation to) an un-mixed state in the bulk.[183] This core-shell behavior is exactly what we observe with our GA algorithm. In fact, the most-favorable chemical ordering and composition for every shape and size AgCu NP was found to be a perfect Ag-shell Cu-core structure.
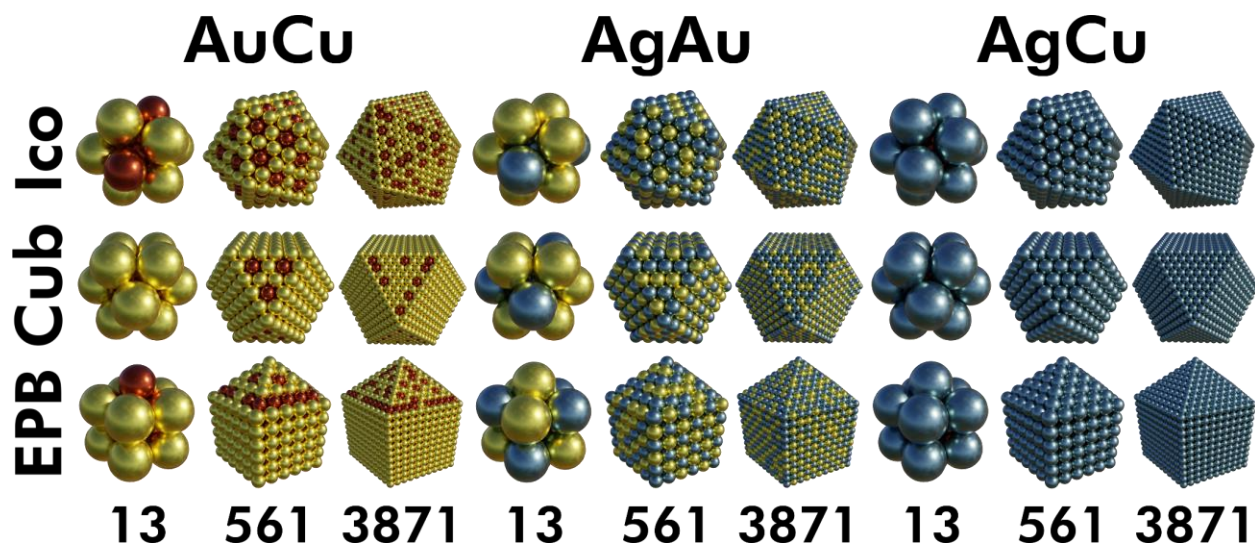
Figure 4.5: Images of the lowest-energy composition and chemical ordering observed for a sample of the NPs investigated, illustrating the breadth of our study. NPs of size 13, 561, and 3871 are shown. Left 3 columns: AuCu. Center 3 columns: AgAu. Right 3 columns: AgCu. Top row: Icosahedrons. Middle row: Cuboctahedrons. Bottom row: EPBs.

4.2.4 Bond-Composition Plots: A New Visualization of Chemical Ordering

Chemical ordering parameters as a function of homo-atomic and hetero-atomic bond counts have been applied by other groups as a useful tool to uncover mixing trends within bimetallic systems.[174, 184, 185] However, these parameters usually reduce the bond counts down to a single value to find correlations with structural properties of interest, like size or composition. Instead of distilling the counts down to a single parameter, we developed a new visualization of NP mixing. The plots shown in Figure 4.6 enable us to depict each bond type along with NP size. As an example on creating the plots for a bimetallic NP made of elements A and B, we calculate each bond type (A-A, B-B, and A-B) as a fraction of the total number of bonds (F_{A-A}, F_{B-B}, F_{A-B}).

Conveniently, each fraction must be between 0 and 1, and all three must add up to 1 (Equation 4.11).

$$F_{A-A} + F_{B-B} + F_{A-B} = 1 \qquad\qquad 4.11$$

Using our calculated bond fractions, we next build a scatter plot of F_{A-A} vs. F_{B-B} and rotate the axis such that the origin is at the top of the graph. By leveraging the unity sum constraint of Equation 4.11, we can then draw horizontal lines of constant mixing (i.e. constant values of F_{A-B}). The bottom-most line in the plot would be a completely unmixed system with no hetero-atomic bonds, as shown in Figure 4.6 a (labeled "$F_{A-B} = 0$"). At the origin would be a system with no bonds between the same atom type (i.e. only hetero-atomic bonds, labeled "$F_{A-B} = 1$" in Figure 4.6 a). As an example, the NP represented as a star in Figure 4.6 a has a hetero-atomic bond fraction of 0.4 due to the homo-atomic bond fractions (shown with thin dashed lines) of 0.2 and 0.4. Since this new representation is normalized by total bond counts, we can plot bimetallic NPs of any size, shape, and composition on a single plot, enabling global mixing trends to emerge.

Our GA-optimized AuCu, AgAu, and AgCu NP results are presented through bond composition plots in Figure 4.6 b – Figure 4.6 d. Figure 4.6 c reveals that AgAu exhibits the highest degree of mixing regardless of NP size, which strongly agrees with experimental observations.[174] AuCu systems yield similar results (Figure 4.6 b), although there is a slight influence by the NP size on the mixing behavior, as depicted by the different heights (degree of mixing) of the colored points (different NP sizes). Interestingly, for the AgCu system, Figure 4.6 d illustrates unfavorable mixing behavior, which clearly contrasts with results from the other d9 metal pairs. These conclusions agree with experimental observations that AgCu NPs tend to minimize the number of hetero-atomic bonds via obtaining a Cu/Ag core/shell morphology.[183] Overall, we show that mixing is diminished as the AgCu NPs grow in size, which suggests that the degree of mixing can

be tuned by controlling NP size. These results can also be analyzed in reference to a "fully random" system, where the bond fractions share equal likelihood and are determined strictly from probability (see Appendix D.3). Using this system as a reference (the dashed line in Figure 4.6), the results reveal that AgAu NPs favor hetero-atomic bonds while AgCu NPs favor homo-atomic bonds.
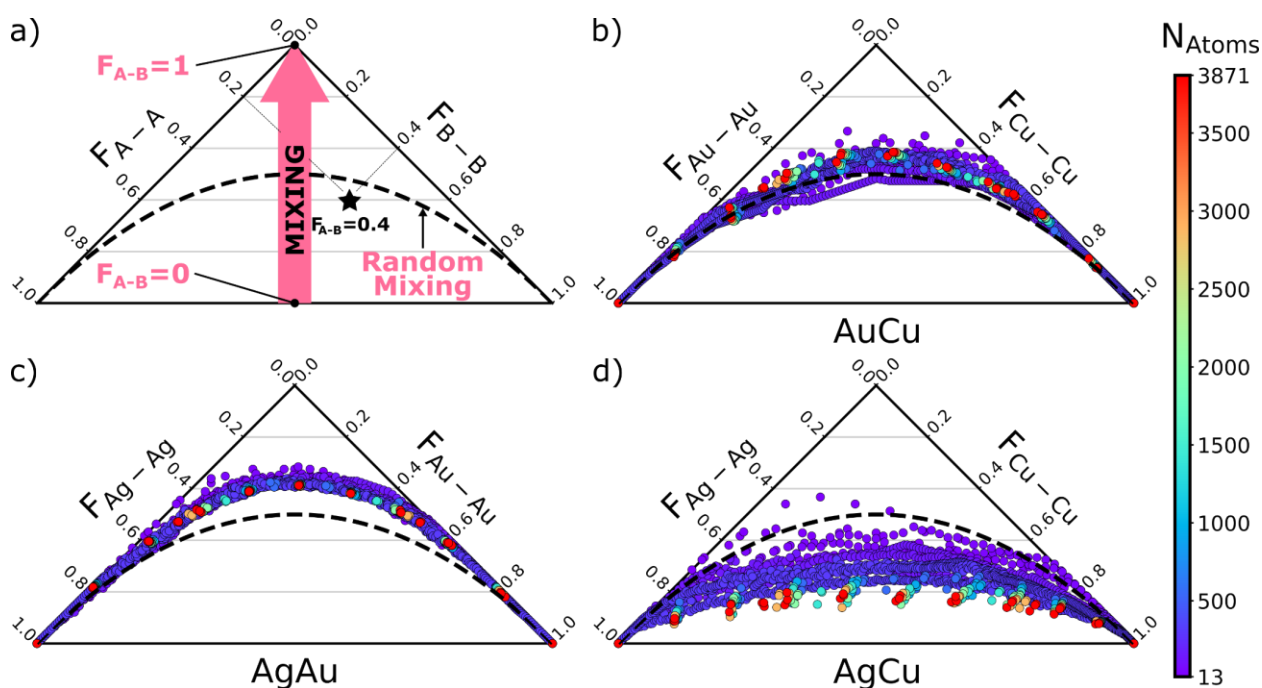
Figure 4.6: Bond composition plots. a) Guiding plot illustrating the theoretical mixing limits from no hetero-atomic bonds ($F_{A-B} = 0$) to no homo-atomic bonds ($F_{A-B} = 1$), with a NP example having 40% hetero-atomic bonds (star). Light gray lines represent constant mixing (i.e. constant fraction, F, of hetero-atomic bonds). The black dashed line indicates a system with fully random mixing (see Appendix D.3 for its derivation). Points on the plots represent b) AuCu, c) AgAu, and d) AgCu NPs at all sizes and compositions studied (5,454 total structures), demonstrating the chemical ordering that is thermodynamically preferred. NP sizes are color-coded by number of total metal atoms (N_{Atoms}).

4.2.5 Thermodynamic Analysis and Comparison with Bulk

We note that the trends demonstrated in Figure 4.6 only show structural trends resulting from the low-energy structures our GA identified within particular alloys, and only show whether an alloy is mixed to a lesser or greater degree than a perfectly random alloy (the dashed line in Figure 4.6). In other words, just because one alloy has a high degree of mixing in these plots does not imply that it is more energetically favorable than another alloy. Instead of rationalizing thermodynamics from structure, we can rationalize structure from thermodynamics. Thermodynamics can be leveraged to rationalize the structural tendency for these materials to either alloy (AgAu and AuCu) or core-shell (AgCu) NPs. In Table 4.1, we list the DFT-calculated enthalpies of formation for bulk crystal cells of AuCu, AgAu, and AgCu (for simplicity, in each case both metals exist in a 1:1 ratio) reported by the OQMD.[37, 38] In cases where the OQMD reports multiple potential structures, we use the one with the most-favorable enthalpy of formation. In our case, energetics (and as a result, structure) at the nanoscale appear to reflect the energetics of the bulk mixing behavior: the AgAu and AuCu bulk alloys listed in Table 4.1 have favorable formation energies; this indicates that mixing is energetically preferred in bulk, and is also what we observe at the nanoscale; this energetic preference causes the GA to give rise to the structural trends (i.e. preference to mix) we see in Figure 4.6 b – Figure 4.6 c.

Table 4.1: Bulk enthalpies of formation for AuCu, AgAu, and AgCu. Data collected from the OQMD.[37, 38]

Alloy	OQMD Entry Number	Enthalpy of Formation (eV/atom)
AuCu	31283	-0.053
AgAu	327735	-0.041
AgCu	307818	+0.111

In contrast, AgCu has an unfavorable formation energy in bulk (Table 4.1), which implies segregation and decomposition is energetically preferred at the bulk scale. This energetic preference to unmix is reflected at the nanoscale (Figure 4.6 d) in the structure of the AgCu NPs, which tend to minimize the formation of Ag-Cu bonds. Indeed, it is generally difficult to mix Ag and Cu; in the solid phase, the maximum solubility of Ag in Cu is only 4.9%, and the maximum solubility of Cu in Ag is 14.5%.[186] Alloys which have higher levels of dissolved Ag or Cu are known to exist, but are metastable.[186] Similarly, although AgCu NPs have been produced, they exhibit issues with stability, and spinodal decomposition has been generally observed to occur above 210-230 °C.[187, 188] Despite this, the presence of AgCu NPs experimentally reveals the importance of assessing entropy: their formation must be entropically driven due to their positive enthalpy of formation.

We apply the regular solution approximation, assessing the configurational entropy of mixing as defined in Equation 4.3. We assume that differences in vibrational entropy across NPs of a given alloy will be small enough that they can be neglected, which is supported by test calculations presented in

Table D.1 and Table D.2. These results show that at constant size, shape, and composition, changes in chemical ordering result in minimal changes to the vibrational entropy of a NP. Additionally, they show that changes in the composition of these systems results in relatively small changes to the vibrational entropy. Although more-accurate methods of approximating NP configurational entropy exist (e.g. using Monte Carlo methods[189, 190]), they would add additional computational complexity, prohibiting an analysis at the scale presented (5,454 NPs ranging from 13 – 3,871 atoms). Moreover, we note that the regular solution approximation is well-known to adequately describe NP mixing behavior and phase trends.[191-195] Thus, we use the regular solution assumption (i.e. assume ideal mixing behavior for entropic contributions) as a proof of concept to approximate composition and temperature effects of competing morphologies within a given alloy.

4.2.6 A Boltzmann Population of Nanoparticles

By combining configurational entropy and enthalpy, we can evaluate morphological preferences via the Gibbs free energy of mixing. From here, we can construct a Boltzmann population (see Computational Methods) of the three morphologies. In Figure 4.7, we visualize the effect of size and composition on the Boltzmann populations at 298K. Three morphologies are considered: icosahedral, cuboctahedral, and EPB. There are of course many other morphologies for these systems beyond those we consider, hence this analysis is primarily intended as a proof-of-concept of our GA's utility in the prediction of nanoscale phase diagrams. Nevertheless, Figure 4.7 reveals important trends between metal pairs as well as the three morphologies. For instance, although the icosahedral morphology can be a significant fraction of the population, it never seems to be the dominant phase. Instead, whenever there is a dominant set of phases, it is typically the cuboctahedron and EPB in competition with each other, which is shown by the lack of red present

in our phase diagrams in Figure 4.7. Additionally, we observe strong size effects on morphology preference within our results. In all three morphologies, as size grows, a greater fraction of the atoms is fully coordinated in the interior of the NP, and all atoms would be fully coordinated at the limit of an infinitely large NP (i.e. they would converge to the bulk). Since the only difference in coordination between the three geometries occurs at the surface, the three shapes are at their most dissimilar when the ratio of surface atoms is maximized. In other words, when comparing similarly sized systems, differences in energy are at their greatest when the NPs are at their smallest. This is especially prevalent in the AuCu system (Figure 4.7 a), where there is a heavy energetic preference towards EPB at small sizes, gradually converging towards a mix of EPB, cuboctahedron, and icosahedron at larger sizes. Overall, the population-based results presented in Figure 4.7 provide a more nuanced view than one which only considers the lowest-energy phase to be the dominant one. Importantly, they show that both size and composition play a central role in governing the NP morphology distribution. Furthermore, we investigated the effect of temperature on our Boltzmann populations, and report in Appendix D.4 variations of Figure 4.7 calculated at 77K (Figure D.1), 640K (Figure D.2), and 1073K (Figure D.3). These results reveal a rich distribution of structures at higher temperatures, and a stronger preference for specific structures at low temperatures, highlighting the importance of temperature on NP morphology distributions.
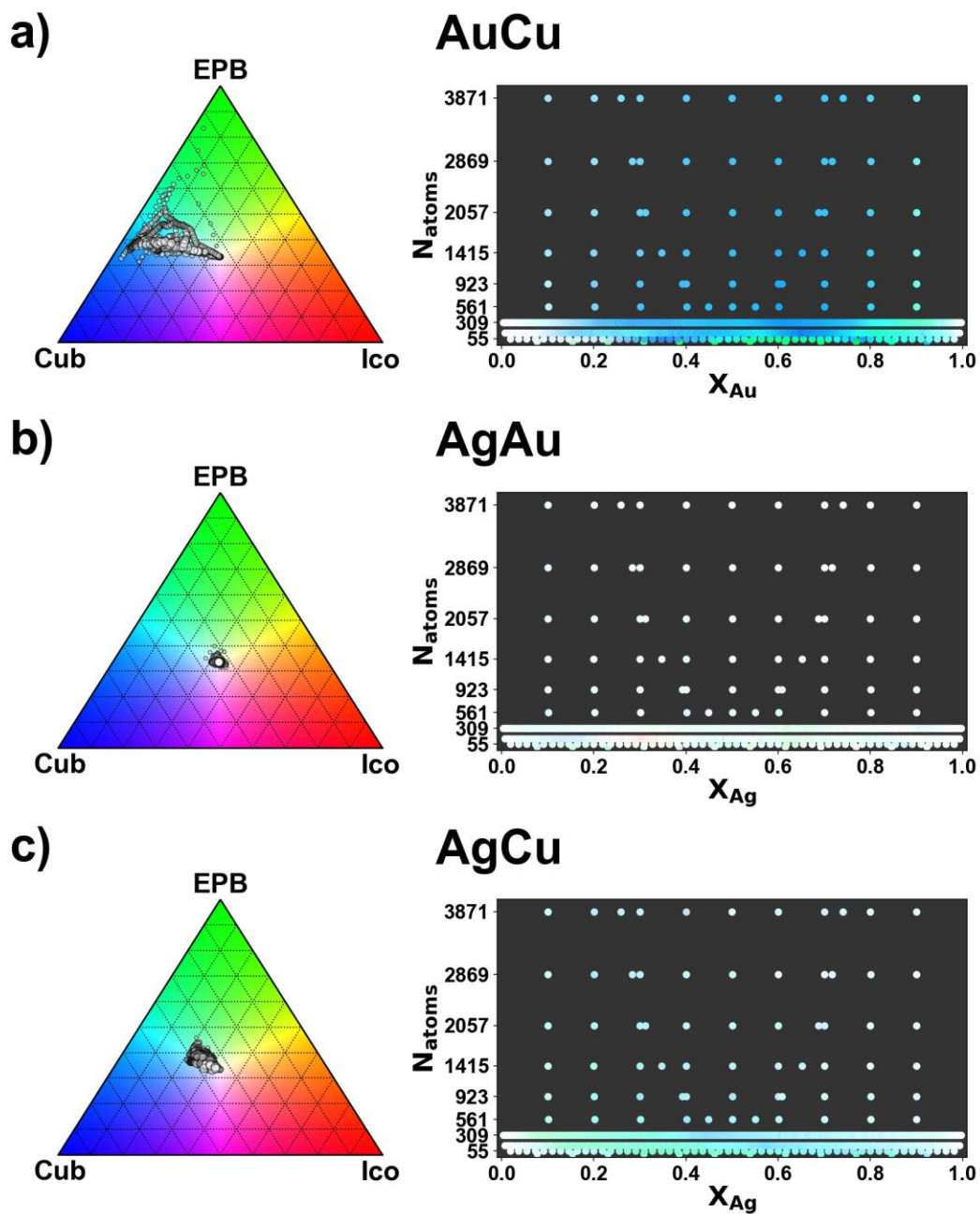
Figure 4.7: Morphology phase diagrams for the a) AuCu, b) AgAu, and c) AgCu systems at 298K. The right-hand column plots the preferred morphological phase as a function of the number of atoms and composition of the system. The legend is given by the ternary diagrams on the left-hand side, which show the percentage of the Boltzmann population taking on cuboctahedral, icosahedral, or EPB morphology. White points, for example, indicate all three morphologies are equally favorable. In addition, the size of the points on the ternary diagrams corresponds to NP size.

4.3 Conclusions

In summary, we have developed a novel methodology which blends the recently developed BCM[90] with machine learning (GA) for the rapid prediction of stable bimetallic NPs of any size, shape, and metal composition. We demonstrated the speed and accuracy of our GA through a benchmark study optimizing the chemical ordering of a 2,869-atom icosahedral AgCu NP. The benchmark shows the effectiveness of the GA in capturing expected experimental trends compared to random guess. We further demonstrated the accuracy and applicability of our approach by predicting the chemical ordering of an experimentally-determined $Fe_{6569}Pt_{16627}$ NP, achieving i) a remarkably close radial distribution of the composition, and ii) an accurate prediction of the surface composition of the NP (important for surface science applications, such as catalysis). Using results from our model, we compared the BCM to EMT[177] and demonstrated that the BCM has superior predictive power towards capturing mixing behavior of bimetallic NPs when compared to DFT calculations. Moreover, we predicted the chemical ordering of 5,454 unique bimetallic AuCu, AgAu, and AgCu NPs in a variety of sizes, shapes, and compositions, rationalizing experimental observations. The developed GA code and NP database are available free of charge on GitHub (http://github.com/mpourmpakis/ce_expansion), and a database of all 5,454 NPs is openly accessible on our newly-developed MetalNanoDB (http://metalnanodb.com; in collaboration with the ADMT lab in the Computer Science Department at the University of Pittsburgh). We introduced a visualization scheme for mixing within a NP, allowing the effective rationalization of mixing behavior between a variety of bimetallic nanostructures (different NP sizes and metal compositions). Using this visualization scheme, we connect mixing behavior to NP size and demonstrated that mixing in AgCu NPs decreases with NP size. In conjunction with Boltzmann Statistics, we developed temperature-dependent size-composition phase diagrams for each of the

three alloy systems to study the distribution of three competing morphological phases. By investigating a variety of alloy compositions, sizes, and shapes, we demonstrate that our optimization scheme accurately captures low-energy NPs without needing to resort to DFT, with the benefit of being applicable to a broad range of NP sizes (simulating tens of thousands of atoms vs. several hundred with DFT). Overall, our methodology allows for the identification of stable bimetallic NPs with atomically precise chemical ordering, which is essential for enabling the simulation of realistic, experimentally relevant NPs. As a result, our work advances the elucidation of the bimetallic NP genome.

5.0 High-Throughput Screening of Bimetallic Nanoparticles for CO_2 Adsorption

This chapter is adapted from an unpublished manuscript, Dean, J.; Mpourmpakis, G. "High-Throughput Screening of Bimetallic Nanoparticles for CO_2 Adsorption." *In Preparation.*

Having elucidated a set of adsorption descriptors and developed a technique to rapidly optimize NP chemical ordering in the previous sections, we now implement a workflow for the high-throughput screening of bimetallic NPs for CO_2 adsorption. We begin by optimizing a series bimetallic NPs of Cu, Ag, Au, Pd, and Pt via CE Expansion, alongside a set of monometallic NPs. A new tool (AutoAdsorbatePlacement) is then developed to automate the identification of binding sites and placement of CO_2, and a dataset of CO_2 adsorption energies is created. A random forest is then trained to predict CO_2 adsorption, and finally we conduct a high-throughput screening of NPs for the adsorption of CO_2 as a proof of concept of our methodology.

5.1 Computational Methods

5.1.1 CE Expansion

CE Expansion[75] was used to optimize the chemical ordering of our NPs. If a NP was not available in the database supplied by the CE Expansion package, it was optimized via the same settings used in its original incarnation, namely: a population of 50 chemical orderings, keeping the best ordering and creating 49 children each generation, continuing until the best ordering is

unchanged for 2,000 generations, along with an 80% mutation rate to maintain genetic diversity (for a more detailed discussion, refer to Section 4.0 of this dissertation).

For any pair of metal atoms (A and B), all possible compositions of the formula $A_X B_{55-X}$ were investigated, and the one minimizing the EE (maximizing thermodynamic stability) was chosen. An initial optimization of this structure was then performed with ASE's[196] built-in EMT calculator and BFGS optimizer (set to halt at fmax = 0.05 eV).

5.1.2 Density-Functional Theory

5.1.2.1 Nanoparticle Adsorption Calculations

CP2K's[197] implementation of Quickstep[117] was used for spin-unrestricted DFT calculations. The PBE functional[119] was used in conjunction Grimme's DFT-D3[198] dispersion correction. Additionally, the MOLOPT basis sets of VandeVondele and Hutter were used[170] as well as the GTH pseudopotentials[120]. A cutoff of 500 Ry was applied.

The SCF optimization was carried out with a convergence criterion of 10^{-7} Ha, and geometries were optimized until they converged to within $4*10^{-4}$ Ha. A computational box of size $30*30*30$ Å^3 was used. Fermi smearing at an electronic temperature of 300K was also leveraged. In the case of adsorption states, the NP atomic positions were frozen and only CO_2 was allowed to relax. BE is calculated using Equation 5.1:

$$E_{Bind} = E_{CO_2+NP} - \left(E_{NP} + E_{CO_2}\right) \qquad 5.1$$

5.1.2.2 Single-Metal Adsorption Calculations

The metal-CO_2 complexes used in the calculation of MADs were assessed via spin-unrestricted DFT calculations in Gaussian 09[199]. The PBE functional[119] was used alongside the

def2-SVP basis set[200], and dispersion was accounted-for via Grimme's DFT-D3[198] correction. The complexes were constructed by placing the metal atom in a series of discrete positions around CO_2, and optimizing via DFT. The configuration minimizing the BE (as defined in Equation 5.1) was taken.

5.1.3 Random Forests

Random forest[201] regression was utilized as implemented in SciKit-Learn[202]. A grid search was used to optimize the model's hyperparameters, such that they minimize the 10-fold CV RMSE of the trained model. Hyperparameters optimized included number of trees in the forest (2, 4, 8, 16, 32, 64, 128, or 256), the maximum number of features a tree could consider at a time (1-9), the minimum number of samples in a node for a split to occur (2-9), and the minimum number of samples a leaf node could contain (1-9).

5.1.4 Neural Networks

Neural Networks were optimized within TensorFlow[203] using the Adam optimizer[204], a batch size of 64 samples, and MSE as the loss function. Due to their many convenient properties, including i) addressing the vanishing gradient problem and ii) the dying neuron problem, as well as iii) their self-normalizing effect, we leverage SELU activation functions for our neurons[203]. Consequently, instead of the default Xavier initialization strategy[205], we apply the initialization strategy of LeCun[206], in which the network's initial weights are selected from a normal distribution with variance equal to the inverse of the number of neurons in the previous layer. As this is a regression problem, the output layer's neuron employs linear activation.

Using SciKit-Learn's built-in CV grid search algorithm[202], we checked MLP architectures containing 1, 2, 3, or 4 hidden layers and 2, 4, 8, or 16 neurons per layer. In each case, 5-fold CV was performed, with MSE as the error metric. For the purposes of identifying an optimal architecture via the CV grid search, 128 training epochs were applied uniformly to all models.

Once an optimal architecture was identified using the 5-fold CV strategy, we split the dataset into 10% validation and 90% training data, and perform a final optimization. Training continued until at least 50 epochs had elapsed without improvement in the validation set's MSE, at which point the model weights which had the best validation set MSE were restored.

5.1.5 AutoAdsorbatePlacement

AutoAdsorbatePlacement is a new code written in Python to automate the detection of surface binding sites (top, bridge, or hollow, as illustrated in Figure 5.1). In addition, it also provides a convenient, object-oriented interface to aid in the algorithmic placement of molecules to these binding sites.
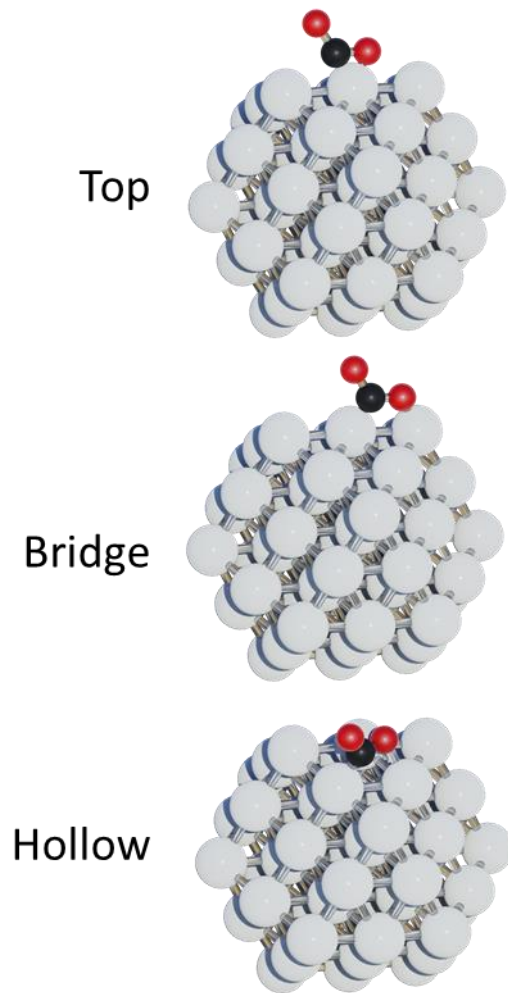
Figure 5.1. Illustration of the three types of binding (top/bridge/hollow) investigated. Key: O = Red, C = Black, Generic metal atoms = grey.

5.1.5.1 The Automated Detection of Nanoparticle Binding Sites

To detect binding sites, we first determine the CN of each atom in the NP, and take under-coordinated atoms (CN < 12) to be surface atoms. For each site, the following hashing strategy is used:

1. The set of first-nearest-neighbors is found, and sorted. Any sorting metric and algorithm is acceptable, as long as it is deterministic. In our case, we sort first-nearest-neighbors by the index of the atom within the molecule.

