

Computational Chemistry and Machine Learning-Assisted Screening of Supported Amorphous Metal Oxide Nanoclusters for Methane Activation

Xijun Wang, Kaihang Shi, Anyang Peng, and Randall Q. Snurr*

Cite This: *ACS Catal.* 2024, 14, 18708–18721

Read Online

ACCESS |



Metrics & More

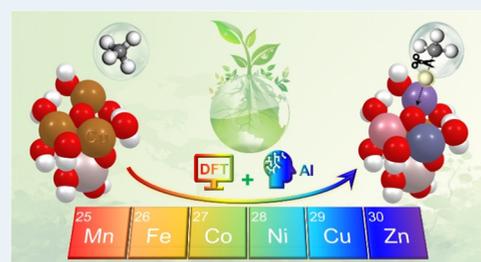


Article Recommendations



Supporting Information

ABSTRACT: Activating the C–H bond in methane represents a cornerstone challenge in catalytic research. While several supported metal oxide nanoclusters (MeO-NCs) have shown promise for this reaction, their optimal composition remains underexplored primarily due to the large number of possible compositions and their amorphous nature. This study addresses these challenges using computational approaches. Leveraging density functional theory (DFT) calculations, we began with a previously studied supported tetra-copper oxide nanocluster and systematically substituted its Cu sites with first-row transition metals (Mn, Fe, Co, Ni, and Zn). This process allowed us to examine the catalytic activity of 162 MeO-NCs with a variety of geometric and electronic structures, leading to 12 new compositions that outperformed the base nanocluster. Exploring the structure–activity relationships with machine learning, our analysis uncovered correlations between the intrinsic electronic and structural properties of the nanoclusters and the free energy barriers for methane activation despite the challenges posed by the structural flexibility of these amorphous nanoclusters. The results offer insights into the optimization of MeO-NCs for methane activation. Additionally, we developed a clustering model capable of distinguishing high-performing nanoclusters from less effective ones with strong tolerance to the interference from the structural flexibility of these amorphous nanoclusters. These findings help narrow down the material design space for more time-consuming high-level quantum chemical calculations, offering a promising pathway toward advancing eco-friendly methane conversion.



KEYWORDS: catalysis, density functional theory, machine learning, oxidation, material discovery

1. INTRODUCTION

Methane, the principal component of natural gas, holds significant potential as a chemical feedstock due to its high abundance.¹ The direct combustion of methane releases carbon dioxide, a greenhouse gas, exacerbating global climate change concerns. Thus, there is a pressing need to utilize methane in a greener manner. For instance, converting methane to liquid products like methanol not only facilitates easier transportation but also offers a pathway to value-added chemicals.^{2–4} However, methane activation, the initial step for methane conversion, is intrinsically challenging. The C–H bonds in methane, with a bond energy of 435 kJ/mol,⁵ are among the strongest in hydrocarbons. Consequently, the development of efficient catalyst materials for methane activation remains critical in energy and chemical research.

Metal oxide nanoclusters (MeO-NCs), especially those based on copper and supported within metal–organic frameworks (MOFs) and zeolites, have recently garnered significant attention in the realm of methane activation. For instance, Lercher et al. efficiently converted methane to methanol under mild conditions using a series of mono- and dicopper oxide nanoclusters dispersed on specific MOFs and zeolites.^{6–10} The structures of these Cu-oxo motifs have been

systematically characterized by using multiple tools such as EXAFS and XANES, and their kinetic behaviors across the full catalytic cycle of methane-to-methanol conversion have been examined. Román-Leshkov et al. also showcased the exceptional capability of [Cu–O–Cu]²⁺ motifs within copper-exchanged zeolites for continuous and selective conversion of methane to methanol.^{11,12} Moreover, several theoretical studies have delved into the catalytic mechanisms of Cu-oxo moieties for this reaction.^{13–17} These pioneering studies highlight the promise of copper oxide nanoclusters in efficient methane activation and conversion and also provide solid foundations for studying variants of copper oxide nanoclusters.

One way to build on the work to date on copper oxide nanoclusters is to explore other compositions.^{18,19} Challenges arise from a limited understanding of the complex structure–

Received: July 6, 2024

Revised: October 15, 2024

Accepted: November 26, 2024

property relationships, which hinders the rational design of nanoclusters. Further complicating the matter is the vast potential material design space, making a “brute force” exploration of all possible compositions and configurations an undesirable path. Moreover, the dynamic behavior of nanoclusters under the reaction conditions often results in amorphous structures. These structures may be characterized by a large number of local minima on the potential energy surface, influencing their catalytic activity and stability. Thus, realizing the full promise of MeO-NCs for methane activation and other reactions necessitates overcoming these intertwined challenges.

First-principles calculations at the level of density functional theory (DFT) have been used extensively to quantify structure–property relationships at the atomic scale, enabling researchers to predict and elucidate the catalytic behaviors of nanoclusters with sufficient accuracy at a reasonable cost.^{20,21} On another front, machine learning excels at uncovering intrinsic correlations between various properties,^{22–25} facilitating deeper insights into DFT-computed data sets. Combining DFT and machine learning methods is a promising strategy for accelerating progress in the design and discovery of new catalysts and improving our understanding of their function. This may be particularly beneficial when potential materials reside beyond the established boundaries of our existing chemical knowledge.