2. For each first-nearest-neighbor in this list, we repeat this operation to find a sorted list of second-nearest-neighbors. In this list, atoms are represented by their atomic symbol. A hashing algorithm is applied to the sorted list. This hashing algorithm is arbitrary, so long as it is deterministic and avoids hash collisions. In our case, we use Python's built-in hashing algorithm for a tuple of strings, which results in an integer. By applying this operation to every first-nearest-neighbor, we arrive at a list of hashes.

3. This list of hashes is once again sorted and hashed. Again, sorting and hashing algorithm choice is arbitrary as long as it is deterministic and avoids hash collisions. In our case, we use Python's built-in hashing algorithm for a tuple of integers. This results in a unique hash describing a binding site, which changes based on CN and atomic identity of the second-nearest-neighbors.

4. If two sites share the same hash, we consider them to be equivalent.

This strategy is based on CE_{local}, which has been shown to be a descriptor of BE[24]. Our justification is that two binding sites will share the same value of CE_{local} if they 1) have the same CN, 2) their first nearest neighbors have the same CN, 3) they share the same atomic identity, and 4) their first nearest neighbors share the same atomic identities.

5.1.5.2 Procedural Placement of Adsorbates to Clusters

AutoAdsorbatePlacement is structured such that there are separate classes (which extend ASE's existing Atoms class) for adsorbates and adsorbants (that is, anything that an adsorbate can adsorb to, such as NP or surface). These classes are organized such that adsorbates and adsorbants

each have "hooks," that is, stored positions which help in determining the correct geometry of a complex. In the case of adsorbants, this is a specific point above the binding site. In the case of adsorbates, this is a specific point within the molecule, chosen such that it aids in creating the adsorbed state geometry.

For CO_2, we set the binding site hook as the midpoint of one C-O bond. For the purposes of this section, we will call this bond the C-O^1 bond, with C-O^2 representing the bond to the other oxygen (O^2). Our algorithm to generate a binding configuration is as follows:

1. The hook, centered on the C-O^1 bond, is placed on the binding site hook, with the rest of the molecule in any arbitrary orientation.

2. Any arbitrary vector perpendicular to the normal direction of the NP surface at that binding site is found, and the C-O^1 bond is rotated into this perpendicular vector.

3. The C-O^1 bond is then rotated about the surface normal such that the average distance between C and the binding site atoms is minimized.

4. The C-O^2 bond is then rotated about the vector defined by the C-O^1 bond, such that its distance to the NP surface is maximized.

5.2 Results and Discussion

5.2.1 Dataset Generation

CE Expansion was used to select and optimize NPs containing 55 atoms. All monometallic systems and bimetallic alloys of Cu, Ag, Au, Pd, or Pt were considered, in a variety of morphologies including cuboctahedra, EPBs, and icosahedra. All 18 considered NPs are illustrated

in Figure 5.2. These structures were then optimized using DFT, followed by application of AutoAdsorbatePlacement to create the CO_2-NP complexes. This resulted in a total of 661 BEs.

EPB Ag$_{27}$Au$_{28}$ Cub Ag$_{25}$Au$_{30}$ Ico Ag$_{29}$Au$_{26}$

Ico Au$_{55}$ Ico Ag$_{55}$ Cub Pt$_{55}$

Cub Au$_{42}$Pt$_{13}$ Cub Ag$_{26}$Pt$_{29}$ Cub Cu$_{26}$Pt29

Ico Au$_{33}$Cu$_{22}$ Ico Ag$_{42}$Cu$_{12}$ Ico Cu$_{42}$Pd$_{13}$

Cub Au$_{12}$Pd$_{43}$ Cub Ag$_{42}$Pd$_{13}$ Ico Cu$_{55}$

Ico Au$_{22}$Pd$_{33}$ Ico Pt$_{13}$Pd$_{42}$ Ico Pd$_{55}$
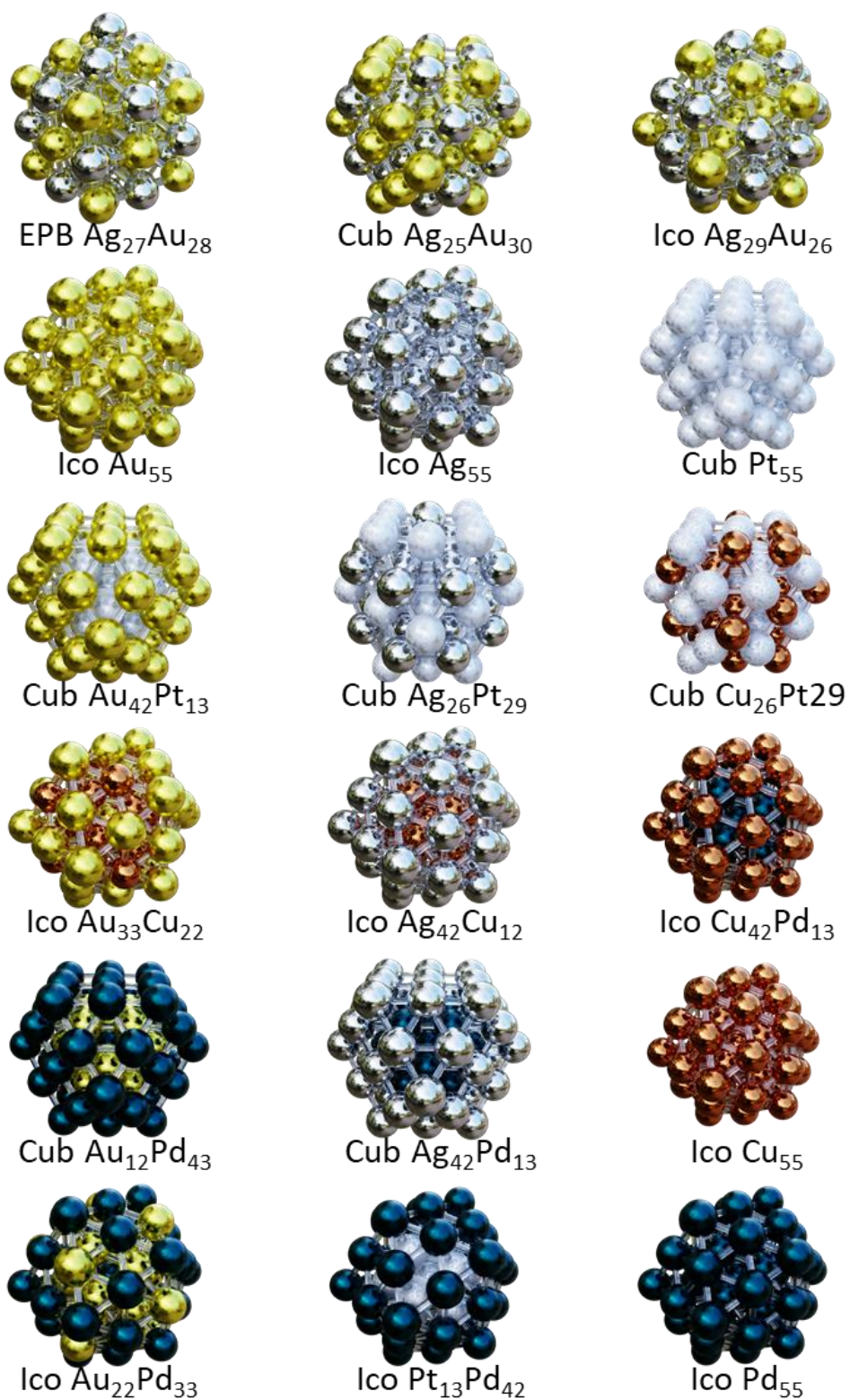
Figure 5.2: NPs optimized in this study. Key: Pd = Blue, Pt = Pearl White, Au = Gold, Ag = Silver, Cu = Copper.

5.2.2 Feature Engineering

A wide range of features were chosen for investigation in our dataset, selected because they are either tabulated, or come from rapid calculations which only need to be performed once to characterize a metal (as in the case of MADs). Special care was taken to extrapolate these features into possible multidentate binding configurations.

In the case of CE_{local}, an average of the original CE_{local} descriptor across all atoms in the binding site was used. This can be intuited as the average energy binding a set of atoms within the NP. MADs, on the other hand was calculated as a summation, as this allows it the differentiation between multidentate binding sites (i.e. bridge and hollow) even when all binding atoms are of the same type. To understand why this is the case, consider an adsorbate which binds to N Cu atoms. The averaged MADs would be equal to $(N * MADs_{Cu}) * \frac{1}{N}$, regardless of whether it were bound to 1, 2, 3, or more Cu atoms. If we sum MADs instead, the value becomes equal to $N * MADs_{Cu}$.

ΔIP[112], which has been shown to be potentially relevant to CO_2 adsorption, was also extended to sites with multiple atoms. This is accomplished by iterating over every atom of the site, and averaging the difference in IP between each atom and each of its neighbors.
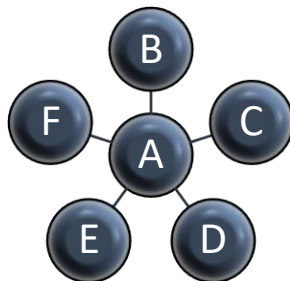
Figure 5.3: Reference image for example of ΔIP. In this case, we calculate the value for an atom A bound to atoms B, C, D, E, and F.

As an example of our extended ΔIP, consider Figure 5.3, where we want to calculate the ΔIP of some atom A, which is is bound to atoms B-F. We first take the difference in IP between A and B, then A and C, and so on until A and F. These differences are all added together, and divided by the CN of A, as depicted in Equation 5.2.

$$\Delta IP = \frac{(IP_A - IP_B) + (IP_A - IP_C) + (IP_A - IP_D) + (IP_A - IP_E) + (IP_A - IP_F)}{CN_A} \qquad 5.2$$

Because the original version[112] of ΔIP was simply the difference in IP between a single dopant atom and the host metal (i.e. in a singly-doped $Cu_{54}M$ NP, the host metal is Cu), this works out to be back compatible, as in that case, the same IP difference is taken CN_M times, then divided by CN_M.

All other features (number of d-[207] or valence[208] electrons, ionization energy[207], crystal radius[207], bulk resistivity[207], bulk melting point[207], bulk conductivity[207], 298K crystal entropy[208], and the 298K gas-phase entropy[208] / enthalpy[208] / Gibbs free energy[208]) were averaged over the atoms present in the binding site.

5.2.3 OLS Regression

We take our first step into this dataset using OLS regression, the results of-which are shown as a parity plot in Figure 5.4. In both cases, we use three terms in the regression model: CE_{local}, IPEA, and MADs.

Using our extended versions of CE_{local} and MADs (see the development in Section 5.2.2, Feature Engineering), we first used the original parameters of our previous adsorption model (Table 5.1). This leads to a relatively poor RMSE of 0.24 eV with an R^2 of -3.78, which indicates the model in this case has worse predictive power than simply taking the mean of the dataset. When we re-fit using just the CO_2 dataset (coefficients in Table 5.1), we see an improvement in RMSE to 0.11 eV, and an improvement in R^2 to 0.04. Overall, although re-fitting marginally improves the quality of the model, it's still not much better than just taking the mean.

Table 5.1. Coefficients used in the OLS portion of this work.

Feature	Original Parameterization (CH₃, CO, OH)	Re-fit to CO₂
CE_{local}	-0.145	-0.013
IPEA	0.332	0.000
MADs	0.679	0.067
Intercept	1.515	-0.238
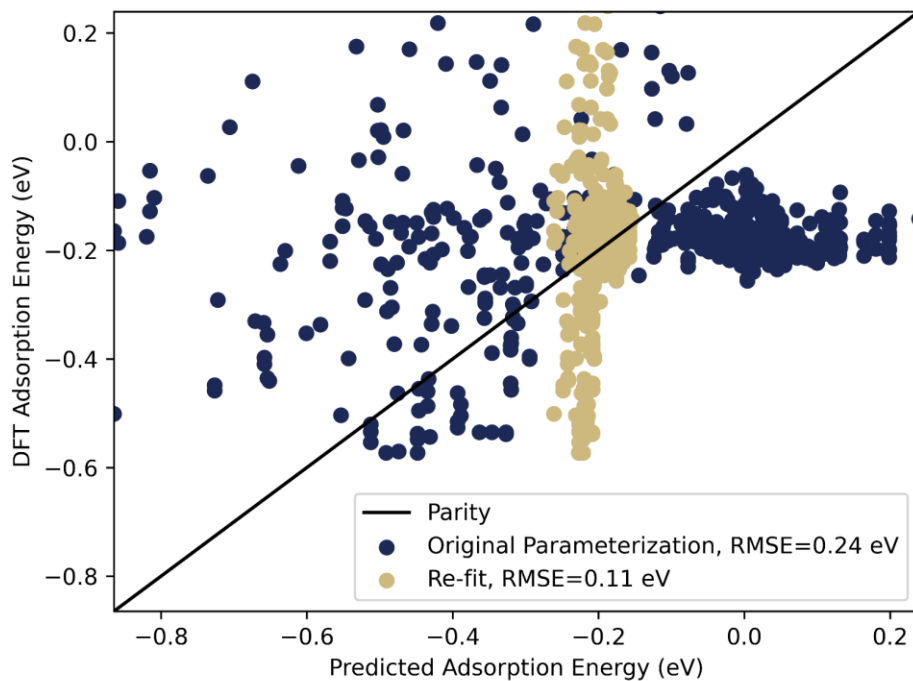
Figure 5.4: Parity plot featuring OLS methods investigated to model CO₂ adsorption.

5.2.4 Random Forest Regression

Next, we train a random forest on all features considered in Section 5.2.2 (Feature Engineering). The purpose of this is to take advantage of the variable importance metric [201] generated by this type of model, which we use for feature-selection purposes. We observe the three most important features (see Appendix E.1, Fitting Information for details on the random forest parameters chosen) to be, in order, 1) CE_{local}, 2) MADs, and 3) ΔIP.

Using these three features to train a new random forest, we find that a random forest with 16 estimators, 1 sample minimum per leaf, and 3 samples lends itself to a 10-fold CV RMSE of

0.07 eV (Figure 5.5, blue points). Overall, this is much better performance than the OLS approach, mostly due to the enhanced flexibility afforded by the random forest approach.

To understand whether a less-complex model is also feasible, we additionally trained a 2-parameter model (Figure 5.5, gold points), and observe a marginally worse 10-fold CV RMSE of 0.08 eV. We also see more bias in the model, with a tendency to predict values closer to approximately -0.15 eV. This is most likely due to the large number of datapoints located in the region between -0.1 to -0.2 eV. To a lesser extent, we also see this present in the 3-parameter model, which predicts binding energies closer to -0.15 eV for weakly-adsorbing sites.
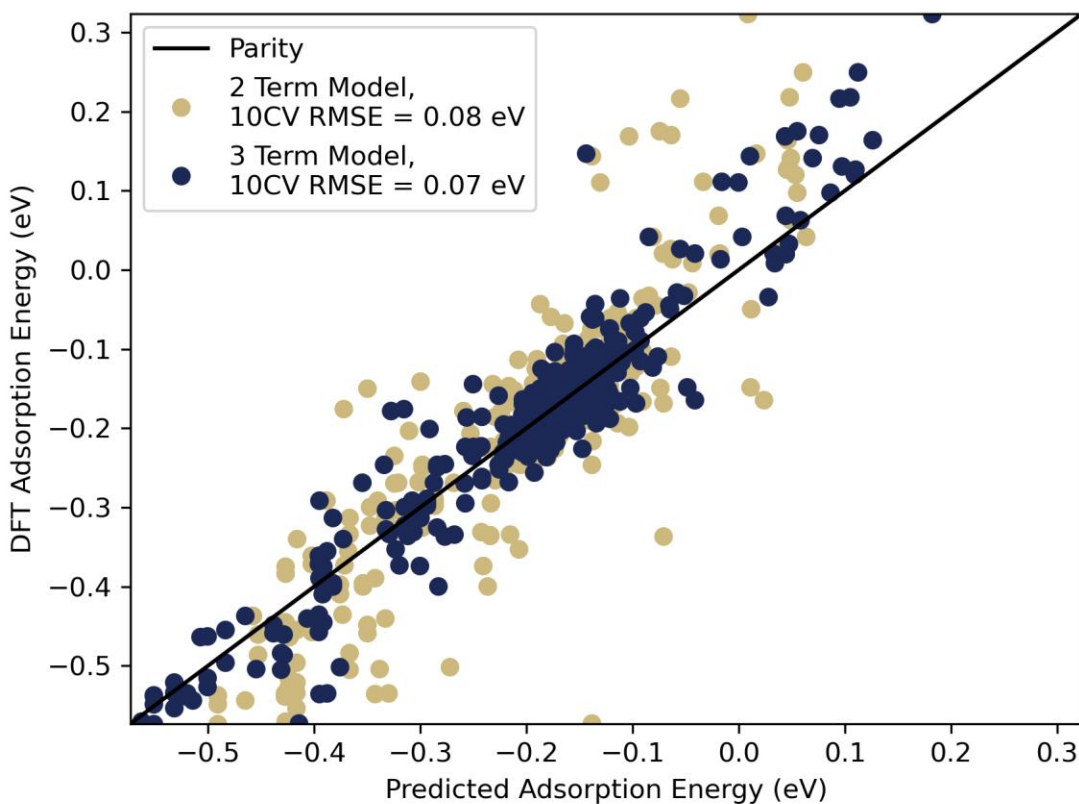
Figure 5.5: Random forest regression results.

5.2.5 Neural Network Regression

Finally, we utilized a neural network regression approach using the three descriptors (CE_{local}, MADs, and ΔIP) identified with our random forest importance scores. Our 5-fold CV hyperparameter optimization found that 4 hidden layers with 8 layers per node were optimal in terms of MSE, with a predicted RMSE of 0.1 eV. This architecture is illustrated in Figure 5.6.

Using this architecture, we fit the neural network to our entire dataset of 661 adsorption energies (using 10% of them as a validation set for early-stopping purposes). The results of this regression can be found in Figure 5.7.
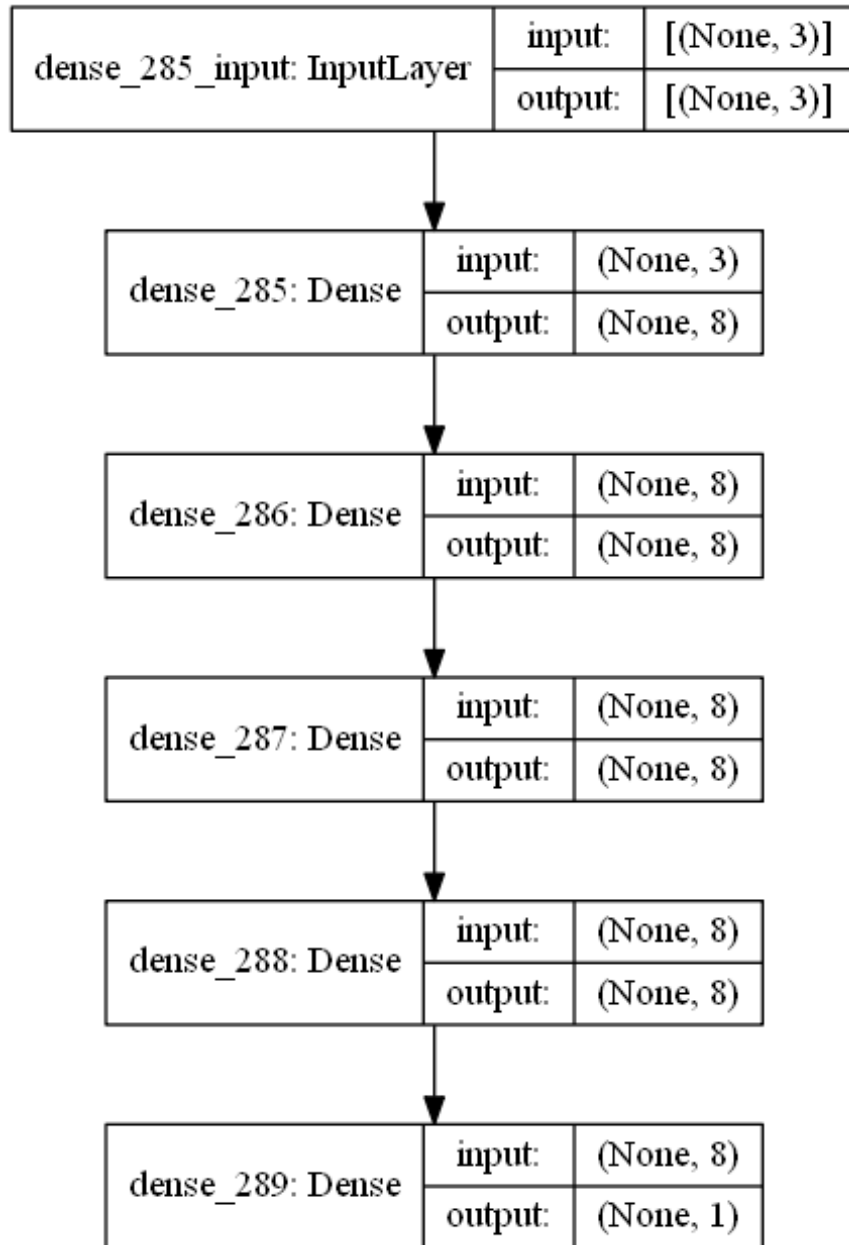
Figure 5.6: Illustration of our final neural network model architecture generated by TensorFlow. Input and output dimensions are shown on the right-hand portion of each layer. Data enters the neural network at the top, flowing in the direction of the arrows, and is output by the output layer on the bottom of the illustration.
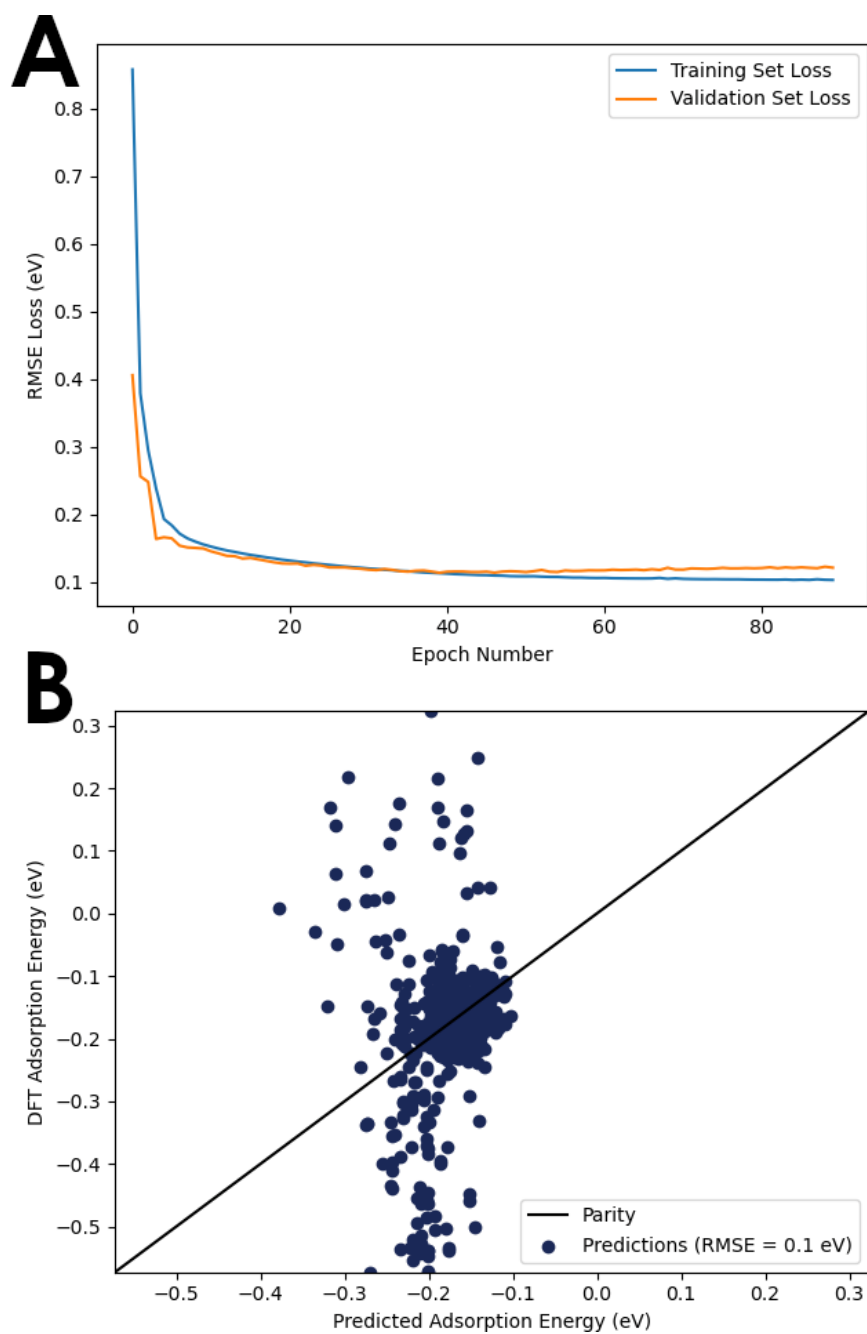
Figure 5.7: Results of our neural network regresion. (A) Plot of the training and validation set learning curves. (B) Parity plot of the model's prediction. The 5-fold CV's predicted RMSE for this architecture of 0.1 eV is shown in the legend.

We see in Figure 5.7 A the neural network learning curve. We observe that most of the learning occurs early-on, leveling out after only a few epochs. This helps justify the approach taken by our hyperparameter optimization, where we took 128 epochs for every architecture: even this, one of the more complex architectures, was adequately trained far before that point was reached.

Unfortunately, we observe similar results to those seen in our OLS regression (Figure 5.4), where the model generally predicts only a few values around the dataset's average BE. A probable cause of this is the sampling imbalance between weak and strong adsorption: half the BEs are weaker than -0.18 eV. Because so many samples have low BE, this likely results in this optimum based on an average appearing in both the OLS and neural network regression methods.

5.2.6 High-Throughput Screening of Bimetallic Nanoparticles for CO_2 Adsorption

In order of increasing performance, we trained an OLS model (RMSE = 0.11 eV), neural network (RMSE = 0.1 eV), a 2-feature random forest (RMSE = 0.08 eV), and a 3-feature random forest (RMSE = 0.07 eV). Notably, we observe our random forests (Figure 5.5) vastly outperform both OLS (Figure 5.4) and the neural network (Figure 5.7). This case is not totally unexpected however, as random forests are known to occasionally outperform more flexible models such as neural networks [50, 209].

As a proof of concept of our machine learning-driven high-throughput screening, we now use this 3-term random forest to screen over several hundred NPs whose chemical ordering were optimized as part of our CE Expansion publication[75].

This specific database contains AgCu, AgAu, and AuCu NPs in icosahedral, cuboctahedral, and EPB morphologies. We investigate clusters with 13 and 55 atoms; in both cases, all possible compositions have already been optimized. This yields a total of 126 13-atom

NPs, and 504 55-atom NPs. We then apply AutoAdsorbatePlacement's binding site detection algorithm to identify all possible top, bridge, and hollow sites on the surface of these NPs. The 13-atom NPs collectively yield a total of 2,292 binding sites, and 44,330 chemically-unique binding sites are identified amongst the 55-atom NPs. A histogram of the BEs of all 46,622 binding sites can be found in Figure 5.8.
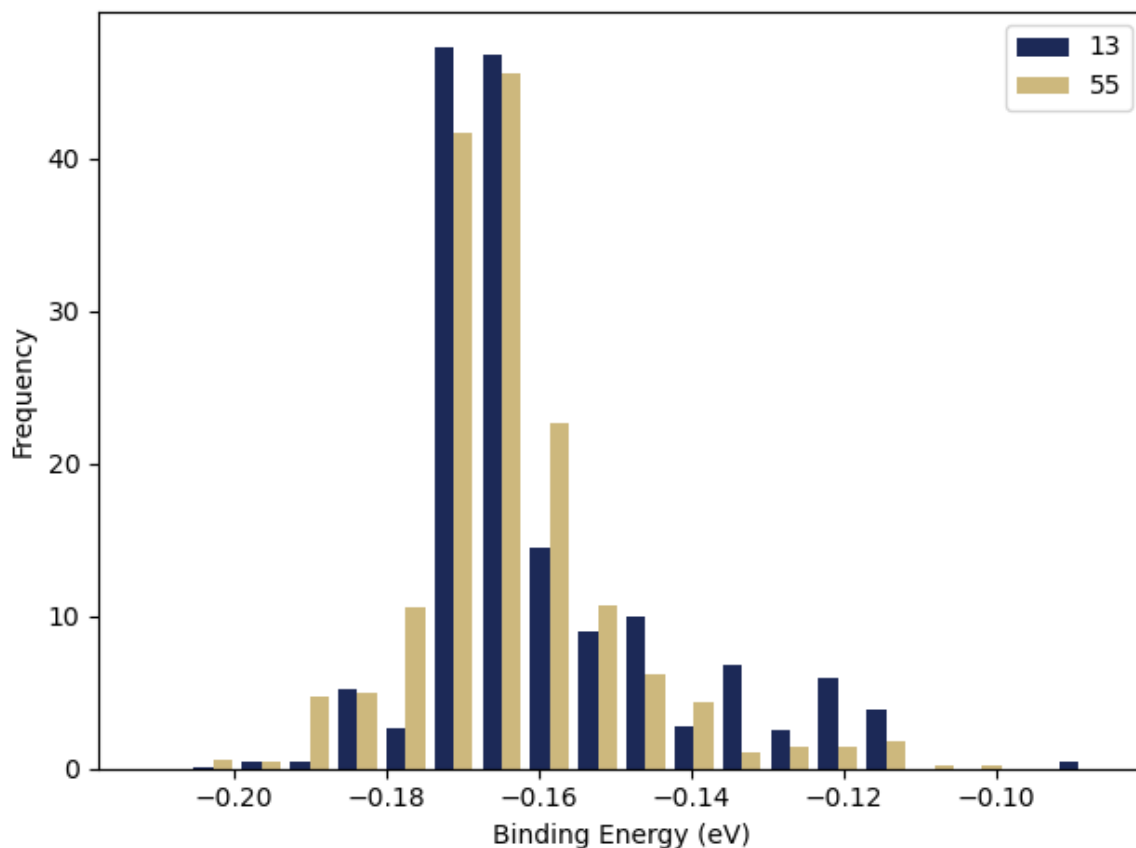
Figure 5.8: Histogoram of 46,622 predicted BEs. In the case of every size, the frequency distribution is normalized such that it integrates to 100%.

We see in this figure that most of the predicted BEs are weak, around -0.17 eV. Indeed, we see from our DFT calculations that alloys between Cu, Ag, and Au generally bind CO_2 weakly. In the case of Cu, this is also consistent with our prior doping study[112], which observed both Ag and

Au doped Cu NPs to only physisorb CO_2. Although these specific NPs don't seem to display very strong binding, their sheer number (over 46,000!) is still useful to demonstrate our methodology of using a small subset to train a model, which is then used to investigate the rest of the chemical space. To underscore the computational benefits of this approach: although our training set was calculated using DFT on a supercomputing facility over the course of hundreds of thousands of core-hours, this screening operation was performed single-threaded on an average desktop computer in just two hours.

5.2.7 Visualizing a Hypothetical Chemical Space

In addition to our high-throughput screening of tens of thousands of NPs, we also leveraged our trained random forest to visualize a hypothetical chemical space. This space is bounded by the minimum / maximum values of CE_{local}, MADs, and specific values of ΔIP. We show in Figure 5.9 several plots, ranging across different values for ΔIP. We consider three different values for $ΔIP$, as in its original version[112], its sign was observed to be relevant to CO_2 adsorption prediction. Specifically, negative values were reported to facilitate strong CO_2 adsorption and activation, and positive values were reported to result in a weak CO_2 physisorption. Figure 5.9 A shows $ΔIP=-1$, B shows $ΔIP=0$, and C shows $ΔIP=1$. To form an intuition about this figure, which visualizes three different slices of the same 3D space, it may help to consider $ΔIP$ as being a direction perpendicular to the page.

We observe in Figure 5.9 a highly complex chemical space predicted by the random forest. The blocky nature of the space is likely an artifact of the decision boundaries being created by the trees in the forest. Additionally, the long rectangles extending outwards from some portions (this is especially apparent in the large ~0.0 eV BE in the upper left of A and B) are likely a result of

the problem random forests have in extrapolation. By their very nature as an average of individual decision trees, the random forest is unable to capture BEs which are higher or lower than the highest or lowest BE observed in the dataset.

Despite these drawbacks, random forests are a powerful tool for interpolation[210]. Overall, they allow us to identify regions in the chemical space which are likely to have strong BEs, and thus which may yield good candidate catalysts to study further for CO_2 conversion.

In all three plots, our model predicts strong binding (dark blue) to occur at least in the region of $-7 \text{ eV} \leq \text{CE}_{local} \leq -5 \text{ eV}$. When ΔIP is positive (Figure 5.9 A), this is predicted to occur when MADs is greater than -6 eV, and when it is close to -9 eV. When ΔIP is negative (Figure 5.9 B), our model predicts a maximal value when MADs is close to this same value of -9 eV, and when it is between -4 eV and -6 eV. Finally, when ΔIP is positive (Figure 5.9 C), our model seems to restrict its strongest predicted binding to MADs greater than -6 eV.

A clear next step in this work would be to combine this model's predictions with the high-throughput screening approach demonstrated in Section 5.2.6, for systems outside the dataset. Even restricting ourselves to FCC metals, there are still plenty of other transition metals alloys which could be considered: in addition to those of Cu, Ag, Au, Pd, and Pt, including bimetallic alloys of Ni, Rh, and Ir are also possible. By optimizing several of these bimetallic NPs using CE Expansion, identifying all binding sites with AutoAdsorbatePlacement, and then filtering out those which are not predicted by this model as exhibiting strong binding, one would arrive at a set of candidate catalysts. These candidates could then be further studied using DFT, and then incorporated back into the training set, to enable the continued, iterative improvement of its predictive capabilities.
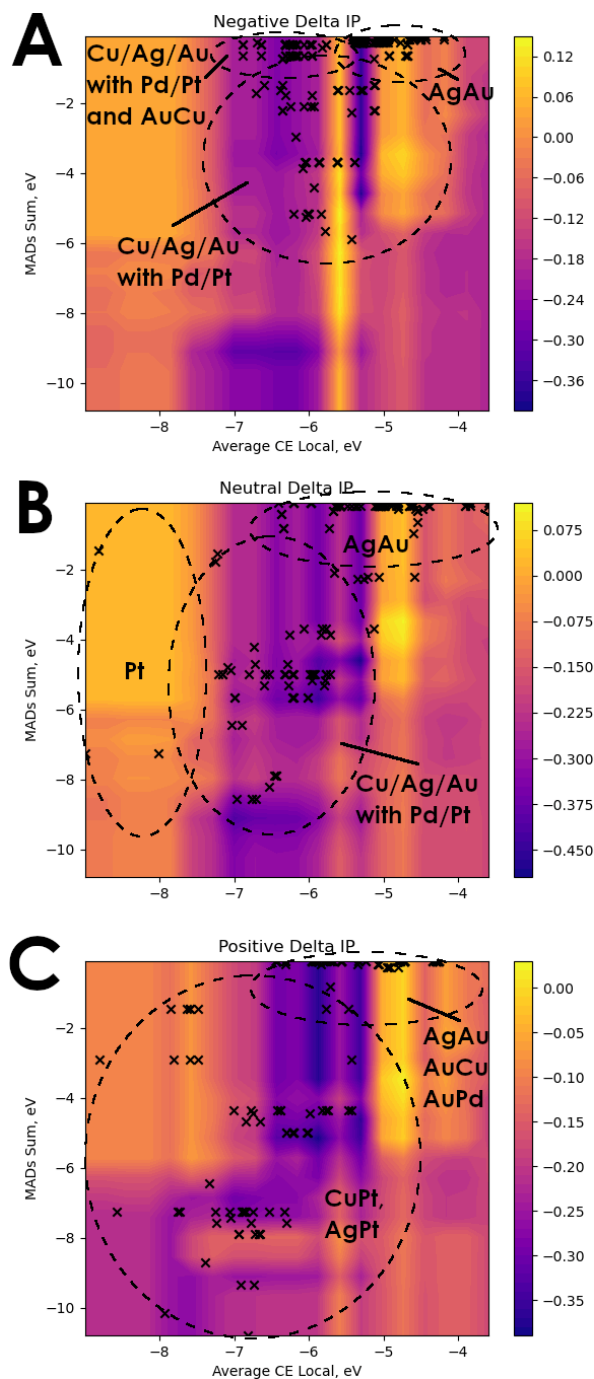
Figure 5.9: Plot of CO₂ BE as a function of CE_local, MADs, and ΔIP, as predicted by the random forest. Predictions are sampled from a 20x20 mesh of points across the CE_loocal and MADs dimensions. Known datapoints are plot with an X. (A) ΔIP=1, and portions of the dataset with ΔIP < -0.2 are shown. (B) ΔIP=0, and portions of the dataset with -0.2 ≤ ΔIP ≤ 0s.2 are shown. (C) ΔIP>0, and portions of the dataset with ΔIP>0.2 are shown.

With regards to which of these alloys to screen first, we note that our DFT showed the strongest binding to occur in alloys containing Pd and Pt. Unfortunately, Pd and Pt are both expensive; they are, however in the same group on the periodic table as Ni – a much cheaper metal. Moreover, there is already evidence that alloying with Ni may be fruitful: past work has observed strong CO_2 adsorption and activation on CuNi NPs in a several different chemical orderings[108]. Therefore, alloying Cu, Ag, and Au with Ni may be a promising first choice in the screening process.

5.2.8 Understanding the Effect of CE$_{local}$, MADs, and ΔIP

Because our random forest shows BE to be a function of the CE$_{local}$, MADs, and ΔIP values, it is relevant to discuss the effect each of these three properties has on BE, which may be helpful in further tuning the binding site. We begin by discussing ΔIP, showing its relation to BE in Figure 5.10.
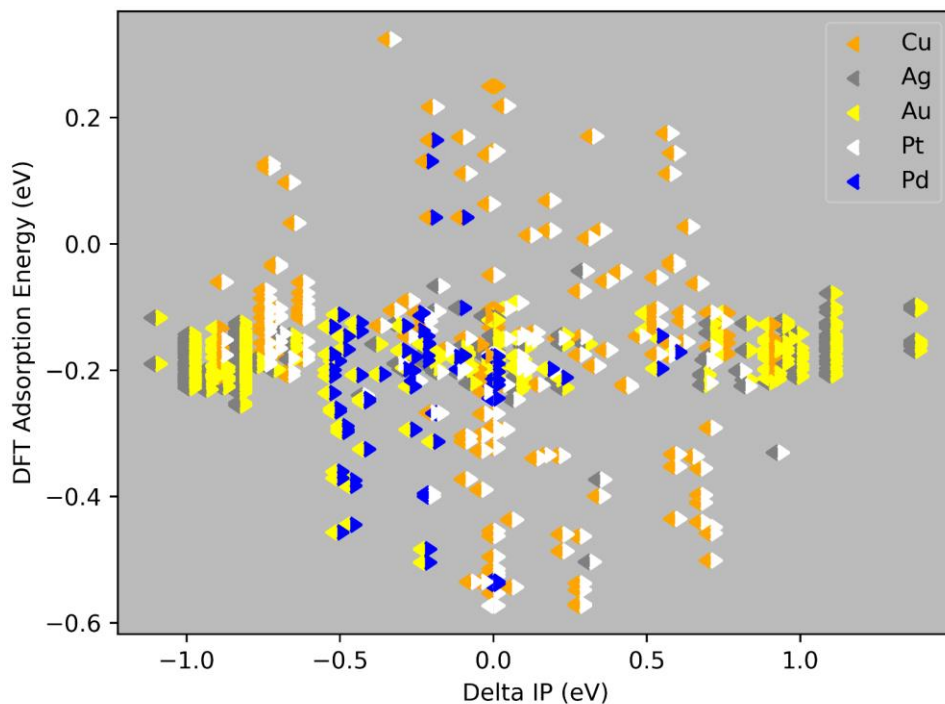
Figure 5.10: CO$_2$ BE as a function of ΔIP. To represent the bimetallic nature of these NPs, symbols are dual-colored. As an example, if a symbol was orange and blue, this would represent a CuPd NP.

We observe the most extreme BEs to be located in the region between -0.5 and 0.75 eV. Surprisingly, this is in contrast to our original development of the ΔIP parameter[112], wherein we observed a negative ΔIP to be necessary for CO$_2$ adsorption. In the current study, both strong and weak binding occurs regardless of the sign of ΔIP.

In its original form, ΔIP helped rationalize the charge transfer between a single dopant atom and the rest of an idealized Cu$_{54}$M (M = dopant) system (see Figure 2.6 and surrounding discussion in Section 2.3.1). Hence, part of this deviation from our original findings may be the result of investigating more-complex systems with intermediate compositions instead of highly-idealized singly-doped systems. Additionally, CO$_2$ can bind in a wide range of adsorption configurations,

and the C and O are oftentimes able to bind to different metal atoms (e.g. C may be bound to a metal of one type while O is simultaneously bound to a metal of another type).

Investigating Figure 5.10 more closely, we also see that most of the deviation is in the CuPt system. Indeed, if this system was not present in the dataset, we would observe a highly similar behavior to the one we found[112] in Section 2.3.1. It is easy to see this in Figure 5.10: if CuPt was not in the dataset, then outside of a handful of points from Pd and AgPt, nearly all of the strong binding would seem to be in regions of negative ΔIP.

Unfortunately, this disruption of the trend also makes regression more challenging: it is more difficult to predict the BE when both the strongest and weakest BEs are located on the same value of ΔIP. Ultimately, this is a likely reason both OLS and our neural network failed to correctly identify the trend in binding. Turning our attention to the other two features in our dataset, in Figure 5.11, we show MADs and CE_{local}. In both cases, we observe generally poor linear trends as well. With CE_{local} (Figure 5.11 B), we see the most negative values of BE occurring at roughly CE_{local} = -6 eV, but this is also where some of the least-favorable values occur. To a lesser extent, even MADs seems to demonstrate this behavior for values below -6 eV and above -4 eV – where a significant portion of the strong BEs are also located.

Overall, this is consistent with the findings in our original OLS work[24] with CH_3, CO, and OH. Part of that study included the training of an adsorption model for all three adsorbates on a variety of different monometallic systems – Cu, Ag, Au, Ni, Pd, Pt, Rh, and Ir (see discussion in Section 3.2.2.2). Specifically, even in the idealized case of monodentate, top-site adsorption to monometallic slabs, we found OLS to still have plenty of trouble in identifying the underlying trends of adsorption. Instead, in that investigation, we recommended models be fit on a per-group basis, as we observed enhanced performance when separate models were trained for adsorption to

114

d7, d8, or d9 metals. We conjectured at that point that further improvements would require the inclusion of a descriptor capable of distinguishing between the different columns of the periodic table – a descriptor we have yet to become aware of. We may be seeing a similar effect here, where some additional descriptor could be necessary to help distinguish between adsorption to different types of bimetallic pair.
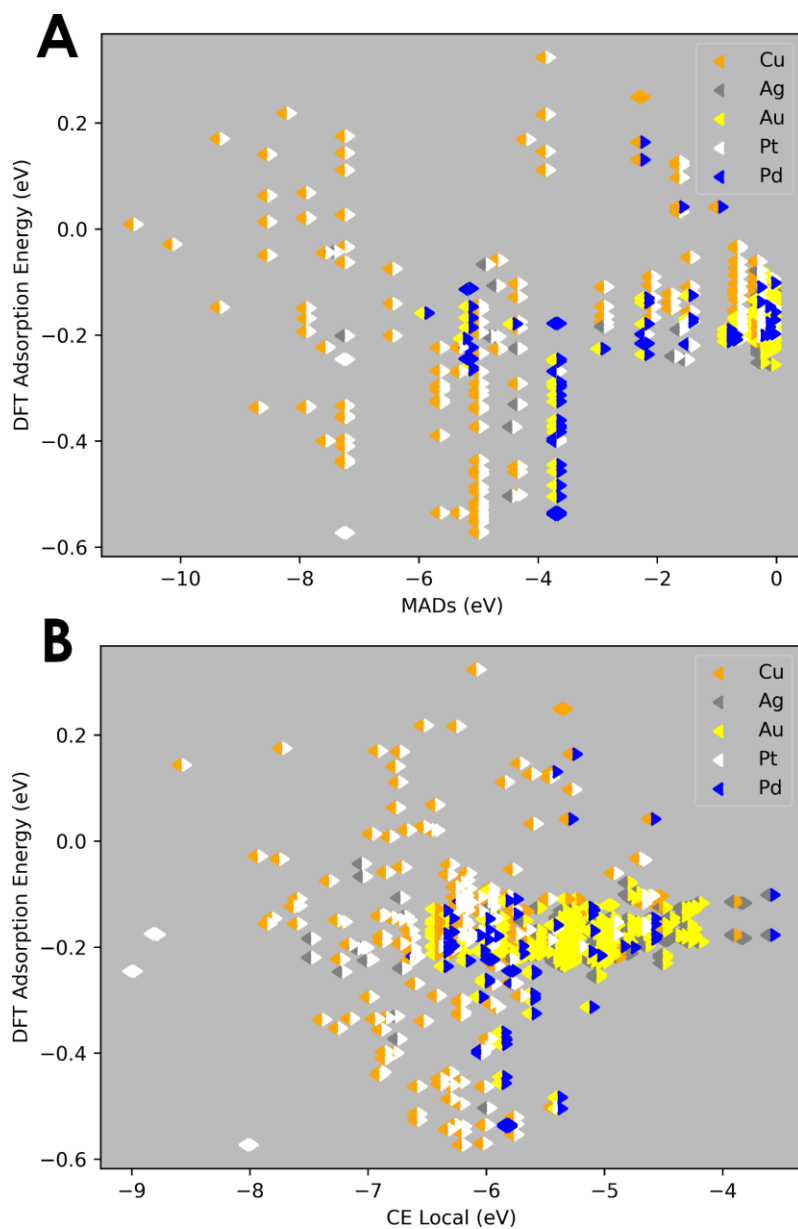
Figure 5.11: CO₂ BE as a function of different descriptors. (A) MADs, and (B) CE_local. To represent the bimetallic nature of these NPs, symbols are dual-colored. As an example, if a symbol was orange and blue, this would represent a CuPd NP.

Overall, it is likely that the flexibility of the random forest to create decision boundaries within the dataset that allows the production of such a high-quality fit where both the OLS and

116

neural network failed. In general, this highlights the many challenges associated with this specific regression problem, where CO_2 can bind to a wide range of sites in a wide range of configurations. Ultimately, improved models of CO_2 adsorption may require new features we have not yet found, or the incorporation of features which are more computationally-challenging to determine (such as an energy predicted by some interatomic potential).

5.3 Conclusions

In this work, we developed a new workflow for the high-throughput screening of metal NPs for any generic property, investigating the adsorption of CO_2 as a proof-of-concept. We begin by developing AutoAdsorbatePlacement, an algorithm to identify potentially unique binding sites in arbitrary clusters. We continue by collecting a dataset of 661 unique adsorption energies across 18 NPs, containing all monometallic and bimetallic systems of Cu, Ag, Au, Pd, and Pt. This enables us to leverage random forest regression to determine the importance of several tabulated descriptors relevant to this problem.

We extend two descriptors in our previous adsorption model[24] and our ΔIP criterion[112] to multi-atom binding sites, and confirm via our random forest that this extension yields a good RMSE. We then train a neural network using these three descriptors to predict BEs of CO_2 to any arbitrary site. Finally, we visualize the hypothetical chemical space defined by the minimum / maximum of our feature space, and using the random forest reveal a highly complex hypothetical chemical space, underscoring the difficulties presented by this regression problem.

As part of our high-throughput screening workflow, we 1) optimize a set of NPs via CE Expansion, 2) automate the calculation some relevant property (adsorption) to generate a training

set, 3) parameterize a machine-learned description of the property of interest (adsorption), and 4) conduct a screening of the chemical space to find regions where NPs with the desired property are likely to exist.

Although at this formulation, our model doesn't identify specific metals which may be of promise, it does show that other bimetallic combinations are likely to exist which can have strong binding similar to those we identified for materials doped with Pt and Pd (see the discussion surrounding Figure 5.9). Specifically, we identify combinations of CE_{local}, MADs, and ΔIP which should be targeted to maximize the CO_2 BE.

6.0 Future Work

In Chapters 2.0 and 3.0, we identified physical properties relevant to adsorption to NPs, and developed a first-of-its-kind universal model of small molecule adsorption to NPs. In Chapter 4.0, we developed CE Expansion, an open-source GA which optimizes the chemical ordering of NPs of any size, shape, or composition. Finally, in Chapter 5.0 we tied together all this work, generating a large dataset of BEs, and trained several machine learning models to predict CO_2 adsorption across tens of thousands of low-energy NP systems.

Although the final developed model does not directly predict specific combinations of metals which are likely to result in strong CO_2 adsorption, it still guides our study toward systems with specific values of CE_{local}, MADs, and ΔIP. The clear next step in this work is to generate a screening approach which will sieve different compositions and chemical orderings toward these targeted physical properties. As a further proof of concept, a simple grid search over several bimetallic alloys of Cu, Ag, Au, Ni, Pd, Pt, Rh, and Ir (choosing ones which were not included in this work) would be of interest. The chosen alloy could then be optimized via CE Expansion and then screened using the random forest model we trained in Section 5.2.4. Additional DFT work on a selection of these systems could also be used to further refine the model. In the end, this could lead to an iterative process where DFT is used to investigate candidates identified by the model, and this DFT data is in turn added to the model's training set. The benefits would be two-fold: this would simultaneously target computationally-expensive DFT calculations to only systems which are the most promising, while simultaneously improving model's predictions over time.

Another natural continuation of this work would be in developing a better understanding of the effects ligands have on NP stability. Specifically, colloidal NPs have become popular

subjects of research in recent years due to their experimental synthesis control, and the choice of ligand is known to have a great effect on NP stability[211]. By further refining our adsorption modeling to the point where ligand adsorption energies can be incorporated, we would be able to use the adsorption energy of the ligand as a descriptor of NP CE. Specifically, a fundamental assumption of the BCM is that a NP's CE can be approximated as the sum of bond energies within the NP. Because ligands have been known to have a stabilizing effect on the NPs surfaces[212], we propose including the BE between a ligand and a NP surface atom into the sum of bond energies considered by the BCM. This would further lower the NP CE, which may help account for the stabilizing effect of ligands. Additionally, this avenue of research could be further enhanced by incorporating the interaction of neighboring ligands – expanding upon an idea which was recently employed to great success by incorporating CO surface coverage to predict the stability of Rh NPs[213].

Overall, we propose further work which continues enhancing both major parts of this thesis: adsorption and stability modeling. By improving our models of NP stability, we are able to arrive at computational model systems which are closer to those of experiment (see Section 4.2.1), thus accelerating materials discovery. Additionally, by improving our modeling capabilities of adsorption, we provide useful tools for the high-throughput screening of NPs based on their physical properties, again accelerating materials discovery by focusing the expensive characterization techniques to only the most-promising candidate systems. Combined, both avenues of improvement continue honing the tools we have created in this thesis to sieve through the tremendously-large space of bimetallic NPs for targeted catalytic applications.

Appendix A List of Abbreviations

ANAKIN-ME .... Accurate NeurAl networK engINe for Molecular Energies

ANI-1 .... ANAKIN-ME

ASE .... Atomic Simulation Environment

BCM .... Bond-Centric Model

BE .... Binding Energy

BFGS .... Broyden-Fletcher-Goldfarb-Shanno Algorithm

CCCBDB .... Computational Chemistry Comparison and Benchmark DataBase

CCSD(T) .... Coupled Cluster Single Double and Perterbative Triple

CE .... Cohesive Energy

CE_{local} .... Local Cohesive Energy

CN .... Coordination Number

CV .... Cross-Validated

ΔIP .... Ionization Potential Difference

DENDRAL .... Dendritic Algorithm

DFT .... Density-Functional Theory

DOF .... Degrees of Freedom

EA .... Electron Affinity

EANN .... Embedded Atom Neural Network

EDS .... Energy Dispersive X-Ray Spectroscopy

EE .... Excess Energy

EMT .... Effective Medium Theory

NMR .... Nuclear Magnetic Resonance

NP .... Nanoparticle

OLS .... Ordinary Least-Squares

OQMD .... Open Quantum Materials Database

PBE .... Perdew-Burke-Ernzerhov

QSAR.. Quantitative Structure-Activity Relation

RI/MARIJ .... Resolution of Identity / Multipole-Accelerated RI

RMSE.. Root Mean Squared Error

RPBE.. Revised PBE

SCF ....Self-Consistent Field

SchNOrb .... SchNet for Orbitals

STD..Standard Deviation

STEM...Scanning Transmission Electron Miscroscopy

TPD... Temperature-Programmed Desorption

USPEX..Universal Structure Predictor: Evolutionary Xtallography

XPS .... X-Ray Photoelectron Spectroscopy

XRD .... X-Ray Diffraction

ZPE ....Zero-Point Energy

Appendix B Supporting Information for "Designing Copper-Based Bimetallic Nanoparticles for CO$_2$ Activation"

The content of this appendix is adapted from the supporting information of Dean, J.; Yang, Y.; Austin, N.; Veser, G.; Mpourmpakis, G. "Design of Copper-Based Bimetallic Nanoparticles for Carbon Dioxide Adsorption and Activation." ChemSusChem 2018, 11 (7) 1169-1179.

Appendix B.1 CO$_2$ TPD on typical support materials
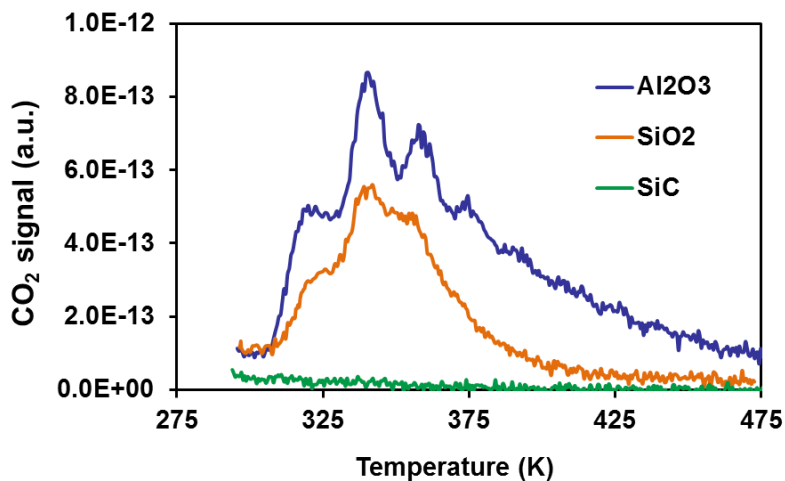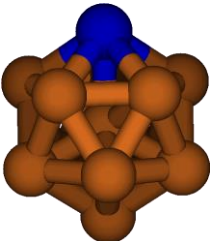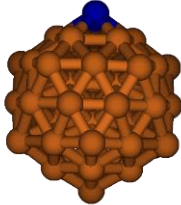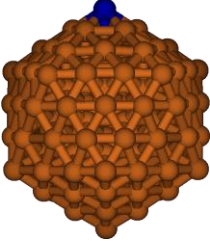
Figure B.1: CO$_2$ TPD spectra of several common catalyst supports, among which only SiC shows no ability to absorb CO$_2$.

In order to investigate CO$_2$ sorption on CuZr bimetallic NPs, CO$_2$ TPD from several common catalyst support materials was first conducted in order to eliminate or minimize support

124

contributions to CO_2 adsorption. As shown in Figure B.1, both silica and alumina supports can adsorb CO_2 to a significant degree, while SiC shows no detectable CO_2 sorption. Therefore, SiC was used as the catalyst support structure to allow for unambiguous attribution of any observable CO_2 adsorption to the supported CuZr NPs.

Appendix B.2 CO_2 Chemisorption onto Differently-Sized CuZr Nanoparticles

Table B.1: CO_2 binding energies (in eV) on icosahedral 13/55/147-atom CuZr (Zr CN=6) NPs.

Species	Structure	CO_2 BE (eV)
$Cu_{12}Zr$		-2.1
$Cu_{54}Zr$		-1.6
$Cu_{146}Zr$		-1.7

Appendix B.3 ZrO₂ (1̄11) Slab Calculations

Since ZrO_2 NPs were experimentally observed in the synthesis of CuZr NPs (see Results and Discussion of the main document), we examined CO_2 adsorption on ZrO_2 by choosing the lowest-energy surface ($\bar{1}11$) termination of monoclinic ZrO_2 from literature[214, 215]. In order to calculate the BE of CO_2 on the ZrO_2 ($\bar{1}11$) surface, the unit cell vectors of ZrO_2 were initially optimized (Table B.1).

Table B.2: Unit cell parameters of the monoclinic ZrO₂ unit cell, in angstrom in the case of A/B/C, and in degrees in the case of $\alpha/\beta/\gamma$.

Cell Parameter (units)	Measurement
A (Å)	5.161
B (Å)	5.228
C (Å)	5.362
α (degree)	90.032
β (degree)	90.573
γ (degree)	90.004

Following the optimization of the unit cell, we next cleaved the ($\bar{1}11$) surface while ensuring ZrO_2 stoichiometry was maintained, and froze the bottom two layers of Zr and O. The slab was allowed to relax with a 10 Å vacuum, resulting in the surface found below in Figure B.2.
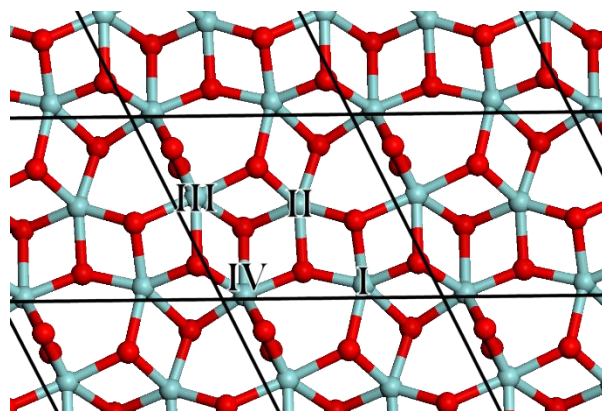
Figure B.2: Top view of the first layer of the optimized ZrO_2 ($\overline{1}11$) surface, with the unit cell overlaid in black. Key: Red=O, Blue=Zr. Numerals "I", "II", "III", and "IV" represent the four unique Zr atoms on the surface. Periodic cell parameters (in Å for A/B/C and degrees for $\alpha/\beta/\gamma$) are as follows: A = 6.796, B = 7.346, C = 22.409 (inclusive of a 10 Å vacuum between each slab), $\alpha = 90$, $\beta = 90$, $\gamma = 63.808$.

Additionally, because the ZrO_2 surface atoms in our synthesized particles may exist in a hydroxylated form, we chose to investigate the hydration of (most acidic) sites III and IV from the dissociation of water. The optimized structures can be found below in Figure B.3.
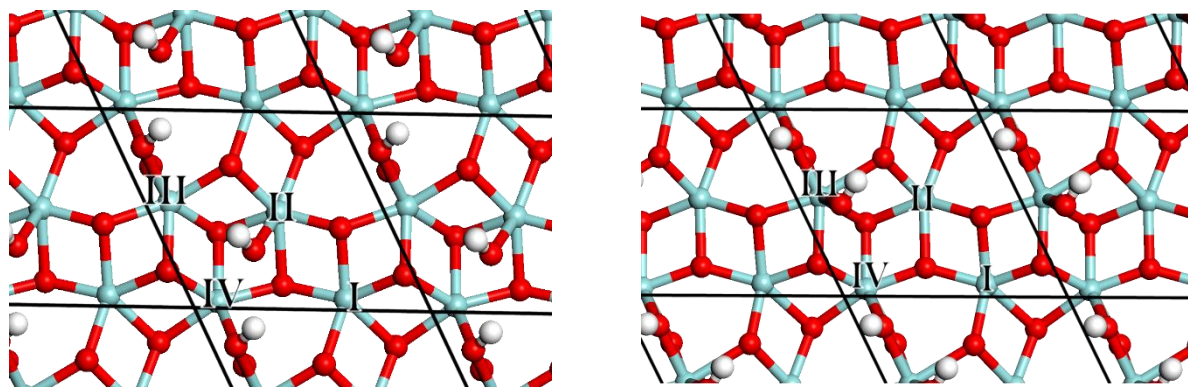
Figure B.3: Top view of the first layer of the optimized hydroxylated ZrO_2 ($\bar{1}11$) surface, with the unit cell overlayed in black. Left: Hydroxylated IV-site. Right: Hydroxylated III site. Key: Red=O, Blue=Zr. Numerals "I", "II", "III", and "IV" represent the four unique Zr atoms on the surface. Periodic cell parameters (in Å for A/B/C and degrees for $\alpha/\beta/\gamma$) are as follows: A = 6.796, B = 7.346, C = 22.409 (inclusive of a 10 Å vacuum between each slab), α = 90, β = 90, γ = 63.808.

Following the optimization of the slabs, we investigated the adsorption of CO_2 onto the non-hydrated surface first. CO_2 was initially placed parallel to the surface approximately 1.5 Å away from Zr atoms I, II, III, and IV. The bottom two layers of Zr and O were again frozen, a vacuum of 10 Å between each slab was included, and the geometry of the system optimized). In all cases, physisorption was observed, with energies ranging from -0.22 eV to -0.27 eV. Following this set of calculations, we proceeded to place CO_2 onto the surface in a bent configuration to form carbonate. Only two sites (III and IV) were able to bind the CO_2 as carbonate; the other two sites (I and II) converged to a physisorbed CO_2. The BE of carbonate can be found in Table B.3.

Table B.3: Lowest-energy CO$_2$ BE (calculated via Equation 2.2) on the ZrO$_2$ ($\bar{1}$11) slab when CO$_2$ is placed in an initially bent configuration. See Figure B.2 for a visual overview of the four unique Zr atoms on the surface of the ZrO$_2$ slab.

Species	Binding Site	CO$_2$ BE (eV)	Zr Charge (e$^-$)
ZrO$_2$ ($\bar{1}$11)	III	-0.61	2.5
ZrO$_2$ ($\bar{1}$11)	IV	-0.81	2.4

Of note is the fact that we observe a favorable formation of the carbonate ion on the surface of ZrO$_2$ in the case of sites III and IV (See Figure B.2 for Site IV). This is in agreement with the results of Larmier et. al[93] who observed the formation of carbonate on the ZrO$_2$ ($\bar{1}$11) surface.

Using a physisorbed CO$_2$ on the ZrO$_2$ ($\bar{1}$11) IV site (BE$_{CO2}$ = -0.23 eV) as the starting point for an NEB calculation and the adsorbed carbonate on the IV site (BE$_{CO2}$ = -0.81 eV) as the ending point, we calculate an energetic barrier of +0.13 eV (relative to the physisorbed state) for this formation of carbonate to occur. These states are shown in Figure B.4 and the reaction pathway in Figure B.5. An NEB calculation was also performed going from a physisorbed CO$_2$ above the III site to a carbonate ion on this position, but this was observed to be barrierless.

-0.23 eV   -0.81 eV   -0.57 eV

Physisorption   Carbonate   Bicarbonate
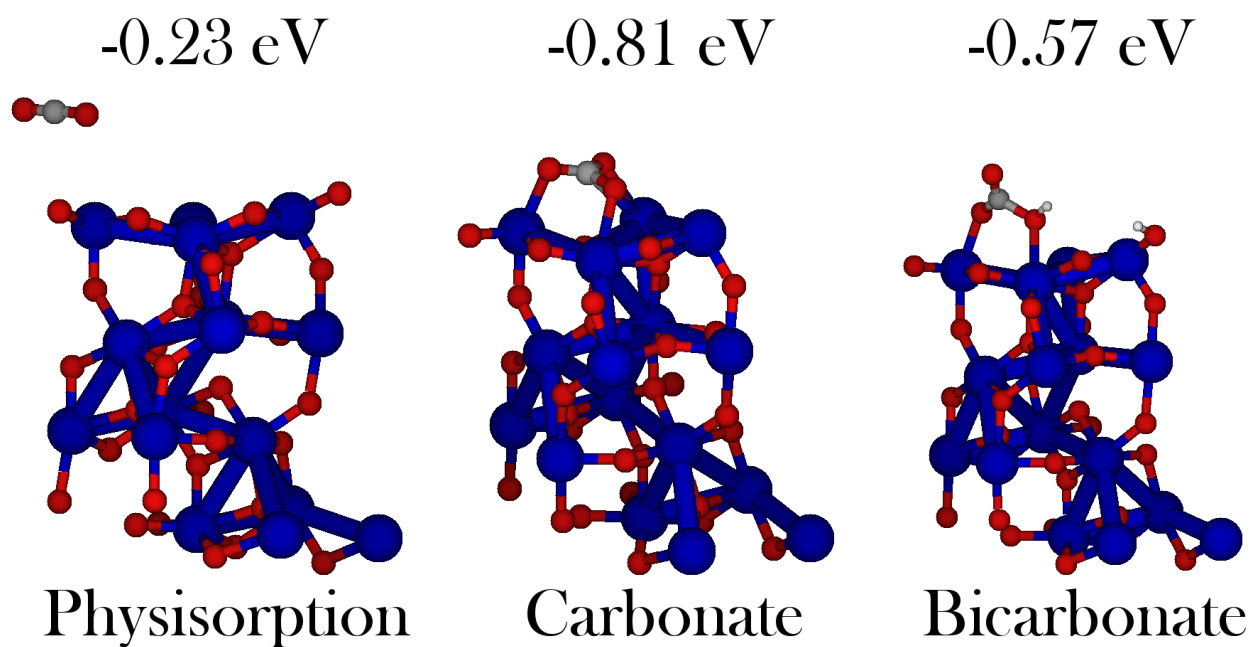
Figure B.4: A focus on the IV site of ZrO$_2$ ($\bar{1}$11), with the physisorbed, carbonate-ion-formation, and bicarbonate formation states labeled below. Above each image we denote the CO$_2$ BE. Key: Red = O, Blue = Zr, Grey = C.
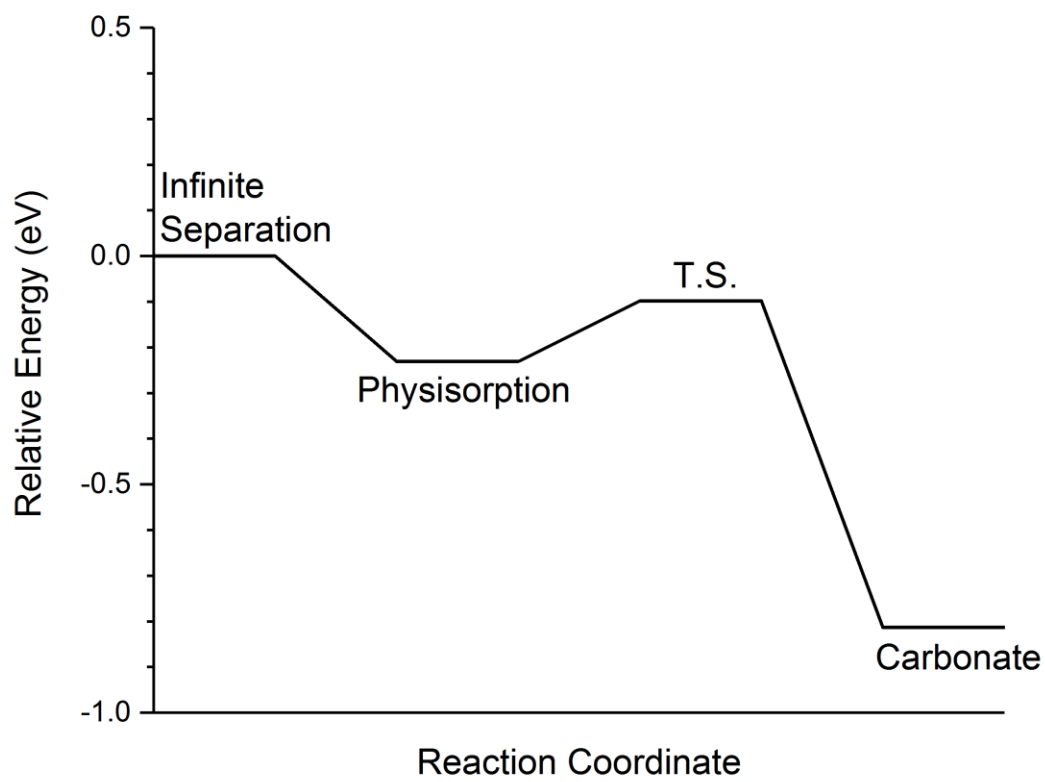
Figure B.5: Reaction coordinate diagram of CO_2 chemisorption onto the ZrO_2 ($\bar{1}11$) IV site. "Infinite separation" represents the infinite separation of the ZrO_2 slab and CO_2, which we set as the zero in the energy diagram. "T.S." stands for transition state. The CO_2 desorption energy from the carbonate to the gas phase is +0.81 on the site IV of ZrO_2 ($\bar{1}11$) surface.

Finally, utilizing the hydrated surface shown in Figure B.3, we bent the CO_2 initially and placed it next to a Zr site in order to form bicarbonate. In the case of the III site, the existence of bicarbonate was not favorable, and CO_2 moved away from the surface to a physisorbed state. In the case of IV site, the BE was calculated to be -0.57 eV (see structure of bicarbonate in Figure B.4).

These results are again in agreement with those of Larmier et al[93], who suggested the formation of bicarbonate on the hydroxylated sites of ZrO_2 ($\bar{1}11$). Finally, utilizing a physisorbed CO_2 on the hydroxylated ZrO_2 ($\bar{1}11$) IV site as the starting point for an NEB calculation, and the bicarbonate on the IV site as the end point, we observe a barrierless adsorption of CO_2 as bicarbonate to the IV site.

Appendix C Supporting Information for "Unfolding Adsorption on Metal Nanoparticles:

Connecting Stability with Catalysis"

The content of this appendix is adapted from the supporting information of Dean, J.; Taylor, M. G.; Mpourmpakis, G. "Unfolding Adsorption on Metal Nanoparticles: Connecting Stability with Catalysis." Science Advances 2019, 5 (9) eaax5101.

Appendix C.1 Thermodynamic Data

In the main document, we outline a proposed model to describe the binding of an adsorbate to a metal (Equation 3.4). Generally, we expect to be a term $Stab_{Ads}$ describing the stability of the adsorbate, $Stab_{NP}$ for the stability of the NP, Int_{Ads-M} for the tendency of the adsorbate and metal to interact, and finally $Stab_{site}$ for the stability of the local site. We choose CE_{local} as our descriptor for the local site stability and outline our decision in the main document. However, this still leaves three other descriptors to account for in our model: stability of the adsorbate and NP, as well as the interaction between the adsorbate and a metal atom. We present a description of our choice of properties for these descriptors, as well as a justification for choosing them as follows.

Appendix C.1.1 $Stab_{Ads}$: Ionization Potential and Electron Affinity (IPEA)

For $Stab_{Ads}$, we relate the adsorbate stability in the gas phase to the chemical potential. It has been shown that within the context of HSAB theory, the negative average of the IP and EA

can be used as a first-order approximation to the chemical potential [153]. Experimental values for both IP and EA for small molecules are tabulated in literature, and even in the absence of experimental data, they can be readily calculated [216]. The data we used to calculate this term comes primarily from tabulated experimental data on the NIST WebBook. In the case of ammonia, in the absence of an experimental EA value in the NIST WebBook, we therefore used the NIST CCCBDB, choosing the entry corresponding to the full CCSD(T) method with a cc-pVQZ basis set. The reference data we used can be found in Table C.1.

Table C.1: IPs and EAs for the methyl radical, carbon monoxide, the hydroxyl radical, and ammonia. IPEA is calculated as the negative of the average of IP and EA. All energies are reported in electron-volts (eV).

Adsorbate	IP (eV)	EA (eV)	IPEA (eV)
Methyl ($\cdot CH_3$)	9.843 [217]	0.08 [218]	-4.961
Carbon Monoxide (CO)	14.0142 [219]	1.32608 [220]	-7.670
Hydroxyl ($\cdot OH$)	1.82767 [221]	13.017 [222]	-7.422
Ammonia (NH_3)	10.02 [223]	-2.338 [41]	-3.841

Appendix C.1.2 Stab$_{NP}$: Cohesive Energy of the Nanoparticle (CE$_{NP}$)

We use the CE of the isolated NP in the gas phase as our descriptor for its stability. CE is defined as the energy required to separate each atom of the NP to an infinite distance, scaled by the number of atoms in the NP. The CE of a monometallic NP is mathematically related to its electronic energy, as shown in Equation C.1.

$$CE_{NP} = \frac{E_{NP} - n * E_M}{n}$$

Equation C.1

In Equation C.1, *n* represents the number of atoms in the NP. *CE* therefore serves as an indirect measure of NP energetics/stability. To calculate CE of the NPs, we use the accelerated BCM, which has been shown to trend with DFT CEs over both monometallic and bimetallic NPs

of different sizes and shapes [151]. The BCM requires no parameter tuning, is orders of magnitude faster than DFT, and is readily extendable to multiple NP systems. In the case of a monometallic NP, the BC-model CE of the NP is represented as Equation C.2.

$$\text{Stab}_{NP} = CE_{NP} = \frac{CE_{bulk}}{n * \sqrt{CN_{bulk}}} \sum_{i=1}^{n} \sqrt{CN_i} \qquad \text{Equation C.2}$$

In Equation C.2, CE_{NP} is the CE of the NP, CE_{bulk} is the bulk CE of the metal, CN_i is the CN of atom i, and CN_{bulk} is the CN of the metal in bulk.

Table C.2: BC-model-calculated cohesive energies of the relevant NPs we investigated.

Morphology	Formula	CE (eV)
Cube	Cu_{172}	-2.946
Cuboctahedron	Cu_{55}	-2.790
Cuboctahedron	Cu_{147}	-2.990
Icosahedron	Cu_{55}	-2.912
Icosahedron	Cu_{147}	-3.082
Cube	Ag_{172}	-2.490
Cuboctahedron	Ag_{55}	-2.359
Cuboctahedron	Ag_{147}	-2.527
Icosahedron	Ag_{55}	-2.466
Icosahedron	Ag_{147}	-2.605
Cube	Au_{172}	-3.216
Cuboctahedron	Au_{55}	-3.046
Cuboctahedron	Au_{147}	-3.264
Icosahedron	Au_{55}	-3.186
Icosahedron	Au_{147}	-3.364
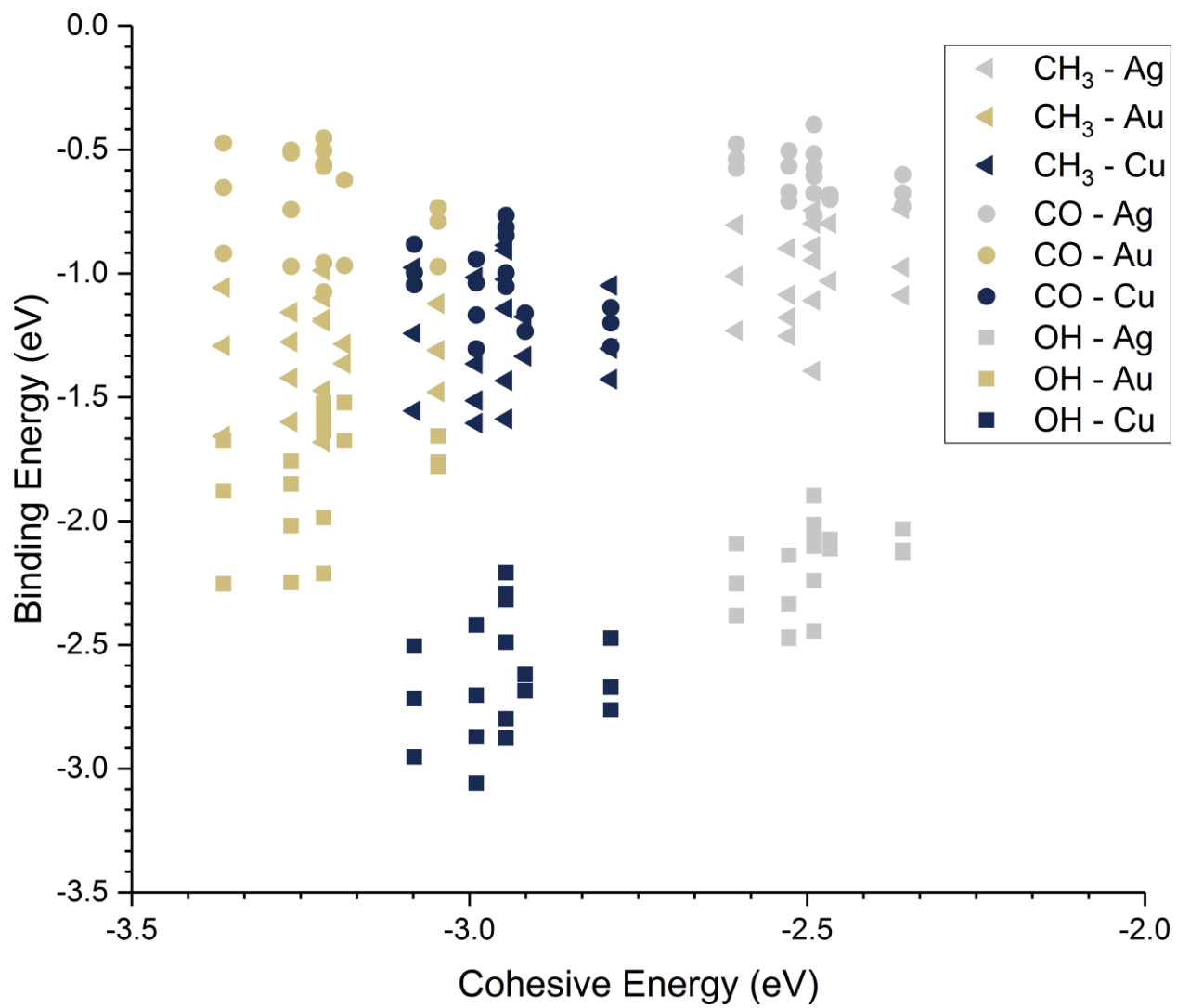
Figure C.1: Adsorbate binding energies versus BC-model-calculated NP cohesive energies.

Appendix C.1.3 IntAds-M: Gas-Phase Metal-Adsorbate BE (MADs)

For the Int_{Ads-M} descriptor we use DFT to calculate a BE in the gas phase between the adsorbates and a single metal atom. This calculation represents a simple but accurate method for screening only metal-adsorbate interactions, as no metal-metal nor adsorbate-adsorbate bonds are present to influence the metal-adsorbate interaction. We highlight that this is the only DFT calculation required to model new metal-adsorbate interactions with the functional form of Equation 3.4. These DFT calculations are significantly faster to calculate than full adsorbate-NP DFT calculations. Therefore, the Int_{Ads-M} descriptor we use is that found in Equation C.3.

$$Int_{Ads-M} = E_{bind,Ads-M} \equiv MADs \qquad \text{Equation C.3}$$

We recognize that MADs, although fast to calculate, may encounter some challenges within DFT. We try to be as accurate as possible in our calculations by investigating several spin states and using an unrestricted Kohn-Sham scheme (see Computational Information). The same descriptor has also been used by Roling et al [157] in the development of their adsorption model. Alternatively, one could use experimental affinities between a metal and adsorbate as the MADs descriptor. The physics (type of descriptors defining the model) should not depend on the particular approach used in determining the descriptors, although the regression coefficients will likely change.

Appendix C.1.4 Stabsite: The local cohesive energy (CElocal)

Beyond simple metal-adsorbate interaction strength, previous work indicates that geometry of the binding site has a strong effect on the binding of single adsorbate (E_{Ads-NP} in Equation 3.1). For example, it was observed that the BE of carbon monoxide to Au NPs depends on both the CN

of the metal participating in the binding interaction as well as the local curvature of the NP surface [23]. For the site stability descriptor, $Stab_{site}$, we choose to take a modified form of the BCM wherein we only consider bonds directly connected to the metal atom participating in the adsorption interaction. We report in Table C.3 calculated values for this descriptor for every binding interaction we investigated, including periodic slab dataset [157].

Table C.3: Calculated CE$_{local}$ for relevant adsorption sites. In case of NPs where more than one site share the same CN (the cubes), we denote different binding sites by including a subscript and a unique identifier for the site. In the case of slabs where more than one site share the same coordination, we only report the CN of the site investigated.

Morphology	Formula	Facet (Slabs Only)	Site	CE$_{local}$ (eV)
Cube	Ag$_{172}$	NA	3	-2.378
Cube	Ag$_{172}$	NA	5	-3.354
Cube	Ag$_{172}$	NA	8_1	-4.811
Cube	Ag$_{172}$	NA	8_2	-4.596
Cube	Ag$_{172}$	NA	8_3	-4.596
Cube	Ag$_{172}$	NA	8_4	-5.057
Cuboctahedron	Ag$_{55}$	NA	5	-3.438
Cuboctahedron	Ag$_{55}$	NA	7	-4.451
Cuboctahedron	Ag$_{55}$	NA	8	-4.679
Cuboctahedron	Ag$_{147}$	NA	5	-3.438
Cuboctahedron	Ag$_{147}$	NA	7	-4.354
Cuboctahedron	Ag$_{147}$	NA	8	-4.638
Cuboctahedron	Ag$_{147}$	NA	9	-5.224
Icosahedron	Ag$_{55}$	NA	6	-3.837
Icosahedron	Ag$_{55}$	NA	8	-4.8
Icosahedron	Ag$_{147}$	NA	6	-3.837
Icosahedron	Ag$_{147}$	NA	8	-4.719
Icosahedron	Ag$_{147}$	NA	9	-5.099
Cube	Au$_{172}$	NA	3	-3.072

Table C.3 (continued).

Cube	Au$_{172}$	NA	5	-4.332
Cube	Au$_{172}$	NA	8_1	-6.214
Cube	Au$_{172}$	NA	8_2	-5.936
Cube	Au$_{172}$	NA	8_3	-5.936
Cube	Au$_{172}$	NA	8_4	-6.531
Cuboctahedron	Au$_{55}$	NA	5	-4.44
Cuboctahedron	Au$_{55}$	NA	7	-5.749
Cuboctahedron	Au$_{55}$	NA	8	-6.044
Cuboctahedron	Au$_{147}$	NA	5	-4.44
Cuboctahedron	Au$_{147}$	NA	7	-5.624
Cuboctahedron	Au$_{147}$	NA	8	-5.99
Cuboctahedron	Au$_{147}$	NA	9	-6.746
Icosahedron	Au$_{55}$	NA	6	-4.956
Icosahedron	Au$_{55}$	NA	8	-6.199
Icosahedron	Au$_{147}$	NA	6	-4.956
Icosahedron	Au$_{147}$	NA	8	-6.095
Icosahedron	Au$_{147}$	NA	9	-6.585
Cube	Cu$_{172}$	NA	3	-2.814
Cube	Cu$_{172}$	NA	5	-3.968
Cube	Cu$_{172}$	NA	8_1	-5.692
Cube	Cu$_{172}$	NA	8_2	-5.438
Cube	Cu$_{172}$	NA	8_3	-5.438
Cube	Cu$_{172}$	NA	8_4	-5.983
Cuboctahedron	Cu$_{55}$	NA	5	-4.067
Cuboctahedron	Cu$_{55}$	NA	7	-5.266
Cuboctahedron	Cu$_{55}$	NA	8	-5.536
Cuboctahedron	Cu$_{147}$	NA	5	-4.067
Cuboctahedron	Cu$_{147}$	NA	7	-5.151
Cuboctahedron	Cu$_{147}$	NA	8	-5.487
Cuboctahedron	Cu$_{147}$	NA	9	-6.18
Icosahedron	Cu$_{55}$	NA	6	-4.54
Icosahedron	Cu$_{55}$	NA	8	-5.679
Icosahedron	Cu$_{147}$	NA	6	-4.54
Icosahedron	Cu$_{147}$	NA	8	-5.583

Table C.3 (continued).

Icosahedron	Cu$_{147}$	NA	9	-6.032
Icosahedron	Cu$_{22}$Ag$_{33}$	NA	6_1	-4.137
Icosahedron	Cu$_{22}$Ag$_{33}$	NA	6_2	-3.993
Icosahedron	Cu$_{22}$Ag$_{33}$	NA	8_1	-5.207
Icosahedron	Cu$_{22}$Ag$_{33}$	NA	8_2	-4.916
Icosahedron	Cu$_{31}$Ag$_{24}$	NA	6_1	-4.148
Icosahedron	Cu$_{31}$Ag$_{24}$	NA	6_2	-4.344
Icosahedron	Cu$_{31}$Ag$_{24}$	NA	8_1	-5.449
Icosahedron	Cu$_{31}$Ag$_{24}$	NA	8_2	-5.02
Cuboctahedron	Rh$_{55}$	NA	5	-6.7
Cuboctahedron	Rh$_{55}$	NA	7	-8.676
Cuboctahedron	Rh$_{55}$	NA	8	-9.121
Icosahedron	Rh$_{55}$	NA	6	-7.479
Icosahedron	Rh$_{55}$	NA	8	-9.356
Slab	Ag	100	4	-2.839
Slab	Ag	100	4	-2.78
Slab	Ag	100	6	-3.858
Slab	Ag	100	6	-3.715
Slab	Ag	100	8	-4.826
Slab	Ag	100	8	-4.596
Slab	Ag	111	3	-2.283
Slab	Ag	111	5	-3.449
Slab	Ag	111	7	-4.3
Slab	Ag	111	9	-4.995
Slab	Ag	211	5	-3.289
Slab	Ag	211	6	-3.79
Slab	Ag	211	7	-4.226
Slab	Au	100	4	-3.666
Slab	Au	100	4	-3.591
Slab	Au	100	6	-4.983
Slab	Au	100	6	-4.798
Slab	Au	100	8	-6.233
Slab	Au	100	8	-5.936
Slab	Au	111	3	-2.948
Slab	Au	111	5	-4.454
Slab	Au	111	7	-5.553
Slab	Au	111	9	-6.452

Table C.3 (continued).

Slab	Au	211	5	-4.248
Slab	Au	211	6	-4.895
Slab	Au	211	7	-5.457
Slab	Cu	100	4	-3.358
Slab	Cu	100	4	-3.289
Slab	Cu	100	6	-4.565
Slab	Cu	100	6	-4.395
Slab	Cu	100	8	-5.71
Slab	Cu	100	8	-5.438
Slab	Cu	111	3	-2.701
Slab	Cu	111	5	-4.08
Slab	Cu	111	7	-5.087
Slab	Cu	111	9	-5.91
Slab	Cu	211	5	-3.891
Slab	Cu	211	6	-4.484
Slab	Cu	211	7	-4.999
Slab	Ir	100	4	-6.678
Slab	Ir	100	4	-6.541
Slab	Ir	100	6	-9.077
Slab	Ir	100	6	-8.74
Slab	Ir	100	8	-11.354
Slab	Ir	100	8	-10.813
Slab	Ir	111	3	-5.371
Slab	Ir	111	5	-8.113
Slab	Ir	111	7	-10.116
Slab	Ir	111	9	-11.752
Slab	Ir	211	5	-7.738
Slab	Ir	211	6	-8.917
Slab	Ir	211	7	-9.941
Slab	Ni	100	4	-4.272
Slab	Ni	100	4	-4.185
Slab	Ni	100	6	-5.807
Slab	Ni	100	6	-5.592
Slab	Ni	100	8	-7.264
Slab	Ni	100	8	-6.918
Slab	Ni	111	3	-3.436
Slab	Ni	111	5	-5.191
Slab	Ni	111	7	-6.472
Slab	Ni	111	9	-7.519
Slab	Ni	211	5	-4.951

Table C.3 (continued).

Slab	Ni	211	6	-5.705
Slab	Ni	211	7	-6.36
Slab	Pd	100	4	-3.743
Slab	Pd	100	4	-3.666
Slab	Pd	100	6	-5.088
Slab	Pd	100	6	-4.899
Slab	Pd	100	8	-6.364
Slab	Pd	100	8	-6.061
Slab	Pd	111	3	-3.01
Slab	Pd	111	5	-4.548
Slab	Pd	111	7	-5.67
Slab	Pd	111	9	-6.587
Slab	Pd	211	5	-4.337
Slab	Pd	211	6	-4.998
Slab	Pd	211	7	-5.572
Slab	Pt	100	4	-5.62
Slab	Pt	100	4	-5.504
Slab	Pt	100	6	-7.638
Slab	Pt	100	6	-7.355
Slab	Pt	100	8	-9.555
Slab	Pt	100	8	-9.099
Slab	Pt	111	3	-4.519
Slab	Pt	111	5	-6.827
Slab	Pt	111	7	-8.512
Slab	Pt	111	9	-9.889
Slab	Pt	211	5	-6.512
Slab	Pt	211	6	-7.504
Slab	Pt	211	7	-8.365
Slab	Rh	100	4	-5.533
Slab	Rh	100	4	-5.419
Slab	Rh	100	6	-7.521
Slab	Rh	100	6	-7.241
Slab	Rh	100	8	-9.407
Slab	Rh	100	8	-8.959
Slab	Rh	111	3	-4.45
Slab	Rh	111	5	-6.722
Slab	Rh	111	7	-8.381
Slab	Rh	111	9	-9.737
Slab	Rh	211	5	-6.411
Slab	Rh	211	6	-7.388

Table C.3 (continued).

Slab	Rh	211	7	-8.236

Appendix C.2 "Leave-One-In" Tests

In order to evaluate the generalizability of our model to additional adsorbate-NP systems, we divided our training set up into several categories based on NP size, shape, and chemical composition, as well as by adsorbate. We choose only one category to train on (e.g. if training on 55-atom NPs, we exclude 147- and 172-atom NPs from the training set). The results of these tests can be found in Table C.4, and parity plots can be found in Figure C.2.

Table C.4: Regression statistics for the various "Leave-One-In" tests we performed. The training and test-set RMSEs have units of eV. The intercept, CE_{local}, IPEA, and $E_{bind, M-Ads}$ coefficient estimates are unitless.

Training Set Description	Training Set 10CV RMSE	Test Set RMSE	Intercept	CE_{Local}	IPEA	$E_{Bind, M-Ads}$
Ag NPs	0.1313	0.2201	1.787	-0.117	0.338	0.766
Cu NPs	0.1789	0.2087	2.156	-0.118	0.386	0.704
Au NPs	0.1441	0.2037	0.733	-0.17	0.264	0.601
Icosahedrons	0.1667	0.1813	1.498	-0.161	0.341	0.696
Cuboctahedrons	0.1632	0.1857	1.581	-0.137	0.341	0.69
Cubes (172-atoms)	0.1658	0.2034	1.362	-0.161	0.314	0.653
55-atom NPs	0.1369	0.1917	1.676	-0.091	0.332	0.624
147-atom NPs	0.1481	0.2128	1.418	-0.194	0.348	0.742
Methyl Radical	0.1837	0.6205	-0.542	-0.126	NA	0.484
Carbon Monoxide	0.1386	0.4742	-0.763	-0.085	NA	0.609
Hydroxyl Radical	0.1795	0.5195	-1.261	-0.188	NA	0.65
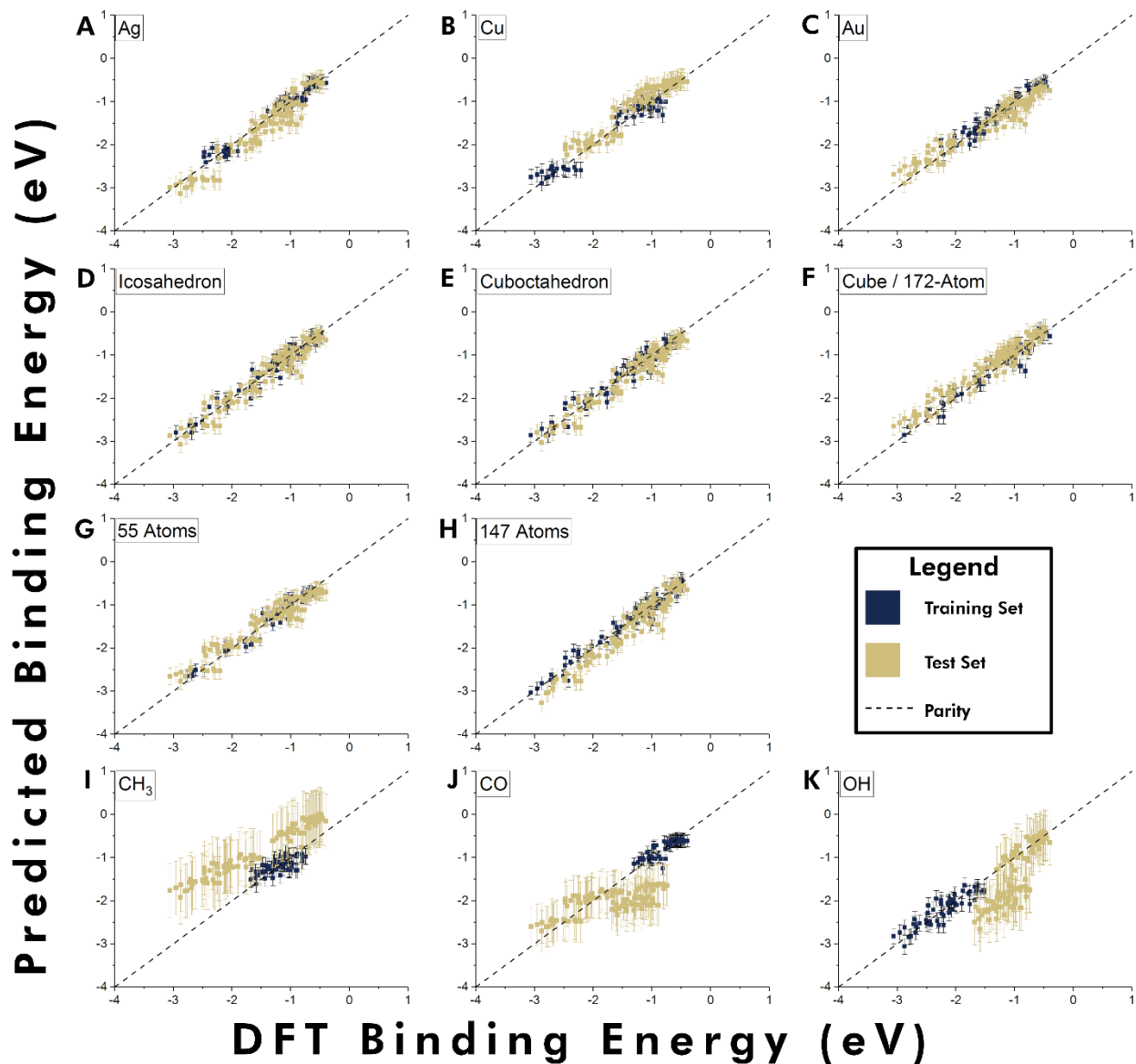
Figure C.2: Parity plots for the various "Leave-One-In" tests we performed. Error bars for the training set are the 10-CV RMSE. Error bars for the test set are the RMSE between the test set model and the DFT prediction. Training sets are exclusively the adsorption interactions on (A) Ag NPs, (B), Cu NPs, (C) Au NPs, (D) Icosahedral NPs, (E) Cuboctahedral NPs, (F) Cube NPs (equivalent with the set of 172-atom NPs), (G) 55-atom NPs, (H) 147-atom NPs, as well as all adsorbate-NP pairs for (I) the methyl radical, (J) carbon monoxide, and (K) the hydroxyl radical.

The larger error bars in the case of the adsorbates can be attributed to the necessity that they underfit. Because only one adsorbate is investigated in S2.1 I, J, and K, it is impossible to sample more than one value of *IPEA*: two chemically-identical adsorbates must have the same value for this descriptor. As a result, there is not enough information in those particular splits of the dataset for the IPEA parameter to be tuned.

Appendix C.3 Investigations on Slabs

Following our initial study of the adsorption of the methyl radical, the hydroxyl radical, and carbon monoxide onto Au, Ag, and Cu NPs, we also investigated the adsorption of these same adsorbate-metal pairs onto the several metallic slabs reported in the slab data[157]. The binding energies are taken from the aforementioned reference, and the local site cohesive energies that we calculated can be found in Table C.3. In order to search for potential descriptors and verify the model descriptors we have already chosen, we utilized three techniques: OLS, LASSO, and Symbolic Regression.
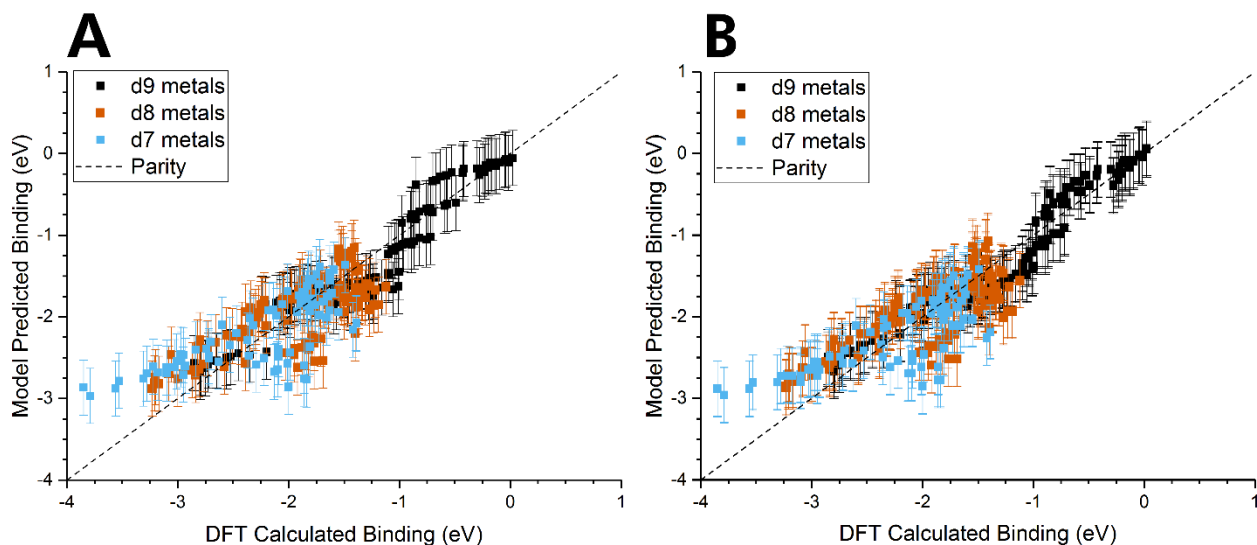
Figure C.3: Characterization of all metal-adsorbate pairs in the slab dataset[157] simultaneously (e.g. there is one training set, which includes all adsorption interactions from the dataset). (A) The model including only an intercept, CE_{Local}, IPEA, and MADs. (B) The model where we have intentionally overfit, utilizing an intercept, the site's CN, the negative average of the IP and EA of the metal, the negative of the average of the IP and EA of the adsorbate, the HSAB-based hardness of the metal, the metal's d-electron configuration, the covalent radius of the metal, the resistivity of the metal, the melting point of the metal, the CE_{local}, and MADs. Error bars in each case are from the 10-fold cross-validated RMSE.

In an attempt to find meaningful descriptors for the three columns of the periodic table investigated in the slab dataset[157], we considered additional descriptors. Values for the EA and IP of the metal (used in the calculation of the chemical potential and chemical hardness under HSAB theory [153]), the covalent radius, the metal resistivity, and the metal melting point all come from the 92nd edition of the CRC Handbook of Chemistry and Physics [161]. Resistivity values used are those reported for 273K.

Table C.5: Regression statistics for intentionally-overfit model plot of Figure C.3 B, with coefficients generated via OLS regression. The final column indicates a coefficient estimate from LASSO at an optimal lambda value of $\lambda = 0.0075$ yielding the lowest MSE.

Term	Coefficient Estimate	Standard Error	p-Value	LASSO Coefficient Estimate
Intercept	-0.279	1.238	0.822	1.229
CN	0.073	0.038	0.054	0.037
Metal Chemical Potential	-0.009	0.064	0.887	-0.009
Metal Chemical Hardness	-0.219	0.14	0.119	0
Metal d-count	0.217	0.123	0.08	0
Metal Covalent Radius	-0.756	0.374	0.044	-0.467
Metal Resistivity	0.001	0.001	0.35	0
Metal Melting Point	0.001	0	0.041	0
CE_{local}	-0.004	0.05	0.931	-0.046
IPEA	0.132	0.015	$8.3*10^{-16}$	0.124
MADS	0.777	0.03	$<2*10^{16}$	0.738

Appendix C.4 LASSO

We observe in an OLS model attempting to fit all possible datapoints that p-values point towards the IPEA and MADs being the most-probable descriptors of the binding interaction, with the only other statistically-significant descriptor being the melting point of the metal. Further, CE_{local} is given a much higher p-value than the CN. Because p-values are only one (often over-relied-upon) piece of model verification [224], in order to get a better description of the importance of each of these variables, we utilized GLMNet [225] as implemented in R[155] to perform feature selection via LASSO [226]. Choosing a value of lambda minimizing the cross-validated mean-squared error, LASSO gives the following parameters nonzero coefficients: An intercept, CN, the chemical potential of the metal, the covalent radius of the metal, CE_{local}, IPEA, and MADS (Table C.5). Of these the highest coefficients are given to an intercept, MADS, IPEA, CN, and CE_{local}.

This offers support for the functional form of our model found in Equation C.4, as although CN seems to have a higher coefficient than CE_{local} by LASSO, CE_{local} captures additional information about the local chemical environment which CN is missing.

Appendix C.5 Symbolic Regression with Eureqa

Because the potential exists for cross terms and higher-order terms in the model, we continued our investigation of these parameters using the software package Eureqa [227], which performs symbolic regression via a GA. Taking the slab dataset [157], we first standardized the data by subtracting the mean and dividing by the standard deviation, on a per-column basis. We allowed the search of the formula space to include the following terms: constants, integer constants, any of the input variables, negations of any constants or variables, and allowed the operators addition, subtraction, multiplication, and division. Since RMSE is not available as an error metric in Eureqa, we instead selected squared error as our error metric. The formula search was allowed to run for approximately 1 million generations, investigating 78 billion potential equations via GA.

Upon investigating the array of potential models Eureqa generated, we observe that nearly all of them have some nonlinear dependence on the IPEA descriptor. The majority of these took the form of $A * IPEA^2$, $\frac{A}{B+IPEA}$, or $\frac{A+B*X}{B+IPEA}$, where A and B are constants, and X is some arbitrary other descriptor from the dataset. Although at first it may appear that this indicates a nonlinear dependence on IPEA, in reality this is an artifact of the dataset only containing three adsorbates. Because there are only three points in this dimension, either a parabola or reciprocal function are flexible-enough to capture them. In other words, this particular set of nonlinear terms featuring IPEA are likely overfitting to the data. For this purpose we utilize Occam's Razor: we focus only

on simple solutions which do not feature a nonlinear relation with IPEA, and ignore the other equations. We are then left with a series of equations all taking the form of Equation C.4.

$$BE = A + B * CE_{local} + C * IPEA + D * MADS \qquad \text{Equation C.4}$$

In Equation C.4, *A-D* are constants. In some simpler equations, some of these coefficients are set to zero, but the best fit occurs when all coefficients are assigned values. In other words, the GA search for equation forms, which was not restricted in its choice of descriptors, has undergone convergent evolution to the equation form we utilize in the main document.

Appendix C.6 Adsorption Configurations

To investigate a variety of chemical environments, we focus on five unique NP morphologies: a 55-atom icosahedron (Figure C.4 A), a 55-atom cuboctahedron (Figure C.4 B), a 147-atom icosahedron (Figure C.4 C), a 147-atom cuboctahedron (Figure C.4 D), and a 172-atom cube (Figure C.4 E).
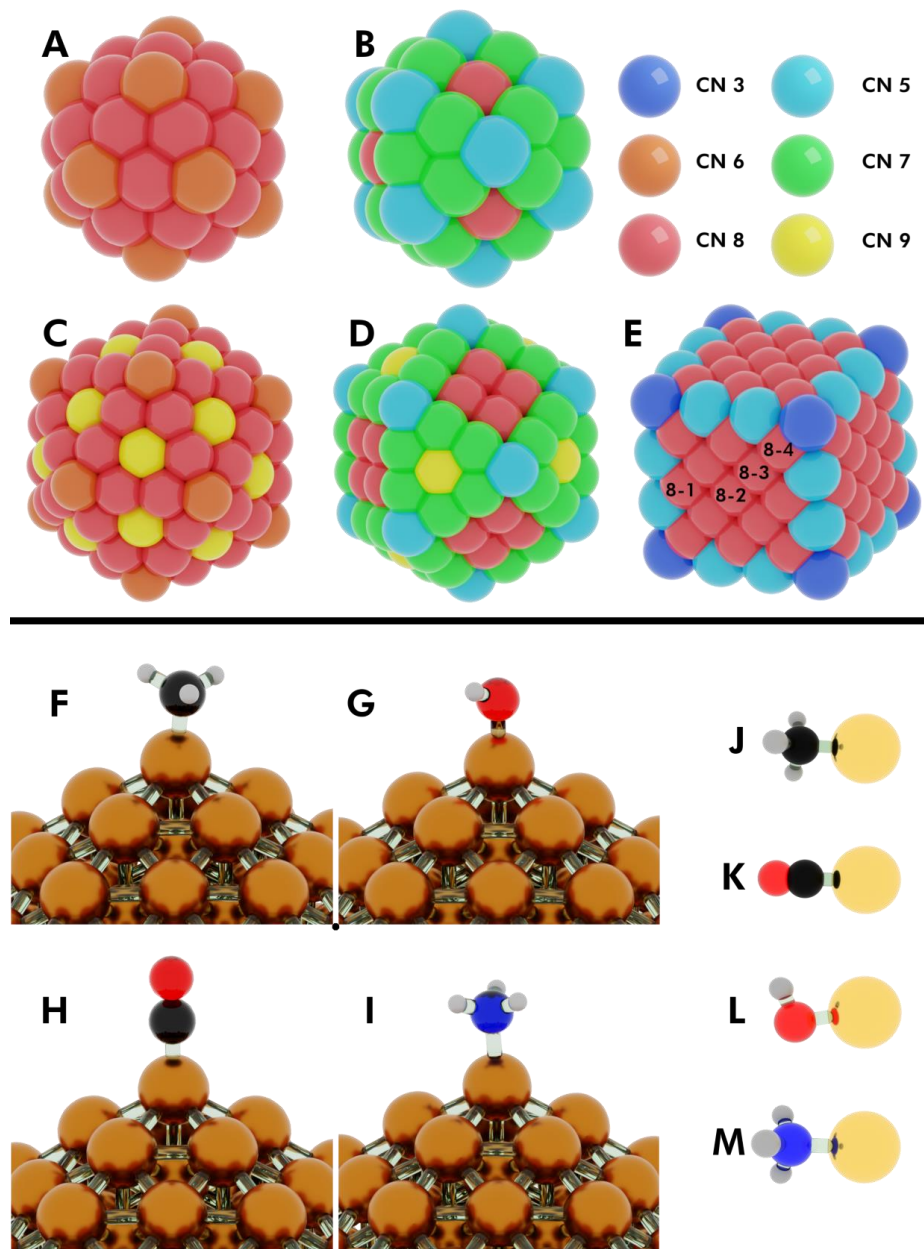
Figure C.4: Illustration of initial configurations for several DFT calculations performed. Upper: CNs on (A) 55-atom icosahedron, (B) 55-atom cuboctahedron, (C) 147-atom icosahedron, (D) 147-atom cuboctahedron, (E) 172-atom cube. In the case of NPs where more than one unique atoms share the same CN, we denote them with numbers 8-1, 8-2, 8-3, 8-4. Lower: Adsorbate-metal complexes in the case of (F) methyl radical, (G) hydroxyl radical, (H) carbon monoxide, (I) ammonia, (J) methyl radical bound to a gas-phase metal atom, (K) carbon monoxide bound to a gas-phase metal atom, (L) hydroxyl radical bound to a gas-phase metal atom, (M) ammonia bound to a gas-phase metal atom.

Due to symmetry, the 55-atom icosahedron (Figure C.4 A) has only two distinct sites, with CN equal to 6 or 8. The 55-atom cuboctahedron (Figure C.4 B) has three unique sites: CN 5, 7, or 8. The 147-atom icosahedron (Figure C.4 C) contains only three distinct sites, with CN equal to 6, 8, or 9. In the 147-atom cuboctahedron (Figure C.4 D), symmetry yields four distinct sites: CN 5, 7, 8, or 9. For the 172-atom cube (Figure C.4 E), six sites are possible: CN 3 or 5, as well as four unique CN 8 sites, which we refer to as 8-1, 8-2, 8-3, and 8-4.

For each of the morphologies shown in Figure C.4 A-E, monometallic Cu, Ag, and Au NPs were constructed and subsequently relaxed via DFT. After relaxation, the adsorbates of interest (CH$_3$, CO, OH) were placed at a "top" adsorption configuration (see Figure C.4 F-I) above each distinct surface site on the NPs. The top adsorption configuration was used as it gives the least number of distinct sites to screen per NP, and has been shown to trend with the BE of other surrounding adsorption sites [156]. In rare cases where the adsorbate moved away from a top configuration to a bridged or hollow position during optimization, the configuration was forced to relax on top. For all adsorption calculations, the NP structure was frozen.

The BE of an adsorbate was calculated as the difference between the adsorbed state and each species at infinite separation (gas-phase) was taken, as shown in Equation C.5.

$$E_{bind,Ads-NP} = E_{Ads-NP} - (E_{Ads} + E_{NP}) \qquad \text{Equation C.5}$$

In Equation C.5, $E_{bind, Ads-NP}$ refers to the BE, E_{Ads-NP} refers to the electronic energy of the NP-adsorbate complex, E_{Ads} refers to the gas-phase electronic energy of the adsorbate, and E_{NP} refers to the gas-phase electronic energy of the NP.

Adsorption calculations on a metal atom for all four adsorbates can be found in Figure C.4 J-M.

Table C.6: DFT-calculated binding energies for all studied adsorbate-NP pairs, ordered first by adsorbate, then by morphology, then by element, and finally by CN. The investigated adsorbates are the methyl radical, carbon dioxide molecule, hydroxyl radical, and ammonia molecule. Investigated morphologies are 172-atom cubes, 55-/147-atom cuboctahedrons, and 55-/147-atom icosahedrons. Investigated metals for these morphologies are Ag, Au, and Cu. In the case of Rh, the 55-atom cuboctahedron and 55-atom icosahedron were investigated. In the case of the bimetallic CuAg alloys, 55-atom icosahedrons were investigated. For NPs where multiple binding sites share the same CN, an underscore followed by a number is used to arbitrarily assign a unique ID to differentiate these sites. We additionally report the gas-phase single-metal-atom BE to each adsorbate; these calculations are indicated with the word "Gas" in the Morphology column.

Morphology	Size	Composition	CN	Adsorbate	DFT BE (eV)
Cube	172	Ag	3	CH3	-1.39
Cube	172	Ag	5	CH3	-1.11
Cube	172	Ag	8_1	CH3	-0.95
Cube	172	Ag	8_2	CH3	-0.74
Cube	172	Ag	8_3	CH3	-0.8
Cube	172	Ag	8_4	CH3	-0.89
Cube	172	Au	3	CH3	-1.68
Cube	172	Au	5	CH3	-1.47
Cube	172	Au	8_1	CH3	-1.18
Cube	172	Au	8_2	CH3	-0.99
Cube	172	Au	8_3	CH3	-1.1
Cube	172	Au	8_4	CH3	-1.19
Cube	172	Cu	3	CH3	-1.59
Cube	172	Cu	5	CH3	-1.43
Cube	172	Cu	8_1	CH3	-1.02
Cube	172	Cu	8_2	CH3	-0.89
Cube	172	Cu	8_3	CH3	-0.91
Cube	172	Cu	8_4	CH3	-1.14
Cuboctahedron	55	Ag	5	CH3	-0.97
Cuboctahedron	55	Ag	7	CH3	-1.09
Cuboctahedron	55	Ag	8	CH3	-0.74
Cuboctahedron	147	Ag	5	CH3	-1.25
Cuboctahedron	147	Ag	7	CH3	-1.18
Cuboctahedron	147	Ag	8	CH3	-0.9
Cuboctahedron	147	Ag	9	CH3	-1.09
Cuboctahedron	55	Au	5	CH3	-1.31

Table C.6 (continued).

Cuboctahedron	55	Au	7	CH3	-1.48
Cuboctahedron	55	Au	8	CH3	-1.12
Cuboctahedron	147	Au	5	CH3	-1.6
Cuboctahedron	147	Au	7	CH3	-1.42
Cuboctahedron	147	Au	8	CH3	-1.16
Cuboctahedron	147	Au	9	CH3	-1.28
Cuboctahedron	55	Cu	5	CH3	-1.3
Cuboctahedron	55	Cu	7	CH3	-1.43
Cuboctahedron	55	Cu	8	CH3	-1.05
Cuboctahedron	147	Cu	5	CH3	-1.6
Cuboctahedron	147	Cu	7	CH3	-1.51
Cuboctahedron	147	Cu	8	CH3	-1.02
Cuboctahedron	147	Cu	9	CH3	-1.36
Cuboctahedron	55	Rh	5	CH3	-2.02
Cuboctahedron	55	Rh	7	CH3	-1.99
Cuboctahedron	55	Rh	8	CH3	-1.9
Icosahedron	55	Ag	6	CH3	-0.8
Icosahedron	55	Ag	8	CH3	-1.03
Icosahedron	147	Ag	6	CH3	-1.23
Icosahedron	147	Ag	8	CH3	-1.01
Icosahedron	147	Ag	9	CH3	-0.8
Icosahedron	55	Au	6	CH3	-1.36
Icosahedron	55	Au	8	CH3	-1.29
Icosahedron	147	Au	6	CH3	-1.66
Icosahedron	147	Au	8	CH3	-1.29
Icosahedron	147	Au	9	CH3	-1.06
Icosahedron	55	Cu	6	CH3	-1.17
Icosahedron	55	Cu	8	CH3	-1.33
Icosahedron	147	Cu	6	CH3	-1.55
Icosahedron	147	Cu	8	CH3	-1.24
Icosahedron	147	Cu	9	CH3	-0.98
Icosahedron	55	$Cu_{22}Ag_{33}$	6_1	CH3	-1.22
Icosahedron	55	$Cu_{22}Ag_{33}$	6_2	CH3	-0.84
Icosahedron	55	$Cu_{22}Ag_{33}$	8_1	CH3	-1.25
Icosahedron	55	$Cu_{22}Ag_{33}$	8_2	CH3	-0.86
Icosahedron	55	$Cu_{31}Ag_{24}$	6_1	CH3	-1.08
Icosahedron	55	$Cu_{31}Ag_{24}$	6_2	CH3	-1.14
Icosahedron	55	$Cu_{31}Ag_{24}$	8_1	CH3	-1.21

Table C.6 (continued).

Icosahedron	55	$Cu_{31}Ag_{24}$	8_2	CH3	-0.9
Icosahedron	55	Rh	6	CH3	-2.08
Icosahedron	55	Rh	8	CH3	-2.09
Cube	172	Ag	3	CO	-0.77
Cube	172	Ag	5	CO	-0.68
Cube	172	Ag	8_1	CO	-0.57
Cube	172	Ag	8_2	CO	-0.4
Cube	172	Ag	8_3	CO	-0.52
Cube	172	Ag	8_4	CO	-0.61
Cube	172	Au	3	CO	-1.07
Cube	172	Au	5	CO	-0.96
Cube	172	Au	8_1	CO	-0.57
Cube	172	Au	8_2	CO	-0.45
Cube	172	Au	8_3	CO	-0.5
Cube	172	Au	8_4	CO	-0.56
Cube	172	Cu	3	CO	-0.81
Cube	172	Cu	5	CO	-1
Cube	172	Cu	8_1	CO	-1.05
Cube	172	Cu	8_2	CO	-0.77
Cube	172	Cu	8_3	CO	-0.85
Cube	172	Cu	8_4	CO	-1
Cuboctahedron	55	Ag	5	CO	-0.73
Cuboctahedron	55	Ag	7	CO	-0.6
Cuboctahedron	55	Ag	8	CO	-0.68
Cuboctahedron	147	Ag	5	CO	-0.71
Cuboctahedron	147	Ag	7	CO	-0.67
Cuboctahedron	147	Ag	8	CO	-0.51
Cuboctahedron	147	Ag	9	CO	-0.57
Cuboctahedron	55	Au	5	CO	-0.97
Cuboctahedron	55	Au	7	CO	-0.73
Cuboctahedron	55	Au	8	CO	-0.79
Cuboctahedron	147	Au	5	CO	-0.97
Cuboctahedron	147	Au	7	CO	-0.74
Cuboctahedron	147	Au	8	CO	-0.51
Cuboctahedron	147	Au	9	CO	-0.5
Cuboctahedron	55	Cu	5	CO	-1.3
Cuboctahedron	55	Cu	7	CO	-1.14
Cuboctahedron	55	Cu	8	CO	-1.2

Table C.6 (continued).

Cuboctahedron	147	Cu	5	CO	-1.3
Cuboctahedron	147	Cu	7	CO	-1.17
Cuboctahedron	147	Cu	8	CO	-0.94
Cuboctahedron	147	Cu	9	CO	-1.04
Cuboctahedron	55	Rh	5	CO	-2.22
Cuboctahedron	55	Rh	7	CO	-2.36
Cuboctahedron	55	Rh	8	CO	-2.25
Icosahedron	55	Ag	6	CO	-0.7
Icosahedron	55	Ag	8	CO	-0.68
Icosahedron	147	Ag	6	CO	-0.57
Icosahedron	147	Ag	8	CO	-0.54
Icosahedron	147	Ag	9	CO	-0.48
Icosahedron	55	Au	6	CO	-0.97
Icosahedron	55	Au	8	CO	-0.62
Icosahedron	147	Au	6	CO	-0.92
Icosahedron	147	Au	8	CO	-0.65
Icosahedron	147	Au	9	CO	-0.47
Icosahedron	55	Cu	6	CO	-1.23
Icosahedron	55	Cu	8	CO	-1.16
Icosahedron	147	Cu	6	CO	-1.04
Icosahedron	147	Cu	8	CO	-1
Icosahedron	147	Cu	9	CO	-0.88
Icosahedron	55	$Cu_{22}Ag_{33}$	6_1	CO	-1.24
Icosahedron	55	$Cu_{22}Ag_{33}$	6_2	CO	-0.68
Icosahedron	55	$Cu_{22}Ag_{33}$	8_1	CO	-1.2
Icosahedron	55	$Cu_{22}Ag_{33}$	8_2	CO	-0.61
Icosahedron	55	$Cu_{31}Ag_{24}$	6_1	CO	-1.14
Icosahedron	55	$Cu_{31}Ag_{24}$	6_2	CO	-1.12
Icosahedron	55	$Cu_{31}Ag_{24}$	8_1	CO	-1.13
Icosahedron	55	$Cu_{31}Ag_{24}$	8_2	CO	-0.55
Icosahedron	55	Rh	6	CO	-2.2
Icosahedron	55	Rh	8	CO	-2.65
Cube	172	Ag	3	OH	-2.44
Cube	172	Ag	5	OH	-2.24
Cube	172	Ag	8_1	OH	-2.1
Cube	172	Ag	8_2	OH	-1.9
Cube	172	Ag	8_3	OH	-2.02
Cube	172	Ag	8_4	OH	-2.04

Table C.6 (continued).

Cube	172	Au	3	OH	-2.21
Cube	172	Au	5	OH	-1.99
Cube	172	Au	8_1	OH	-1.64
Cube	172	Au	8_2	OH	-1.52
Cube	172	Au	8_3	OH	-1.57
Cube	172	Au	8_4	OH	-1.63
Cube	172	Cu	3	OH	-2.88
Cube	172	Cu	5	OH	-2.8
Cube	172	Cu	8_1	OH	-2.49
Cube	172	Cu	8_2	OH	-2.29
Cube	172	Cu	8_3	OH	-2.21
Cube	172	Cu	8_4	OH	-2.32
Cuboctahedron	55	Ag	5	OH	-2.12
Cuboctahedron	55	Ag	7	OH	-2.13
Cuboctahedron	55	Ag	8	OH	-2.03
Cuboctahedron	147	Ag	5	OH	-2.47
Cuboctahedron	147	Ag	7	OH	-2.47
Cuboctahedron	147	Ag	8	OH	-2.14
Cuboctahedron	147	Ag	9	OH	-2.33
Cuboctahedron	55	Au	5	OH	-1.76
Cuboctahedron	55	Au	7	OH	-1.78
Cuboctahedron	55	Au	8	OH	-1.66
Cuboctahedron	147	Au	5	OH	-2.25
Cuboctahedron	147	Au	7	OH	-2.02
Cuboctahedron	147	Au	8	OH	-1.76
Cuboctahedron	147	Au	9	OH	-1.85
Cuboctahedron	55	Cu	5	OH	-2.76
Cuboctahedron	55	Cu	7	OH	-2.67
Cuboctahedron	55	Cu	8	OH	-2.47
Cuboctahedron	147	Cu	5	OH	-3.06
Cuboctahedron	147	Cu	7	OH	-2.87
Cuboctahedron	147	Cu	8	OH	-2.42
Cuboctahedron	147	Cu	9	OH	-2.7
Cuboctahedron	55	Rh	5	OH	-3.33
Cuboctahedron	55	Rh	7	OH	-3
Cuboctahedron	55	Rh	8	OH	-2.77
Icosahedron	55	Ag	6	OH	-2.07
Icosahedron	55	Ag	8	OH	-2.11

Table C.6 (continued).

Icosahedron	147	Ag	6	OH	-2.38
Icosahedron	147	Ag	8	OH	-2.25
Icosahedron	147	Ag	9	OH	-2.09
Icosahedron	55	Au	6	OH	-1.68
Icosahedron	55	Au	8	OH	-1.52
Icosahedron	147	Au	6	OH	-2.25
Icosahedron	147	Au	8	OH	-1.88
Icosahedron	147	Au	9	OH	-1.68
Icosahedron	55	Cu	6	OH	-2.69
Icosahedron	55	Cu	8	OH	-2.62
Icosahedron	147	Cu	6	OH	-2.95
Icosahedron	147	Cu	8	OH	-2.72
Icosahedron	147	Cu	9	OH	-2.5
Icosahedron	55	$Cu_{22}Ag_{33}$	6_1	OH	-2.64
Icosahedron	55	$Cu_{22}Ag_{33}$	6_2	OH	-2.09
Icosahedron	55	$Cu_{22}Ag_{33}$	8_1	OH	-2.56
Icosahedron	55	$Cu_{22}Ag_{33}$	8_2	OH	-1.99
Icosahedron	55	$Cu_{31}Ag_{24}$	6_1	OH	-2.6
Icosahedron	55	$Cu_{31}Ag_{24}$	6_2	OH	-2.68
Icosahedron	55	$Cu_{31}Ag_{24}$	8_1	OH	-2.53
Icosahedron	55	$Cu_{31}Ag_{24}$	8_2	OH	-2
Icosahedron	55	Rh	6	OH	-3.35
Icosahedron	55	Rh	8	OH	-3.15
Cuboctahedron	55	Ag	5	NH3	-0.53
Cuboctahedron	55	Ag	7	NH3	-0.39
Cuboctahedron	55	Ag	8	NH3	-0.39
Icosahedron	55	Ag	6	NH3	-0.51
Icosahedron	55	Ag	8	NH3	-0.41
Cuboctahedron	147	Ag	5	NH3	-0.52
Cuboctahedron	147	Ag	7	NH3	-0.43
Cuboctahedron	147	Ag	8	NH3	-0.31
Cuboctahedron	147	Ag	9	NH3	-0.32
Icosahedron	147	Ag	6	NH3	-0.48
Icosahedron	147	Ag	8	NH3	-0.38
Icosahedron	147	Ag	9	NH3	-0.3
Cube	172	Ag	3	NH3	-0.61
Cube	172	Ag	5	NH3	-0.52
Cube	172	Ag	8_1	NH3	-0.36

Table C.6 (continued).

Cube	172	Ag	8_2	NH3	-0.28
Cube	172	Ag	8_3	NH3	-0.33
Cube	172	Ag	8_4	NH3	-0.34
Cuboctahedron	147	Au	5	NH3	-0.64
Cuboctahedron	147	Au	7	NH3	-0.51
Cuboctahedron	147	Au	8	NH3	-0.39
Cuboctahedron	147	Au	9	NH3	-0.4
Cuboctahedron	55	Au	5	NH3	-0.65
Cuboctahedron	55	Au	7	NH3	-0.48
Cuboctahedron	55	Au	8	NH3	-0.56
Icosahedron	55	Au	6	NH3	-0.64
Icosahedron	55	Au	8	NH3	-0.42
Icosahedron	147	Au	6	NH3	-0.63
Icosahedron	147	Au	8	NH3	-0.48
Icosahedron	147	Au	9	NH3	-0.37
Cube	172	Au	3	NH3	-0.71
Cube	172	Au	5	NH3	-0.65
Cube	172	Au	8_1	NH3	-0.41
Cube	172	Au	8_2	NH3	-0.38
Cube	172	Au	8_3	NH3	-0.37
Cube	172	Au	8_4	NH3	-0.38
Cuboctahedron	55	Cu	5	NH3	-0.76
Cuboctahedron	55	Cu	7	NH3	-0.54
Cuboctahedron	55	Cu	8	NH3	-0.48
Icosahedron	55	Cu	6	NH3	-0.73
Icosahedron	55	Cu	8	NH3	-0.56
Cuboctahedron	147	Cu	5	NH3	-0.74
Cuboctahedron	147	Cu	7	NH3	-0.59
Cuboctahedron	147	Cu	8	NH3	-0.29
Cuboctahedron	147	Cu	9	NH3	-0.38
Icosahedron	147	Cu	6	NH3	-0.69
Icosahedron	147	Cu	8	NH3	-0.51
Icosahedron	147	Cu	9	NH3	-0.34
Cube	172	Cu	3	NH3	-0.76
Cube	172	Cu	5	NH3	-0.76
Cube	172	Cu	8_1	NH3	-0.28
Cube	172	Cu	8_2	NH3	-0.22
Cube	172	Cu	8_3	NH3	-0.27

Table C.6 (continued).

Cube	172	Cu	8_4	NH3	-0.39
Icosahedron	55	Rh	6	NH3	-1.15
Icosahedron	55	Rh	8	NH3	-1.09
Cuboctahedron	55	Rh	5	NH3	-1.11
Cuboctahedron	55	Rh	7	NH3	-0.88
Cuboctahedron	55	Rh	8	NH3	-0.82
Icosahedron	55	$Cu_{22}Ag_{33}$	6_1	NH3	-0.73
Icosahedron	55	$Cu_{22}Ag_{33}$	6_2	NH3	-0.51
Icosahedron	55	$Cu_{22}Ag_{33}$	8_1	NH3	-0.52
Icosahedron	55	$Cu_{22}Ag_{33}$	8_2	NH3	-0.34
Icosahedron	55	$Cu_{31}Ag_{24}$	6_1	NH3	-0.63
Icosahedron	55	$Cu_{31}Ag_{24}$	6_2	NH3	-0.66
Icosahedron	55	$Cu_{31}Ag_{24}$	8_1	NH3	-0.44
Icosahedron	55	$Cu_{31}Ag_{24}$	8_2	NH3	-0.32
Gas	1	Ag	N/A	CH3	-2.1
Gas	1	Au	N/A	CH3	-2.8
Gas	1	Cu	N/A	CH3	-2.97
Gas	1	Ir	N/A	CH3	-3.65
Gas	1	Ni	N/A	CH3	-3.09
Gas	1	Pd	N/A	CH3	-2.43
Gas	1	Pt	N/A	CH3	-3.23
Gas	1	Rh	N/A	CH3	-3.01
Gas	1	Ag	N/A	CO	-0.4
Gas	1	Au	N/A	CO	-0.62
Gas	1	Cu	N/A	CO	-1.2
Gas	1	Ir	N/A	CO	-4.04
Gas	1	Ni	N/A	CO	-2.67
Gas	1	Pd	N/A	CO	-2.58
Gas	1	Pt	N/A	CO	-3.65
Gas	1	Rh	N/A	CO	-3.05
Gas	1	Ag	N/A	OH	-2.57
Gas	1	Au	N/A	OH	-2.55
Gas	1	Cu	N/A	OH	-3.58
Gas	1	Ir	N/A	OH	-4.22
Gas	1	Ni	N/A	OH	-3.92
Gas	1	Pd	N/A	OH	-2.82
Gas	1	Pt	N/A	OH	-3.49
Gas	1	Rh	N/A	OH	-3.78

Table C.6 (continued).

Gas	1	Ag	N/A	NH3	-0.54
Gas	1	Au	N/A	NH3	-0.81
Gas	1	Cu	N/A	NH3	-1.07
Gas	1	Rh	N/A	NH3	-1.64

Appendix D Supporting Information for "Rapid Prediction of Bimetallic Mixing Behavior at the Nanoscale"

The content of this appendix is adapted from the supporting information of Dean, J.; Cowan, M. J.; Estes, J.; Ramadan, M.; Mpourmpakis, G. "Rapid Prediction of Bimetallic Mixing Behavior at the Nanoscale." ACS Nano 2020. 14 (7) 8171-8180.

Appendix D.1 Connecting Excess Energy with Enthalpy of Mixing

In their development of the BCM, Yan et al[90] showed that for a bimetallic NP, the EE of an arbitrary alloy is related to the CE of that alloy and its monometallic counterparts, scaled by the fraction (F_A and F_B) of that component in the NP (Equation D.1).

$$EE_{A_xB_y} = CE_{A_xB_y} - \left(F_A CE_{A_{x+y}} + F_B CE_{B_{x+y}}\right) \qquad \text{Equation D.1}$$

Recalling the definition of enthalpy (Equation D.2), U is related to H *via* the product of pressure (P) and volume (V). When pressure is 0, H and U are therefore equivalent.

$$H = U + PV \qquad \text{Equation D.2}$$

The enthalpy of mixing for a binary system is defined as the difference between the enthalpy of the mixed system (H_{ij}) and the product of the enthalpy of the pure systems (H_i and H_j) with their fractions (F_i and F_j) in the mixed system (Equation D.3).

$$\Delta H_{mix} = H_{ij} - \left(F_i H_i + F_j H_j\right) \qquad \text{Equation D.3}$$

At 0 pressure, H=U as a consequence of Equation D.2. This allows us to find Equation D.4.

$$\Delta H_{mix} = U_{AB} - (F_A U_B + F_A U_B) \qquad \text{Equation D.4}$$

Under the framework of the BCM, the energy holding a NP together is exclusively defined as the sum of the bonds in the NP. We can use this to approximate the internal energy of the system as the product of the number of atoms in the system (N) and the CE (Equation D.5). This comes from the definition of CE as the average energy (on a per-atom basis) holding the system together.

$$N * CE_{A_x B_y} \approx U_{A_x B_y} \qquad \text{Equation D.5}$$

We can then plug the approximation found in Equation D.5 into Equation D.1 to generate Equation D.6.

$$EE_{alloy} \approx \left(\frac{U_{A_x B_y}}{N} - \left(\frac{F_A U_{A_N}}{N} + \frac{F_B U_{B_N}}{N} \right) \right) = \frac{U_{A_x B_y} - \left(F_A U_{A_N} + F_B U_{B_N} \right)}{N} \qquad \text{Equation D.6}$$

Recalling the relationship between ΔH_{mix} and U derived in Equation D.4, we can further simplify Equation D.6, yielding a relationship between the EE and enthalpy of mixing (Equation D.7).

$$EE_{alloy} \approx \frac{\Delta H_{mix}}{N} \qquad \text{Equation D.7}$$

Overall, this allows us to relate the property calculated by the BCM (CE) with the important thermodynamic property ΔH_{mix}, which allows us to then approximate the free energy of mixing in the main document.

Appendix D.2 Statistical Thermodynamics

The ZPE was calculated *via* Equation D.8,[228] where *h* represents Planck's constant and ν represents the vibrational frequency. It is a summation over the N_{modes} vibrational modes.

$$ZPE = \sum_{i=1}^{N_{modes}} \frac{h\nu_i}{2} \qquad \text{Equation D.8}$$

The constant-volume heat capacity due to vibrational motion ($C_{V,vib}$) was calculated *via* Equation D.9, where T refers to temperature and k_b refers to Boltzmann's constant. Utilizing the energy of each vibrational mode,[171] we sum over all modes to find the vibrational heat capacity.

$$\int_0^T C_{V,vib} dT = \sum_{i=1}^{N_{modes}} \frac{h\nu_i}{e^{\frac{h\nu_i}{k_b T}} - 1} \qquad \text{Equation D.9}$$

The entropic contribution from entropy (S_V) was calculated *via* Equation D.10.[228]

$$S_V = \sum_{i=1}^{N_{modes}} \frac{h\nu_i}{k_b T \left(e^{\frac{h\nu_i}{k_b T}} - 1 \right)} - \ln \left[1 - e^{-\frac{h\nu_i}{k_b T}} \right] \qquad \text{Equation D.10}$$

Finally, Gibb's free energy (G) can be calculated *via* Equation D.11.

$$G = E_{DFT} + ZPE + C_V T - TS_V$$

$$\text{Equation D.11}$$

Table D.1: Thermodynamic data for several chemical orderings randomly sampled in various Ag$_X$Au$_{55-X}$ systems.

System	Sample	DFT (eV)	ZPE (eV)	C$_{V,vib}$T (eV), 298K	TS (eV), 298K	G (eV), 298K
Ag$_{50}$Au$_5$	1	-54880	0.914	3.249	7.769	-54883.6
Ag$_{50}$Au$_5$	2	-54879.5	0.905	3.256	7.808	-54883.1
Ag$_{50}$Au$_5$	3	-54880	0.915	3.248	7.765	-54883.6
Ag$_{25}$Au$_{30}$	1	-52310.7	0.836	3.316	8.226	-52314.7
Ag$_{25}$Au$_{30}$	2	-52310.8	0.828	3.323	8.286	-52314.9
Ag$_{25}$Au$_{30}$	3	-52311.1	0.838	3.315	8.232	-52315.1
Ag$_{20}$Au$_{35}$	1	-51796.8	0.827	3.324	8.295	-51801
Ag$_{20}$Au$_{35}$	2	-51796.3	0.819	3.331	8.344	-51800.5
Ag$_{20}$Au$_{35}$	3	-51796.8	0.83	3.322	8.286	-51800.9
Ag$_5$Au$_{50}$	1	-50250.9	0.76	3.381	8.676	-50255.4
Ag$_5$Au$_{50}$	2	-50251.5	0.769	3.373	8.618	-50255.9
Ag$_5$Au$_{50}$	3	-50251.5	0.774	3.37	8.727	-50256

Table D.2: Thermodynamic data for several chemical orderings randomly sampled in various Ag$_X$Au$_{13-X}$ systems.

System	Sample	DFT (eV)	ZPE (eV)	C$_{V,vib}$T (eV), 298K	TS (eV), 298K	G (eV), 298K
Ag$_{11}$Au$_2$	1	-12882.3	0.185	0.679	1.663	-12883.1
Ag$_{11}$Au$_2$	2	-12881.7	0.18	0.682	1.671	-12882.5
Ag$_6$Au$_7$	1	-12368.7	0.174	0.689	1.761	-12369.6
Ag$_6$Au$_7$	2	-12368.6	0.177	0.686	1.735	-12369.5
Ag$_6$Au$_7$	3	-12368.3	0.174	0.688	1.718	-12369.1
Ag$_6$Au$_7$	4	-12369.4	0.182	0.682	1.686	-12370.2
Ag$_2$Au$_{11}$	1	-11955.6	0.156	0.703	1.783	-11956.5
Ag$_2$Au$_{11}$	2	-11955.6	0.156	0.703	1.786	-11956.5
Ag$_2$Au$_{11}$	3	-11955.9	0.161	0.7	1.84	-11956.9

Appendix D.3 Comparing Nanoparticle Mixing Behavior to Fully Random Mixing

For bimetallic systems, we define fully random mixing as the equal likelihood of forming any bond type (i.e. no energetic preference between bond types). To determine the bond fractions (F_{A-A}, F_{B-B}, and F_{A-B}) of a bimetallic NP that exhibits fully random mixing, we can calculate the probability of consecutively selecting two atoms from the group in order to form a bond. First, we imagine a system of N_{Atoms} atoms consisting of A and B atom types, where the composition is defined by $x_A + x_B = 1$. The probability of creating an A-A bond (p_{A-A}) is the chance of randomly choosing an A atom followed by choosing a second A atom, as shown in Equation D.12. The same reasoning can be used to determine the fraction of B-B bonds (Equation D.13).

$$F_{A\text{-}A} = p_{A\text{-}A} = x_A \times \left(\frac{N_A - 1}{N_{Atoms} - 1}\right) = x_A \times \left(\frac{N_A}{N_{Atoms} - 1} - \frac{1}{N_{Atoms} - 1}\right) \qquad \text{Equation D.12}$$

$$F_{B\text{-}B} = p_{B\text{-}B} = x_B \times \left(\frac{N_B - 1}{N_{Atoms} - 1}\right) = x_B \times \left(\frac{N_B}{N_{Atoms} - 1} - \frac{1}{N_{Atoms} - 1}\right) \qquad \text{Equation D.13}$$

This same reasoning can be used to develop an expression of the total number of hetero-atomic bonds in the system (Equation D.14 – Equation D.16).

$$p_{A-B} = x_A \left(\frac{N_B}{N_{Atoms} - 1}\right) \qquad \text{Equation D.14}$$

$$p_{B-A} = x_B \left(\frac{N_A}{N_{Atoms} - 1}\right) \qquad \text{Equation D.15}$$

$$F_{A-B} = F_{A-B\|B-A} = p_{A-B} + p_{B-A} = x_A \left(\frac{N_B}{N_{Atoms} - 1} \right) + x_B \left(\frac{N_A}{N_{Atoms} - 1} \right) \quad \text{Equation D.16}$$

16

For a sufficiently large NP (i.e. $N_{Atoms} \gg 1$), we can simplify several terms in Equation D.12 – Equation D.16 using Equation D.17 and Equation D.18.

$$\lim_{N_{Atoms} \to \infty} \frac{N_A}{N_{Atoms} - 1} = \frac{N_A}{N_{Atoms}} = x_A \quad \text{Equation D.17}$$

$$\lim_{N_{Atoms} \to \infty} \frac{1}{N_{Atoms} - 1} = 0 \quad \text{Equation D.18}$$

Thus, for sufficiently large NPs, we can simplify Equation D.12, Equation D.13, and Equation D.16 to Equations 19-21.

$$F_{A-A} = x_A^2 \quad \text{Equation D.19}$$

$$F_{B-B} = x_B^2 \quad \text{Equation D.20}$$

$$F_{A-B} = x_A x_B + x_B x_A = 2 x_A x_B \quad \text{Equation D.21}$$

This formulation is equivalent to applying a mean-field approximation with equal pairwise energies in a regular solution binary mixture.

Appendix D.4 Boltzmann Statistics at Various Temperatures

In Figure 6 of the main document, we present morphology phase diagrams at 298K for AuCu, AgAu, and AgCu systems with icosahedral, cuboctahedral, or EPB structures. In this

section, we report the same type of diagram, at 77K (Figure D.1), 640K (Figure D.2), and 1073K (Figure D.3). Overall, we observe that as temperature increases, the structural trend tends toward an even mix of all three morphologies. At low temperatures, the thermodynamically-most-favorable structures tend to dominate the distribution.
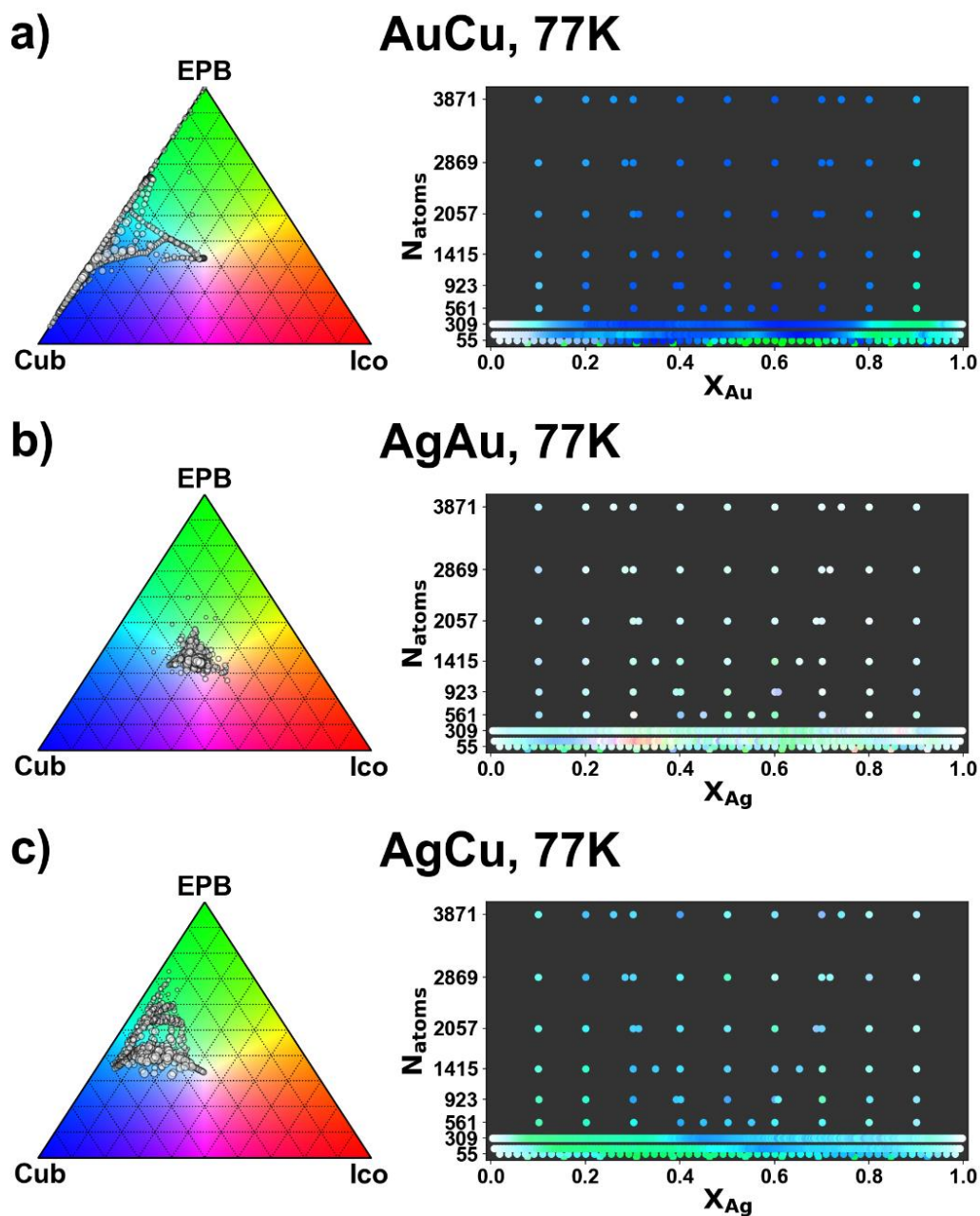
Figure D.1: Morphology phase diagrams for the a) AuCu, b) AgAu, and c) AgCu systems at 77K. The right-hand column plots the preferred morphological phase as a function of the number of atoms and composition of the system. The legend is given by the ternary diagrams on the left-hand side, which show the percentage of the Boltzmann population taking on cuboctahedral, icosahedral, or EPB. White points, for example, indicate all three morphologies are equally favorable. In addition, the size of the points on the ternary diagrams corresponds to NP size.
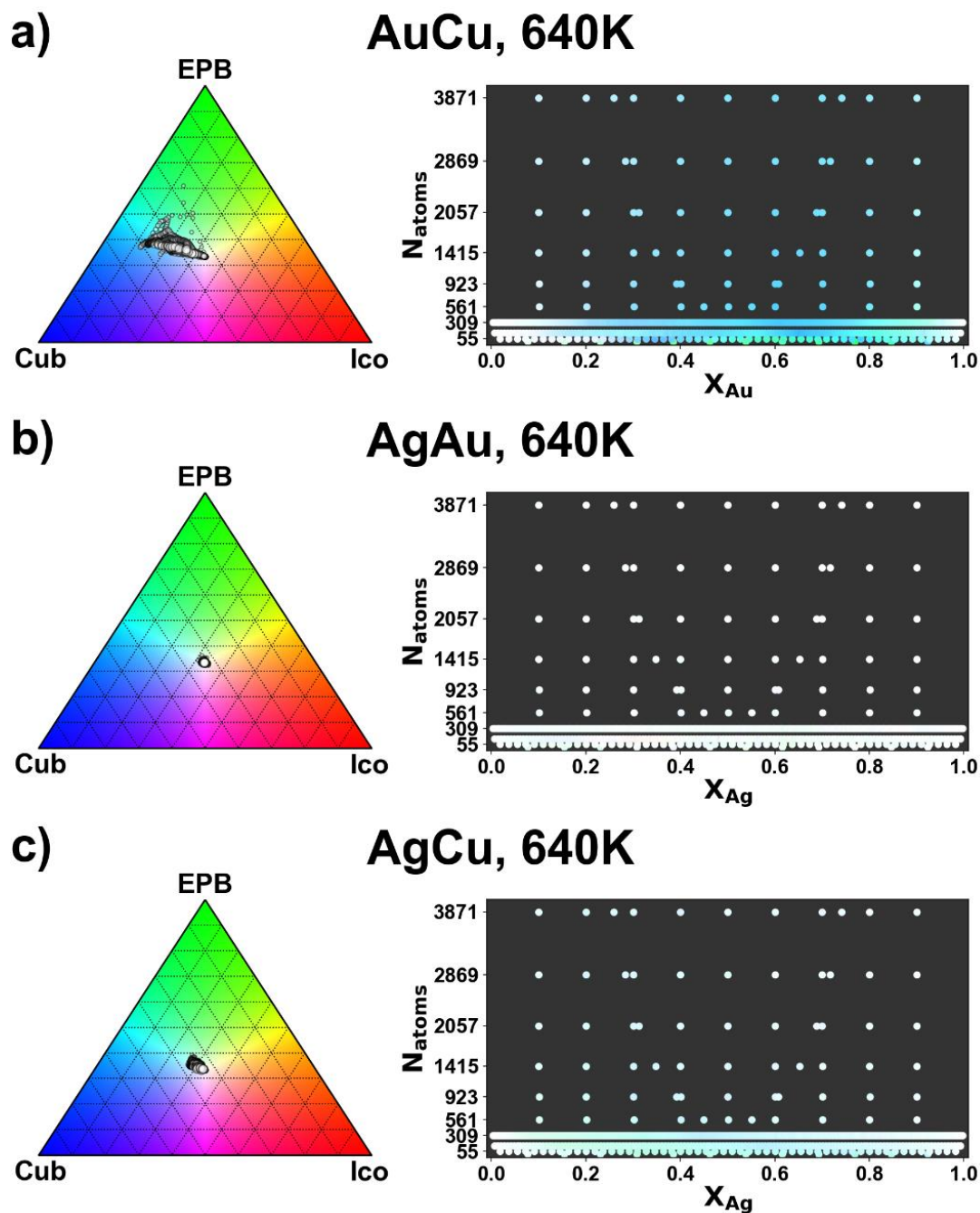
Figure D.2: Morphology phase diagrams for the a) AuCu, b) AgAu, and c) AgCu systems at 640K. The right-hand column plots the preferred morphological phase as a function of the number of atoms and composition of the system. The legend is given by the ternary diagrams on the left-hand side, which show the percentage of the Boltzmann population taking on cuboctahedral, icosahedral, or EPB. White points, for example, indicate all three morphologies are equally favorable. In addition, the size of the points on the ternary diagrams corresponds to NP size.
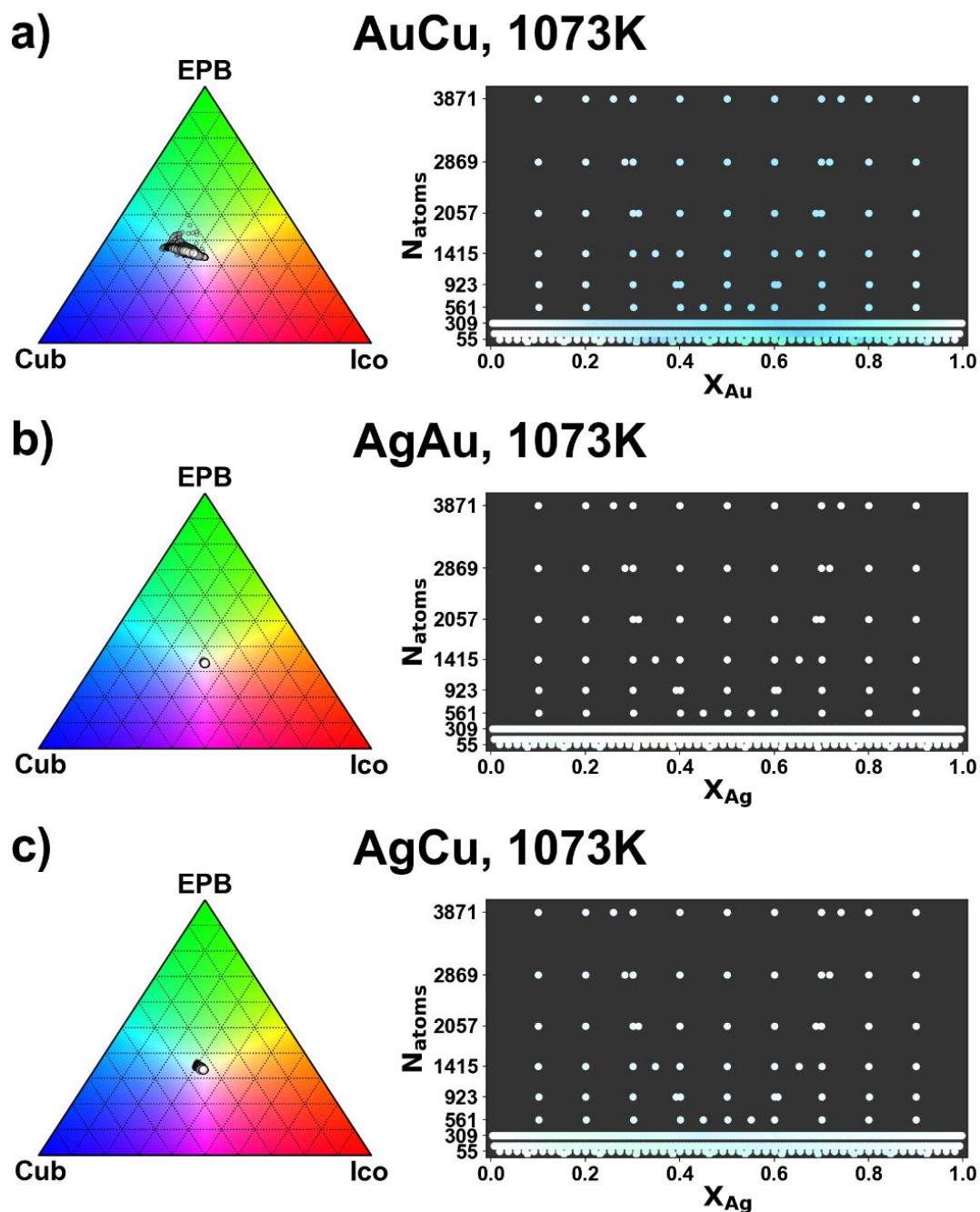
Figure D.3: Morphology phase diagrams for the a) AuCu, b) AgAu, and c) AgCu systems at 1073K. The right-hand column plots the preferred morphological phase as a function of the number of atoms and composition of the system. The legend is given by the ternary diagrams on the left-hand side, which show the percentage of the Boltzmann population taking on cuboctahedral, icosahedral, or EPB morphology. White points, for example, indicate all three morphologies are equally favorable. In addition, the size of the points on the ternary diagrams corresponds to NP size.

Appendix E Supporting Information for "High-Throughput Screening of Bimetallic

Nanoparticles for CO2 Adsorption"

The content of this appendix is adapted from the supporting information of the unpublished manuscript Dean, J.; Mpourmpakis, G. "High-Throughput Screening of Bimetallic Nanoparticles for CO_2 Adsorption." *In Preparation.*

Appendix E.1 Fitting Information
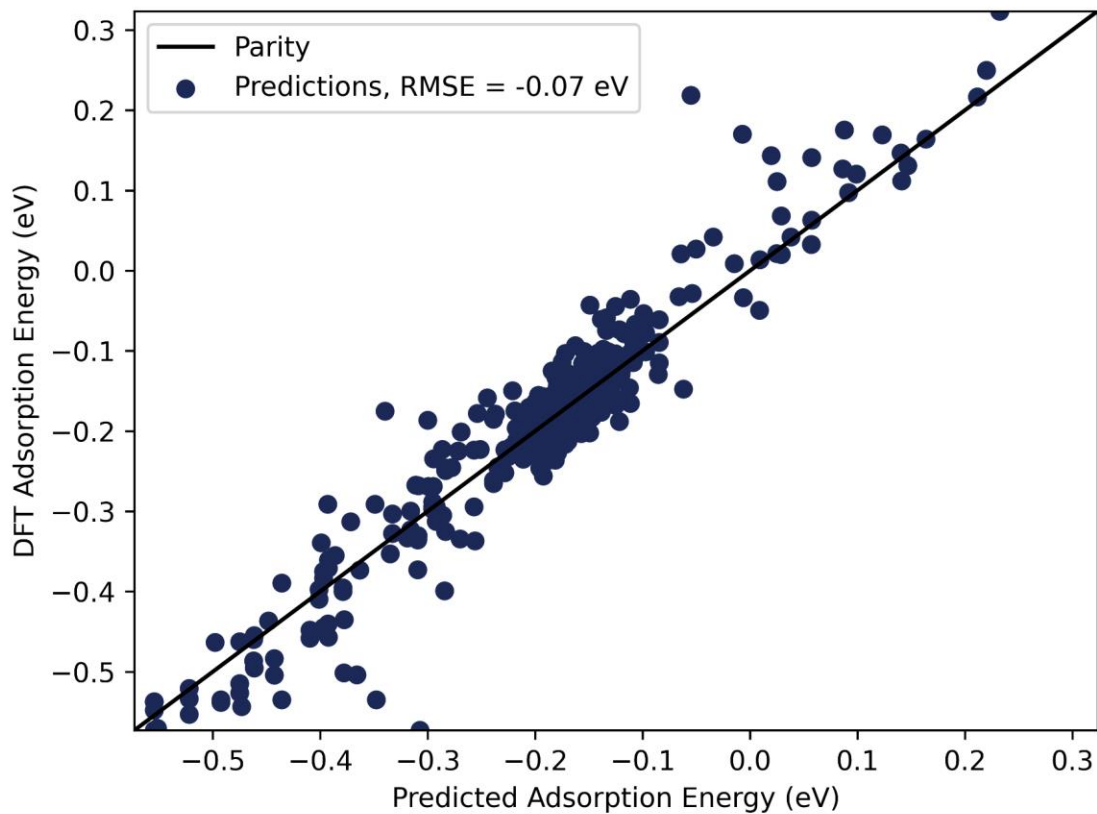
Figure E.1. Parity plot for a the random forest trained on all features, with a 10-fold CV RMSE of 0.07 eV. Parameters are: 32 estimators, 3 samples minimum for a split to occur, 1 sample minimum per leaf, no maximum depth or number of leaf nodes, and 9 features considered per split.

Table E.1. Variable importance scores for all features considered in the random forest.

Feature	Importance
MADs (summation)	0.29
CE_{local} (average)	0.19
ΔIP	0.14
Crystal Entropy (298K)	0.06
Resistivity (298K)	0.03
Conductivity (298K)	0.03
IP	0.02
Gas-Phase Entropy (298K)	0.02
Number of valence electrons	0.02
Gas-phase Gibbs Free Energy (298K)	0.02
Number of d electrons	0.00
IPEA	0.00

Bibliography

1. H., L. B., The Story of Adams' Catalyst: Platinum Oxide in Catalytic Reductions. *Platinum Metals Rev* **1962,** *6*, 41,150-152.

2. Wolfenden, J. H., The role of chance in chemical investigation. ACS Publications: 1967.

3. Nernst, W., Über das Ammoniakgleichgewicht. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie* **1907,** *13* (32), 521-524.

4. Mittasch, A.; Frankenburger, W., The historical development and theory of ammonia synthesis. *Journal of Chemical Education* **1929,** *6*, 2097.

5. Erisman, J. W.; Sutton, M. A.; Galloway, J.; Klimont, Z.; Winiwarter, W., How a century of ammonia synthesis changed the world. *Nature Geoscience* **2008,** *1* (10), 636-639.

6. Jacobsen, C. J. H.; Dahl, S.; Clausen, B. S.; Bahn, S.; Logadottir, A.; Nørskov, J. K., Catalyst Design by Interpolation in the Periodic Table: Bimetallic Ammonia Synthesis Catalysts. *Journal of the American Chemical Society* **2001,** *123* (34), 8404-8405.

7. Levy, M., Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc Natl Acad Sci U S A* **1979,** *76* (12), 6062-5.

8. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Physical Review* **1964,** *136* (3B), B864-B871.

9. Vignale, G.; Rasolt, M., Density-functional theory in strong magnetic fields. *Physical Review Letters* **1987,** *59* (20), 2360-2363.

10. Hartree, D. R.; Hartree, W., Self-consistent field, with exchange, for beryllium. *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences* **1935,** *150* (869), 9-33.

11. Schaller, R. R., Moore's law: past, present and future. *IEEE Spectrum* **1997,** *34* (6), 52-59.

12. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965,** *140* (4A), A1133-A1138.

13. Kohn, W., Density Functional Theory of Metal Surfaces and of Chemisorption on Metals. In *The Physical Basis for Heterogeneous Catalysis*, Drauglis, E.; Jaffee, R. I., Eds. Springer US: Boston, MA, 1975; pp 279-293.

14. Niu, S.; Hall, M. B., Theoretical Studies on Reactions of Transition-Metal Complexes. *Chemical Reviews* **2000,** *100* (2), 353-406.

15. Chermette, H., Density functional theory: A powerful tool for theoretical studies in coordination chemistry. *Coordination Chemistry Reviews* **1998,** *178-180*, 699-721.

16. Wimmer, E.; Fu, C. L.; Freeman, A. J., Catalytic Promotion and Poisoning: All-Electron Local-Density-Functional Theory of CO on Ni(001) Surfaces Coadsorbed with K or S. *Physical Review Letters* **1985,** *55* (23), 2618-2621.

17. Stampfl, C.; Scheffler, M., Anomalous Behavior of Ru for Catalytic Oxidation: A Theoretical Study of the Catalytic ReactionCO+12O2→CO2. *Physical Review Letters* **1997,** *78* (8), 1500-1503.

18. Ravikovitch, P. I.; Haller, G. L.; Neimark, A. V., Density functional theory model for calculating pore size distributions: pore structure of nanoporous catalysts. *Advances in Colloid and Interface Science* **1998,** *76-77*, 203-226.

19. Nicholas, J. B., Density functional theory studies of zeolite structure, acidity, and reactivity. *Topics in Catalysis* **1997,** *4* (1), 157-171.

20. Alavi, A.; Hu, P.; Deutsch, T.; Silvestrelli, P. L.; Hutter, J., CO Oxidation on Pt(111): An Ab Initio Density Functional Theory Study. *Physical Review Letters* **1998,** *80* (16), 3650-3653.

21. Langley, P., The changing science of machine learning. Springer: 2011.

22. Reyes, K. G.; Maruyama, B., The machine learning revolution in materials? *MRS Bulletin* **2019,** *44* (7), 530-537.

23. Mpourmpakis, G.; Andriotis, A. N.; Vlachos, D. G., Identification of Descriptors for the CO Interaction with Metal Nanoparticles. *Nano Lett* **2010,** *10* (3), 1041-1045.

24. Dean, J.; Taylor, M. G.; Mpourmpakis, G., Unfolding adsorption on metal nanoparticles: Connecting stability with catalysis. *Science Advances* **2019,** *5* (9), eaax5101.

25. Magela E Silva, G.; Acioli, P. H.; Pedroza, A. C., Estimating correlation energy of diatomic molecules and atoms with neural networks. *Journal of Computational Chemistry* **1997,** *18* (11), 1407-1414.

26. Dral, P. O., Quantum Chemistry in the Age of Machine Learning. *The Journal of Physical Chemistry Letters* **2020,** *11* (6), 2336-2347.

27. Behler, J.; Parrinello, M., Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters* **2007,** *98* (14), 146401.

28. Krizhevsky, A.; Sutskever, I.; Hinton, G. E., Imagenet classification with deep convolutional neural networks. *Communications of the ACM* **2017,** *60* (6), 84-90.

29. Smith, J. S.; Isayev, O.; Roitberg, A. E., ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science* **2017,** *8* (4), 3192-3203.

30. Custódio, C. A.; Filletti, É. R.; França, V. V., Artificial neural networks for density-functional optimizations in fermionic systems. *Scientific reports* **2019,** *9* (1), 1-7.

31. Zhang, Y.; Hu, C.; Jiang, B. Embedded Atom Neural Network Potentials: Efficient and Accurate Machine Learning with a Physically Inspired Representation 2019, p. arXiv:1907.06159. https://ui.adsabs.harvard.edu/abs/2019arXiv190706159Z (accessed July 01, 2019).

32. 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. *The Journal of Physical Chemistry Letters* **2019,** *10* (15), 4401-4408.

33. Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K. R.; Maurer, R. J., Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature Communications* **2019,** *10* (1), 5024.

34. Gastegger, M.; McSloy, A.; Luya, M.; Schütt, K. T.; Maurer, R. J., A deep neural network for molecular wave functions in quasi-atomic minimal basis representation. *The Journal of Chemical Physics* **2020,** *153* (4), 044123.

35. Hu, C.; Zhang, Y.; Jiang, B., Dynamics of H2O Adsorption on Pt(110)-(1 × 2) Based on a Neural Network Potential Energy Surface. *The Journal of Physical Chemistry C* **2020,** *124* (42), 23190-23199.

36. Gao, P.; Zhang, J.; Sun, Y.; Yu, J., Toward Accurate Predictions of Atomic Properties via Quantum Mechanics Descriptors Augmented Graph Convolutional Neural Network: Application of This Novel Approach in NMR Chemical Shifts Predictions. *The Journal of Physical Chemistry Letters* **2020,** *11* (22), 9812-9818.

37. Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C., Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013,** *65* (11), 1501-1509.

38. Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C., The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *Npj Computational Materials* **2015,** *1*, 15010.

39. Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A., Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013,** *1* (1), 011002.

40. Winther, K. T.; Hoffmann, M. J.; Boes, J. R.; Mamun, O.; Bajdich, M.; Bligaard, T., Catalysis-Hub.org, an open electronic structure database for surface reactions. *Scientific Data* **2019,** *6* (1), 75.

41. Johnson, R. D., III, NIST Computational Chemistry Comparison and Benchmark Database. 2018.

42. de Pablo, J. J.; Jackson, N. E.; Webb, M. A.; Chen, L.-Q.; Moore, J. E.; Morgan, D.; Jacobs, R.; Pollock, T.; Schlom, D. G.; Toberer, E. S., New frontiers for the materials genome initiative. *npj Computational Materials* **2019,** *5* (1), 41.

43. Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A., DENDRAL: a case study of the first expert system for scientific hypothesis formation. *Artif. Intell.* **1993,** *61* (2), 209–261.

44. Peris, M., An Overview of Recent Expert System Applications in Analytical Chemistry. *Critical Reviews in Analytical Chemistry* **1996,** *26* (4), 219-237.

45. Zolotarev, P. N.; Arshad, M. N.; Asiri, A. M.; Al-amshany, Z. M.; Blatov, V. A., A Possible Route toward Expert Systems in Supramolecular Chemistry: 2-Periodic H-Bond Patterns in Molecular Crystals. *Crystal Growth & Design* **2014,** *14* (4), 1938-1949.

46. de Santana, F. B.; Borges Neto, W.; Poppi, R. J., Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. *Food Chemistry* **2019,** *293*, 323-332.

47. Zhang, Q.; Sun, X.; Feng, K.; Wang, S.; Zhang, Y.-H.; Wang, S.; Lu, L.; Cai, Y.-D., Predicting Citrullination Sites in Protein Sequences Using mRMR Method and Random Forest Algorithm. *Combinatorial Chemistry & High Throughput Screening* **2017,** *20* (2), 164-173.

48. Keller, C. A.; Evans, M. J., Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. *Geoscientific Model Development Discussions* **2019**, 1209-1225.

49. Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. O., Random Forest Models To Predict Aqueous Solubility. *Journal of Chemical Information and Modeling* **2007,** *47* (1), 150-158.

50. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P., Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences* **2003,** *43* (6), 1947-1958.

51. Sheridan, R. P., Three Useful Dimensions for Domain Applicability in QSAR Models Using Random Forest. *Journal of Chemical Information and Modeling* **2012,** *52* (3), 814-823.

52. Lombardo, F.; Obach, R. S.; DiCapua, F. M.; Bakken, G. A.; Lu, J.; Potter, D. M.; Gao, F.; Miller, M. D.; Zhang, Y., A Hybrid Mixture Discriminant Analysis−Random Forest Computational Model for the Prediction of Volume of Distribution of Drugs in Human. *Journal of Medicinal Chemistry* **2006,** *49* (7), 2262-2267.

53. Zilian, D.; Sotriffer, C. A., SFCscoreRF: A Random Forest-Based Scoring Function for Improved Affinity Prediction of Protein–Ligand Complexes. *Journal of Chemical Information and Modeling* **2013,** *53* (8), 1923-1933.

54. Cano, G.; Garcia-Rodriguez, J.; Garcia-Garcia, A.; Perez-Sanchez, H.; Benediktsson, J. A.; Thapa, A.; Barr, A., Automatic selection of molecular descriptors using random forest: Application to drug discovery. *Expert Systems with Applications* **2017,** *72*, 151-159.

55. Polishchuk, P. G.; Muratov, E. N.; Artemenko, A. G.; Kolumbin, O. G.; Muratov, N. N.; Kuz'min, V. E., Application of Random Forest Approach to QSAR Prediction of Aquatic Toxicity. *Journal of Chemical Information and Modeling* **2009,** *49* (11), 2481-2488.

56. Goldschmidt, V. M., Die Gesetze der Krystallochemie. *Naturwissenschaften* **1926,** *14* (21), 477-485.

57. Hautier, G.; Fischer, C.; Ehrlacher, V.; Jain, A.; Ceder, G., Data Mined Ionic Substitutions for the Discovery of New Compounds. *Inorganic Chemistry* **2011,** *50* (2), 656-663.

58. Jain, A.; Hautier, G.; Moore, C. J.; Ping Ong, S.; Fischer, C. C.; Mueller, T.; Persson, K. A.; Ceder, G., A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science* **2011,** *50* (8), 2295-2310.

59. Oganov, A. R.; Glass, C. W., Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of Chemical Physics* **2006,** *124* (24), 244704.

60. Oganov, A. R.; Lyakhov, A. O.; Valle, M., How Evolutionary Crystal Structure Prediction Works—and Why. *Accounts Chem. Res.* **2011,** *44* (3), 227-237.

61. Lyakhov, A. O.; Oganov, A. R.; Stokes, H. T.; Zhu, Q., New developments in evolutionary structure prediction algorithm USPEX. *Computer Physics Communications* **2013,** *184* (4), 1172-1182.

62. Maddox, J., Crystals from first principles. *Nature* **1988,** *335* (6187), 201-201.

63. Zhang, W.; Oganov, A. R.; Goncharov, A. F.; Zhu, Q.; Boulfelfel, S. E.; Lyakhov, A. O.; Stavrou, E.; Somayazulu, M.; Prakapenka, V. B.; Konôpková, Z., Unexpected Stable Stoichiometries of Sodium Chlorides. *Science* **2013,** *342* (6165), 1502-1505.

64. Oganov, A. R.; Ono, S., The high-pressure phase of alumina and implications for Earth's D″ layer. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (31), 10828-10831.

65. Gao, G.; Oganov, A. R.; Ma, Y.; Wang, H.; Li, P.; Li, Y.; Iitaka, T.; Zou, G., Dissociation of methane under high pressure. *The Journal of Chemical Physics* **2010,** *133* (14), 144508.

66. Oganov, A. R.; Ono, S., Theoretical and experimental evidence for a post-perovskite phase of MgSiO3 in Earth's D″ layer. *Nature* **2004,** *430* (6998), 445-448.

67. Oganov, A. R.; Martoňák, R.; Laio, A.; Raiteri, P.; Parrinello, M., Anisotropy of Earth's D″ layer and stacking faults in the MgSiO3 post-perovskite phase. *Nature* **2005,** *438* (7071), 1142-1144.

68. Rapetti, D.; Ferrando, R., Density functional theory global optimization of chemical ordering in AgAu nanoalloys. *J. Alloy. Comp.* **2019,** *779*, 582-589.

69. Isenberg, N. M.; Taylor, M. G.; Yan, Z.; Hanselman, C. L.; Mpourmpakis, G.; Gounaris, C. E., Identification of optimally stable nanocluster geometries via mathematical optimization and density-functional theory. *Mol. Syst. Des. Eng.* **2019**.

70. Mamatkulov, M.; Yudanov, I. V.; Bukhtiyarov, A. V.; Prosvirin, I. P.; Bukhtiyarov, V. I.; Neyman, K. M., Pd Segregation on the Surface of Bimetallic PdAu Nanoparticles Induced by Low Coverage of Adsorbed CO. *J. Phys. Chem. C* **2019,** *123* (13), 8037-8046.

71. Farsi, L.; Deskins, N. A., First principles analysis of surface dependent segregation in bimetallic alloys. *Phys. Chem. Chem. Phys.* **2019,** *21* (42), 23626-23637.

72. Yu, Y.; Xiao, W.; Wang, J.; Wang, L., Understanding the surface segregation behavior of transition metals on Ni(111): a first-principles study. *Phys. Chem. Chem. Phys.* **2016,** *18* (38), 26616-26622.

73. Wang, L.-L.; Johnson, D. D., Predicted Trends of Core−Shell Preferences for 132 Late Transition-Metal Binary-Alloy Nanoparticles. *J. Am. Chem. Soc.* **2009,** *131* (39), 14023-14029.

74. Ruban, A. V.; Skriver, H. L.; Nørskov, J. K., Surface segregation energies in transition-metal alloys. *Phys. Rev. B* **1999,** *59* (24), 15990-16000.

75. Dean, J.; Cowan, M. J.; Estes, J.; Ramadan, M.; Mpourmpakis, G., Rapid Prediction of Bimetallic Mixing Behavior at the Nanoscale. *ACS Nano* **2020,** *14* (7), 8171-8180.

76. Leardi, R., Genetic algorithms in chemometrics and chemistry: a review. *Journal of Chemometrics* **2001,** *15* (7), 559-569.

77. Broadhurst, D.; Goodacre, R.; Jones, A.; Rowland, J. J.; Kell, D. B., Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Analytica Chimica Acta* **1997,** *348* (1), 71-86.

78. Jouan-Rimbaud, D.; Massart, D.-L.; Leardi, R.; De Noord, O. E., Genetic Algorithms as a Tool for Wavelength Selection in Multivariate Calibration. *Analytical Chemistry* **1995,** *67* (23), 4295-4301.

79. Bangalore, A. S.; Shaffer, R. E.; Small, G. W.; Arnold, M. A., Genetic Algorithm-Based Method for Selecting Wavelengths and Model Size for Use with Partial Least-Squares Regression: Application to Near-Infrared Spectroscopy. *Analytical Chemistry* **1996,** *68* (23), 4200-4212.

80. Deaven, D. M.; Ho, K. M., Molecular Geometry Optimization with a Genetic Algorithm. *Physical Review Letters* **1995,** *75* (2), 288-291.

81. Panteleev, J.; Gao, H.; Jia, L., Recent applications of machine learning in medicinal chemistry. *Bioorganic & Medicinal Chemistry Letters* **2018,** *28* (17), 2807-2815.

82. Häse, F.; Roch, L. M.; Kreisbeck, C.; Aspuru-Guzik, A., Phoenics: A Bayesian Optimizer for Chemistry. *ACS Central Science* **2018,** *4* (9), 1134-1145.

83. Tarafder, A.; Lee, B. C. S.; Ray, A. K.; Rangaiah, G. P., Multiobjective Optimization of an Industrial Ethylene Reactor Using a Nondominated Sorting Genetic Algorithm. *Industrial & Engineering Chemistry Research* **2005,** *44* (1), 124-141.

84. Rajesh, J. K.; Gupta, S. K.; Rangaiah, G. P.; Ray, A. K., Multiobjective Optimization of Steam Reformer Performance Using Genetic Algorithm. *Industrial & Engineering Chemistry Research* **2000,** *39* (3), 706-717.

85. Ebrahimi, S.; Ghorbani, B.; Vijayaraghavan, K., Optimization of catalyst distribution along PEMFC channel through a numerical two-phase model and genetic algorithm. *Renewable Energy* **2017,** *113*, 846-854.

86. Altınten, A.; Erdoğan, S.; Hapoğlu, H.; Alpbaz, M., Control of a polymerization reactor by fuzzy control method with genetic algorithm. *Computers & Chemical Engineering* **2003,** *27* (7), 1031-1040.

87. Altınten, A.; Ketevanlioğlu, F.; Erdoğan, S.; Hapoğlu, H.; Alpbaz, M., Self-tuning PID control of jacketed batch polystyrene reactor using genetic algorithm. *Chemical Engineering Journal* **2008,** *138* (1), 490-497.

88. Alade, I. O.; Abd Rahman, M. A.; Saleh, T. A., Modeling and prediction of the specific heat capacity of Al2 O3/water nanofluids using hybrid genetic algorithm/support vector regression model. *Nano-Structures & Nano-Objects* **2019,** *17*, 103-111.

89. Vakili, M.; Khosrojerdi, S.; Aghajannezhad, P.; Yahyaei, M., A hybrid artificial neural network-genetic algorithm modeling approach for viscosity estimation of graphene nanoplatelets nanofluid using experimental data. *International Communications in Heat and Mass Transfer* **2017,** *82*, 40-48.

90. Yan, Z.; Taylor, M. G.; Mascareno, A.; Mpourmpakis, G., Size-, Shape-, and Composition-Dependent Model for Metal Nanoparticle Stability Prediction. *Nano Letters* **2018,** *18* (4), 2696-2704.

91. Wang, Z.; Hu, P., Towards rational catalyst design: a general optimization framework. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2016,** *374* (2061), 20150078.

92. Arrhenius, S., XXXI. On the influence of carbonic acid in the air upon the temperature of the ground. *Philosophical Magazine* **1896,** *41* (251), 237-276.

93. Larmier, K.; Liao, W. C.; Tada, S.; Lam, E.; Verel, R.; Bansode, A.; Urakawa, A.; Comas-Vives, A.; Coperet, C., CO2-to-Methanol Hydrogenation on Zirconia-Supported Copper Nanoparticles: Reaction Intermediates and the Role of the Metal-Support Interface. *Angew Chem Int Edit* **2017,** *56* (9), 2318-2323.

94. Fisher, I. A.; Bell, A. T., In-situ infrared study of methanol synthesis from H-2/CO2 over Cu/SiO2 and Cu/ZrO2/SiO2. *J Catal* **1997,** *172* (1), 222-237.

95. Fiorani, G.; Guo, W. S.; Kleij, A. W., Sustainable conversion of carbon dioxide: the advent of organocatalysis. *Green Chem.* **2015,** *17* (3), 1375-1389.

96. Wang, W.; Wang, S. P.; Ma, X. B.; Gong, J. L., Recent advances in catalytic hydrogenation of carbon dioxide. *Chem Soc Rev* **2011,** *40* (7), 3703-3727.

97. Song, C. S., Global challenges and strategies for control, conversion and utilization of CO2 for sustainable development involving energy, catalysis, adsorption and chemical processing. *Catalysis Today* **2006,** *115* (1-4), 2-32.

98. Gotz, M.; Lefebvre, J.; Mors, F.; Koch, A. M.; Graf, F.; Bajohr, S.; Reimert, R.; Kolb, T., Renewable Power-to-Gas: A technological and economic review. *Renew. Energy* **2016,** *85*, 1371-1390.

99. Sabatier, P., Direct hydrogenation of oxides of carbon in presence of various finely divided metals. *C. R. Acad. Sci.* **1902,** *134*, 689-691.

100. Newsome, D. S., The Water-Gas Shift Reaction. *Catalysis Reviews* **1980,** *21* (2), 275-318.

101. Austin, N.; Ye, J.; Mpourmpakis, G., CO2 activation on Cu-based Zr-decorated nanoparticles. *Catalysis Science & Technology* **2017,** *7* (11), 2245-2251.

102. Porosoff, M. D.; Baldwin, J. W.; Peng, X.; Mpourmpakis, G.; Willauer, H. D., Potassium-Promoted Molybdenum Carbide as a Highly Active and Selective Catalyst for CO(2) Conversion to CO. *ChemSusChem* **2017,** *10* (11), 2408-2415.

103. Zhao, S.; Austin, N.; Li, M.; Song, Y.; House, S. D.; Bernhard, S.; Yang, J. C.; Mpourmpakis, G.; Jin, R., Influence of Atomic-Level Morphology on Catalysis: The Case of Sphere and Rod-Like Gold Nanoclusters for CO2 Electroreduction. *ACS Catalysis* **2018,** *8* (6), 4996-5001.

104. Austin, N.; Zhao, S.; McKone, J. R.; Jin, R.; Mpourmpakis, G., Elucidating the active sites for CO2 electroreduction on ligand-protected Au25 nanoclusters. *Catalysis Science & Technology* **2018,** *8* (15), 3795-3805.

105. Nagarajan, A. V.; Juarez-Mosqueda, R.; Cowan, M. J.; Jin, R.; Kauffman, D. R.; Mpourmpakis, G., Elucidating the stability of ligand-protected Au nanoclusters under electrochemical reduction of CO2. *SN Applied Sciences* **2020,** *2* (4), 680.

106. Li, S.; Alfonso, D.; Nagarajan, A. V.; House, S. D.; Yang, J. C.; Kauffman, D. R.; Mpourmpakis, G.; Jin, R., Monopalladium Substitution in Gold Nanoclusters Enhances CO2 Electroreduction Activity and Selectivity. *ACS Catalysis* **2020,** *10* (20), 12011-12016.

107. Dixit, M.; Peng, X.; Porosoff, M.; Willauer, H.; Mpourmpakis, G., Elucidating the Role of Oxygen Coverage in CO 2 Reduction on Mo 2 C. *Catal. Sci. Technol.* **2017,** *7*.

108. Austin, N.; Butina, B.; Mpourmpakis, G., CO2 activation on bimetallic CuNi nanoparticles. *Progress in Natural Science: Materials International* **2016,** *26* (5), 487-492.

109. Gao, D.; Arán-Ais, R. M.; Jeon, H. S.; Roldan Cuenya, B., Rational catalyst and electrolyte design for CO2 electroreduction towards multicarbon products. *Nature Catalysis* **2019,** *2* (3), 198-210.

110. Dang, S.; Qin, B.; Yang, Y.; Wang, H.; Cai, J.; Han, Y.; Li, S.; Gao, P.; Sun, Y., Rationally designed indium oxide catalysts for CO2 hydrogenation to methanol with high activity and selectivity. *Sci Adv* **2020,** *6* (25), eaaz2060.

111. Klinkova, A.; De Luna, P.; Dinh, C.-T.; Voznyy, O.; Larin, E. M.; Kumacheva, E.; Sargent, E. H., Rational Design of Efficient Palladium Catalysts for Electroreduction of Carbon Dioxide to Formate. *ACS Catalysis* **2016,** *6* (12), 8115-8120.

112. Dean, J.; Yang, Y.; Austin, N.; Veser, G.; Mpourmpakis, G., Design of Copper-Based Bimetallic Nanoparticles for Carbon Dioxide Adsorption and Activation. *ChemSusChem* **2018,** *11* (7), 1169-1178.

113. Walsh, A. D., 467. The electronic orbitals, shapes, and spectra of polyatomic molecules. Part II. Non-hydride AB2 and BAC molecules. *Journal of the Chemical Society (Resumed)* **1953,** (0), 2266-2288.

114. Yang, Y.; Chen, C.-C.; Scott, M. C.; Ophus, C.; Xu, R.; Pryor, A.; Wu, L.; Sun, F.; Theis, W.; Zhou, J.; Eisenbach, M.; Kent, P. R. C.; Sabirianov, R. F.; Zeng, H.; Ercius, P.; Miao, J., Deciphering chemical order/disorder and material properties at the single-atom level. *Nature* **2017,** *542*, 75.

115. Anderson, T. R.; Hawkins, E.; Jones, P. D., CO2, the greenhouse effect and global warming: from the pioneering work of Arrhenius and Callendar to today's Earth System Models. *Endeavour* **2016,** *40* (3), 178-187.

116. Rodriguez, J. A.; Liu, P.; Stacchiola, D. J.; Senanayake, S. D.; White, M. G.; Chen, J. G., Hydrogenation of CO2 to Methanol: Importance of Metal–Oxide and Metal–Carbide Interfaces in the Activation of CO2. *ACS Catalysis* **2015,** *5* (11), 6696-6706.

117. VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J., Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Computer Physics Communications* **2005,** *167* (2), 103-128.

118. Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J., CP2K: atomistic simulations of condensed matter systems. *Wires Comput Mol Sci* **2014,** *4* (1), 15-25.

119. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996,** *77* (18), 3865-3868.

120. Goedecker, S.; Teter, M.; Hutter, J., Separable dual-space Gaussian pseudopotentials. *Physical Review B* **1996,** *54* (3), 1703-1710.

121. VandeVondele, J.; Hutter, J., Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J Chem Phys* **2007,** *127* (11), 114105.

122. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* **2010,** *132* (15), 154104.

123. Tang, W.; Sanville, E.; Henkelman, G., A grid-based Bader analysis algorithm without lattice bias. *J. Phys.: Condens. Matter* **2009,** *21* (8), 084204.

124. Ruban, A. V.; Skriver, H. L.; Norskov, J. K., Surface segregation energies in transition-metal alloys. *Phys Rev B* **1999,** *59* (24), 15990-16000.

125. Ko, J.; Kim, B. K.; Han, J. W., Density Functional Theory Study for Catalytic Activation and Dissociation of CO2 on Bimetallic Alloy Surfaces. *J Phys Chem C* **2016,** *120* (6), 3438-3447.

126. Wang, S. G.; Liao, X. Y.; Cao, D. B.; Huo, C. F.; Li, Y. W.; Wang, J. G.; Jiao, H. J., Factors controlling the interaction of CO2 with transition metal surfaces. *J Phys Chem C* **2007,** *111* (45), 16934-16940.

127. Freund, H. J.; Roberts, M. W., Surface chemistry of carbon dioxide. *Surf Sci Rep* **1996,** *25* (8), 225-273.

128. Iczkowski, R.; Margrave, J. L., Electronegativity. *J Am Chem Soc* **1961,** *83* (17), 3547-3551.

129. Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E., Electronegativity - Density Functional Viewpoint. *J. Chem. Phys.* **1978,** *68* (8), 3801-3807.

130. Dean, J. A.; Lange, N. A., *Lange's handbook of chemistry*. 14th ed.; McGraw-Hill: New York, 1992.

131. Hammer, B.; Norskov, J. K., Theoretical surface science and catalysis - Calculations and concepts. *Adv Catal* **2000,** *45*, 71-129.

132. Hammer, B.; Norskov, J. K., Electronic factors determining the reactivity of metal surfaces. *Surf Sci* **1995,** *343* (3), 211-220.

133. Andriotis, A. N.; Mpourmpakis, G.; Broderick, S.; Rajan, K.; Datta, S.; Sunkara, M.; Menon, M., Informatics guided discovery of surface structure-chemistry relationships in catalytic nanoparticles. *J Chem Phys* **2014,** *140* (9), 094705.

134. Hong, Q. J.; Liu, Z. P., Mechanism of CO2 hydrogenation over Cu/ZrO2((2)over-bar12) interface from first-principles kinetics Monte Carlo simulations. *Surf Sci* **2010,** *604* (21-22), 1869-1876.

135. Zhao, N. N.; Pan, D. C.; Nie, W.; Ji, X. L., Two-phase synthesis of shape-controlled colloidal zirconia nanocrystals and their characterization. *J. Am. Chem. Soc.* **2006,** *128* (31), 10118-10124.

136. Kasatkin, I.; Girgsdies, F.; Ressler, T.; Caruso, R. A.; Schattka, J. H.; Urban, J.; Weiss, K., HRTEM observation of the monoclinic-to-tetragonal (m-t) phase transition in nanocrystalline ZrO2. *Journal of Materials Science* **2004,** *39* (6), 2151-2157.

137. Saha, S.; Abd Hamid, S. B., CuZrO3 nanoparticles catalyst in aerobic oxidation of vanillyl alcohol. *Rsc Advances* **2017,** *7* (16), 9914-9925.

138. Sinhamahapatra, A.; Jeon, J. P.; Kang, J.; Han, B.; Yu, J. S., Oxygen-Deficient Zirconia (ZrO2-x): A New Material for Solar Light Absorption. *Scientific Reports* **2016,** *6*.

139. Rasmussen, P. B.; Taylor, P. A.; Chorkendorff, I., The Interaction of Carbon-Dioxide with Cu(100). *Surf. Sci.* **1992,** *269*, 352-359.

140. Ernst, K. H.; Schlatterbeck, D.; Christmann, K., Adsorption of carbon dioxide on Cu(110) and on hydrogen and oxygen covered Cu(110) surfaces. *PCCP* **1999,** *1* (17), 4105-4112.

141. Redhead, P. A., Thermal desorption of gases. *Vacuum* **1962,** *12* (4), 203-211.

142. Bachiller-Baeza, B.; Rodriguez-Ramos, I.; Guerrero-Ruiz, A., Interaction of carbon dioxide with the surface of zirconia polymorphs. *Langmuir* **1998,** *14* (13), 3556-3564.

143. Pokrovski, K.; Jung, K. T.; Bell, A. T., Investigation of CO and CO2 adsorption on tetragonal and monoclinic zirconia. *Langmuir* **2001,** *17* (14), 4297-4303.

144. Jung, K. T.; Bell, A. T., The effects of synthesis and pretreatment conditions on the bulk structure and surface properties of zirconia. *Journal of Molecular Catalysis a-Chemical* **2000,** *163* (1-2), 27-42.

145. Goedecker, S.; Teter, M.; Hutter, J., Separable dual-space Gaussian pseudopotentials. *Phys Rev B Condens Matter* **1996,** *54* (3), 1703-1710.

146. Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R., RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chemical Physics Letters* **1998,** *294* (1-3), 143-152.

147. Weigend, F.; Häser, M., RI-MP2: first derivatives and global consistency. *Theoretical Chemistry Accounts* **1997,** *97* (1-4), 331-331.

148. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C., Electronic-structure calculations on workstation computers - the program system Turbomole. *Chemical Physics Letters* **1989,** *162* (3), 165-169.

149. Taylor, M. G.; Austin, N.; Gounaris, C. E.; Mpourmpakis, G., Catalyst Design Based on Morphology- and Environment-Dependent Adsorption on Metal Nanoparticles. *Acs Catal* **2015,** *5* (11), 6296-6301.

150. Wassenaar, J.; Jansen, E.; van Zeist, W.-J.; Bickelhaupt, F. M.; Siegler, M. A.; Spek, A. L.; Reek, J. N. H., Catalyst selection based on intermediate stability measured by mass spectrometry. *Nature Chemistry* **2010,** *2*, 417.

151. Yan, Z.; Taylor, M. G.; Mascareno, A.; Mpourmpakis, G., Size-, Shape-, and Composition-Dependent Model for Metal Nanoparticle Stability Prediction. *Nano Lett* **2018,** *18* (4), 2696-2704.

152. Austin, N.; Johnson, J. K.; Mpourmpakis, G., Au-13: CO Adsorbs, Nanoparticle Responds. *J Phys Chem C* **2015,** *119* (32), 18196-18202.

153. Delchev, Y. I.; Kuleff, A. I.; Maruani, J.; Mineva, T.; Zahariev, F. In *Strutinsky's Shell-Correction Method in the Extended Kohn-Sham Scheme: Application to the Ionization Potential, Electron Affinity, Electronegativity, and Chemical Hardness of Atoms*, Dordrecht, Springer Netherlands: Dordrecht, 2006; pp 159-176.

154. Kuhn, M., Building Predictive Models in R Using the caret Package. *2008* **2008,** *28* (5), 26.

155. R Core Team *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: Vienna, Austria, 2018.

156. 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.

157. Roling, T.; Abild-Pedersen, F., Structure-Sensitive Scaling Relations: Adsorption Energies from Surface Site Stability. *ChemCatChem* **2018,** *10* (7), 1643-1650.

158. Hammer, B.; Hansen, L. B.; Nørskov, J. K., Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys Rev B* **1999,** *59* (11), 7413-7421.

159. Zhao, Z.; Chen, Z.; Zhang, X.; Lu, G., Generalized Surface Coordination Number as an Activity Descriptor for CO2 Reduction on Cu Surfaces. *The Journal of Physical Chemistry C* **2016,** *120* (49), 28125-28130.

160. Tran, K.; Ulissi, Z. W., Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and H2 evolution. *Nature Catalysis* **2018,** *1* (9), 696-703.

161. Haynes, W. M., *CRC Handbook of Chemistry and Physics, 92nd Edition*. CRC Press: Hoboken, 2011.

162. Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F., Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **2017,** *355* (6321).

163. Andersen, M.; Levchenko, S. V.; Scheffler, M.; Reuter, K., Beyond Scaling Relations for the Description of Catalytic Materials. *Acs Catal* **2019,** *9* (4), 2752-2759.

164. Ma, X.; Xin, H., Orbitalwise Coordination Number for Predicting Adsorption Properties of Metal Nanocatalysts. *Physical Review Letters* **2017,** *118* (3), 036101.

165. Mehta, P.; Barboun, P.; Herrera, F. A.; Kim, J.; Rumbach, P.; Go, D. B.; Hicks, J. C.; Schneider, W. F., Overcoming ammonia synthesis scaling relations with plasma-enabled catalysis. *Nature Catalysis* **2018,** *1* (4), 269-275.

166. Vojvodic, A.; Medford, A. J.; Studt, F.; Abild-Pedersen, F.; Khan, T. S.; Bligaard, T.; Nørskov, J. K., Exploring the limits: A low-pressure, low-temperature Haber–Bosch process. *Chemical Physics Letters* **2014,** *598*, 108-112.

167. Porosoff, M. D.; Yan, B.; Chen, J. G., Catalytic reduction of CO2 by H2 for synthesis of CO, methanol and hydrocarbons: challenges and opportunities. *Energ Environ Sci* **2016,** *9* (1), 62-73.

168. Flytzani-Stephanopoulos, M., Gold Atoms Stabilized on Various Supports Catalyze the Water–Gas Shift Reaction. *Accounts Chem. Res.* **2014,** *47* (3), 783-792.

169. Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J., CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014,** *4* (1), 15-25.

170. VandeVondele, J.; Hutter, J., Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics* **2007,** *127* (11), 114105.

171. Kittel, C., *Introduction to solid state physics*. 8th ed.; Wiley: Hoboken, NJ, 2005; p 680.

172. McQuarrie, D. A., *Statistical Mechanics*. University Science Books: 2000; p 72.

173. Back, T., *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press, Inc.: 1996; p 314.

174. Ferrando, R.; Jellinek, J.; Johnston, R. L., Nanoalloys: from theory to applications of alloy clusters and nanoparticles. *Chem. Rev.* **2008,** *108* (3), 845-910.

175. Yang, Y.; Chen, C. C.; Scott, M. C.; Ophus, C.; Xu, R.; Pryor, A.; Wu, L.; Sun, F.; Theis, W.; Zhou, J.; Eisenbach, M.; Kent, P. R.; Sabirianov, R. F.; Zeng, H.; Ercius, P.; Miao, J., Deciphering chemical order/disorder and material properties at the single-atom level. *Nature* **2017,** *542* (7639), 75-79.

176. Larsen, P. M.; Jacobsen, K. W.; Schiøtz, J., Rich Ground-State Chemical Ordering in Nanoparticles: Exact Solution of a Model for Ag-Au Clusters. *Phys. Rev. Lett.* **2018,** *120* (25), 256101.

177. Jacobsen, K. W.; Stoltze, P.; Nørskov, J. K., A semi-empirical effective medium theory for metals and alloys. *Surf. Sci.* **1996,** *366* (2), 394-402.

178. Shevchenko, V. Y.; Madison, A. E., Structure of Nanoparticles: I. Generalized Crystallography of Nanoparticles and Magic Numbers. *Glass Phys. Chem.* **2002,** *28* (1), 40-43.

179. Zhang, Q.; Xie, J.; Yang, J.; Lee, J. Y., Monodisperse Icosahedral Ag, Au, and Pd Nanoparticles: Size Control Strategy and Superlattice Formation. *ACS Nano* **2009,** *3* (1), 139-148.

180. Radi, A.; Pradhan, D.; Sohn, Y.; Leung, K. T., Nanoscale Shape and Size Control of Cubic, Cuboctahedral, and Octahedral Cu−Cu2O Core−Shell Nanoparticles on Si(100) by One-Step, Templateless, Capping-Agent-Free Electrodeposition. *ACS Nano* **2010,** *4* (3), 1553-1560.

181. Jones, R.; Draheim, R.; Roldo, M., Silver Nanowires: Synthesis, Antibacterial Activity and Biomedical Applications. *Appl. Sci.* **2018,** *8*, 673.

182. Lewis, E. A.; Slater, T. J. A.; Prestat, E.; Macedo, A.; O'Brien, P.; Camargo, P. H. C.; Haigh, S. J., Real-time imaging and elemental mapping of AgAu nanoparticle transformations. *Nanoscale* **2014,** *6* (22), 13598-13605.

183. Bochicchio, D.; Ferrando, R., Size-Dependent Transition to High-Symmetry Chiral Structures in AgCu, AgCo, AgNi, and AuNi Nanoalloys. *Nano Lett.* **2010,** *10* (10), 4211-4216.

184. Batista, K. E. A.; Da Silva, J. L. F.; Piotrowski, M. J., Adsorption of CO, NO, and H_2 on the Pd_nAu_{55-n} Nanoclusters: A Density Functional Theory Investigation within the van der Waals D3 Corrections. *J. Phys. Chem. C* **2019,** *123* (12), 7431-7439.

185. Lopez, G. E.; Freeman, D. L., A study of low temperature heat capacity anomalies in bimetallic alloy clusters using J-walking Monte Carlo methods. *J. Chem. Phys.* **1993,** *98* (2), 1428-1435.

186. Subramanian, P. R.; Perepezko, J. H., The ag-cu (silver-copper) system. *J. Phase Equilib.* **1993,** *14* (1), 62-75.

187. Sopoušek, J.; Zobač, O.; Buršík, J.; Roupcová, P.; Vykoukal, V.; Brož, P.; Pinkas, J.; Vřešt'ál, J., Heat-induced spinodal decomposition of Ag–Cu nanoparticles. *Phys. Chem. Chem. Phys.* **2015,** *17* (42), 28277-28285.

188. Sopousek, J.; Pinkas, J.; Broz, P.; Bursik, J.; Vykoukal, V.; Skoda, D.; Styskalik, A.; Zobac, O.; Vrest'al, J.; Hrdlicka, A.; Simbera, J., Ag-Cu Colloid Synthesis: Bimetallic Nanoparticle Characterisation and Thermal Treatment. *J. Nanomater.* **2014,** *2014*, 1-13.

189. Davari, S. A.; Mukherjee, D., Kinetic Monte Carlo simulation for homogeneous nucleation of metal nanoparticles during vapor phase synthesis. *AIChE J.* **2018,** *64* (1), 18-28.

190. Reich, M.; Becker, U., First-principles calculations of the thermodynamic mixing properties of arsenic incorporation into pyrite and marcasite. *Chem. Geol.* **2006,** *225* (3), 278-290.

191. Guisbiers, G.; Khanal, S.; Ruiz-Zepeda, F.; Roque de la Puente, J.; José-Yacaman, M., Cu–Ni nano-alloy: mixed, core–shell or Janus nano-particle? *Nanoscale* **2014,** *6* (24), 14630-14635.

192. Shirinyan, A. S.; Wautelet, M., Phase separation in nanoparticles. *Nanotechnology* **2004,** *15* (12), 1720-1731.

193. Jayaganthan, R.; Chow, G. M., Thermodynamics of surface compositional segregation in Ni–Co nanoparticles. *Mater. Sci. Eng. B* **2002,** *95* (2), 116-123.

194. Vallée, R.; Wautelet, M.; Dauchot, J. P.; Hecq, M., Size and segregation effects on the phase diagrams of nanoparticles of binary systems. *Nanotechnology* **2001,** *12* (1), 68-74.

195. Christensen, A.; Stoltze, P.; Norskov, J. K., Size dependence of phase separation in small bimetallic clusters. *J. Phys. Condens. Matter* **1995,** *7* (6), 1047-1057.

196. Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Bjerre Jensen, P.; Kermode, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W., The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017,** *29* (27), 273002.

197. Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J., cp2k: atomistic simulations of condensed matter systems. *WIREs Computational Molecular Science* **2014,** *4* (1), 15-25.

198. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010,** *132* (15), 154104.

199. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc: Wallingford, CT, 2013.

200. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005,** *7* (18), 3297-3305.

201. Breiman, L., Random Forests. *Machine Learning* **2001,** *45* (1), 5-32.

202. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python 2012, p. arXiv:1201.0490. https://ui.adsabs.harvard.edu/abs/2012arXiv1201.0490P (accessed January 01, 2012).

203. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S., Self-Normalizing Neural Networks. *CoRR* **2017,** *abs/1706.02515.*

204. Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization 2014, p. arXiv:1412.6980. https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K (accessed December 01, 2014).

205. Glorot, X.; Bengio, Y., Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Yee Whye, T.; Mike, T., Eds. PMLR: Proceedings of Machine Learning Research, 2010; Vol. 9, pp 249--256.

206. LeCun, Y. A.; Bottou, L.; Orr, G. B.; Müller, K.-R., Efficient BackProp. In *Neural Networks: Tricks of the Trade: Second Edition*, Montavon, G.; Orr, G. B.; Müller, K.-R., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp 9-48.

207. Kittel, C., *Introduction to solid state physics*. 8th ed.; Wiley: Hoboken, NJ, 2005; p xix, 680 p.

208. Rumble, J. R.; Lide, D. R.; Bruno, T. J., *CRC handbook of chemistry and physics : a ready-reference book of chemical and physical data*. 2018.

209. Singh, B.; Sihag, P.; Singh, K., Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems and Environment* **2017,** *3* (3), 999-1004.

210. Li, J.; Heap, A. D.; Potter, A.; Daniell, J. J., Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software* **2011,** *26* (12), 1647-1659.

211. Cowan, M. J.; Mpourmpakis, G., Towards elucidating structure of ligand-protected nanoclusters. *Dalton Transactions* **2020,** *49* (27), 9191-9202.

212. Vanithakumari, S. C.; Nanda, K. K., Phenomenological Predictions of Cohesive Energy and Structural Transition of Nanoparticles. *The Journal of Physical Chemistry B* **2006,** *110* (2), 1033-1037.

213. Cheula, R.; Maestri, M.; Mpourmpakis, G., Modeling Morphology and Catalytic Activity of Nanoparticle Ensembles Under Reaction Conditions. *ACS Catalysis* **2020,** *10* (11), 6149-6158.

214. Bazhenov, A. S.; Honkala, K., Understanding Structure and Stability of Monoclinic Zirconia Surfaces from First-Principles Calculations. *Top Catal* **2017,** *60* (6-7), 382-391.

215. Christensen, A.; Carter, E. A., First-principles study of the surfaces of zirconia. *Phys Rev B* **1998,** *58* (12), 8050-8064.

216. Zhan, C. G.; Nichols, J. A.; Dixon, D. A., Ionization potential, electron affinity, electronegativity, hardness, and electron excitation energy: Molecular properties from density functional theory orbital energies. *J Phys Chem A* **2003,** *107* (20), 4184-4195.

217. Berkowitz, J.; Ellison, G. B.; Gutman, D., Three methods to measure RH bond energies. *The Journal of Physical Chemistry* **1994,** *98* (11), 2744-2765.

218. Ellison, G. B.; Engelking, P. C.; Lineberger, W. C., An experimental determination of the geometry and electron affinity of methyl radical. *J Am Chem Soc* **1978,** *100* (8), 2556-2558.

219. Erman, P.; Karawajczyk, A.; Rachlew-Källne, E.; Strömholm, C.; Larsson, J.; Persson, A.; Zerne, R., Direct determination of the ionization potential of CO by resonantly enhanced multiphoton ionization mass spectroscopy. *Chemical Physics Letters* **1993,** *215* (1), 173-178.

220. Refaey, K. M. A.; Franklin, J. L., Endoergic ion—molecule-collision processes of negative ions. III. Collisions of I− on O2, CO, and CO2. *International Journal of Mass Spectrometry and Ion Physics* **1976,** *20* (1), 19-32.

221. Schulz, P. A.; Mead, R. D.; Jones, P. L.; Lineberger, W. C., OH− and OD− threshold photodetachment. *The Journal of Chemical Physics* **1982,** *77* (3), 1153-1165.

222. Wiedmann, R. T.; Tonkyn, R. G.; White, M. G.; Wang, K.; McKoy, V., Rotationally resolved threshold photoelectron spectra of OH and OD. *The Journal of Chemical Physics* **1992,** *97* (2), 768-772.

223. Qi, F.; Sheng, L. S.; Zhang, Y. W.; Yu, S. Q.; Li, W. K., Experimental and Theoretical-Study of the Dissociation-Energies D-0(H2n-H) and D-0(H2n+-H) and Other Related Quantities. *Chemical Physics Letters* **1995,** *234* (4-6), 450-454.

224. Wasserstein, R. L.; Lazar, N. A., The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician* **2016,** *70* (2), 129-133.

225. Friedman, J.; Hastie, T.; Tibshirani, R., Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software* **2010,** *33* (1), 1-22.

226. Tibshirani, R., Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met* **1996,** *58* (1), 267-288.

227. Schmidt, M.; Lipson, H., Eureqa. **2014**.

228. Cramer, C. J., *Essentials of computational chemistry : theories and models*. J. Wiley: West Sussex, England ; New York, 2002; p 542.