Our previous research systematically examined the coordination environments, configurations, and catalytic behaviors of various supported di- and tetra-copper nanoclusters for methane activation.^{26–28} It is envisioned that these well-defined nanoclusters could be synthesized by atomic layer deposition (ALD) on supports containing isolated –OH groups using the experimental procedure outlined by Avila et al.,²⁹ and our DFT thermodynamic analysis²⁸ supports the feasibility of their synthesis. These nanoclusters are, thus, members of an extended class of possible nanoclusters of varying numbers of metal atoms and varying composition.

Herein, we selected a representative tetra-copper oxide nanocluster from our previous work²⁸ as a starting point. This cluster's activity in methane activation was confirmed using DFT transition state calculations. The chosen cluster features a tricoordinated oxygen site surrounded by three Cu sites, labeled M_1 , M_2 , and M_3 , with a fourth Cu site positioned further from the oxygen (see Figure 1 and Section 1 in the Supporting Information). The stability of this configuration under practical conditions (200 °C) has been corroborated by *ab initio* molecular dynamic (AIMD) simulations, ensuring a solid base for further modifications. Here, we substituted the three Cu sites with six different first-row transition metals (Mn, Fe, Co, Ni, and Zn), resulting in 216 distinct MeO-NCs ($6 \times 6 \times 6 = 216$). Each structure is named according to the metal types at the M_1 , M_2 , and M_3 sites. For instance, the base nanocluster is termed **CuCuCu** (Figure S1). We performed these strategic substitutions with the aim of reducing the free energy barrier associated with methane activation at the tricoordinated oxygen site. After investigating methane activation on these nanoclusters using DFT, we employed supervised machine learning algorithms to find correlations between the structural and electronic properties of the reactant cluster and the free energy barriers for methane activation. We also used an unsupervised machine learning clustering algorithm to classify the nanoclusters into two distinct groups: one for the high-performing nanoclusters and one for the less

reactive nanoclusters. This approach can be expanded in the future to efficiently screen out less promising nanoclusters—for example, with other metals or larger sizes—reducing the number of computationally intensive transition state (TS) calculations and narrowing the material search space.

2. COMPUTATIONAL DETAILS

2.1. DFT Calculations. All DFT calculations for the MeO-NC models were implemented using Gaussian 16.³⁰ We optimized geometries, computed vibrational frequencies, and searched for transition states using the M06-L functional³¹ with ultrafine integration grids. Carbon, hydrogen, and oxygen were modeled using the def2-SVP basis set,³² while transition metals were represented with the def2-TZVP basis set³³ and its corresponding effective core potential. This choice of functional and basis sets is known for its balance of accuracy and computational cost, as evidenced by previous studies on MeO-NC systems,^{6,26–28,34} where it provided accurate geometric and electronic structure insights. The accuracy of the def2-SVP basis set was validated through test calculations comparing the free energy barriers of methane activation for five test systems using the more accurate def2-TZVP basis set for all of the atoms. The results, shown in Table S1, closely align with our original calculations. This confirms the reliability of using the def2-SVP basis set in our study. Additionally, we incorporated the D3 dispersion correction with zero damping to account for dispersion forces.³⁵

To investigate magnetic couplings, we explored potential high spin (ferromagnetic), ferrimagnetic, and antiferromagnetic broken symmetry states for **CuCuCu**, **CoMnFe**, **CuZnFe**, **CuZnNi**, and **FeFeZn** as test cases. The results revealed a preference for ferromagnetic coupling in most cases, with **CuZnNi** being the only exception, where the ferrimagnetic state is marginally more stable (by 0.03 eV) than the ferromagnetic state (Figure S2 and Table S2). These findings are consistent with previous studies on the magnetic couplings in copper oxide nanoclusters and nanoparticles.^{26,36–38} Accurately examining all possible spin states for all 216 MeO-NCs would require a prohibitive amount of computing resources. To make this study feasible, we adopted ferromagnetic couplings for all of the MeO-NCs. We carefully examined a range of possible spin multiplicities during geometric optimization to determine the ground spin state, as detailed in Table S3.

The nature of all stationary points on the potential energy landscape, including both minima and transition states, was confirmed through harmonic vibrational analysis, which was also used to calculate the thermal corrections. To estimate the Gibbs free energy, we considered thermal and entropic factors derived from vibrational frequency analysis using the harmonic oscillator model. We adjusted frequencies under 50 cm^{-1} to a standard value of 50 cm^{-1} , to maintain the validity of this model at low frequencies.^{39,40} For all nanoclusters and bound species, we intentionally disregarded the translational and rotational contributions to mimic the behavior of the solid-state species.

2.2. Sure Independence Screening and Sparsifying Operator (SISSO) method. We employed the SISSO algorithm⁴¹ to establish the mathematical relationship between free energy barriers and input electronic and geometric features. SISSO is a powerful tool capable of uncovering the mathematical correlations between a set of input features and target properties by employing a framework rooted in

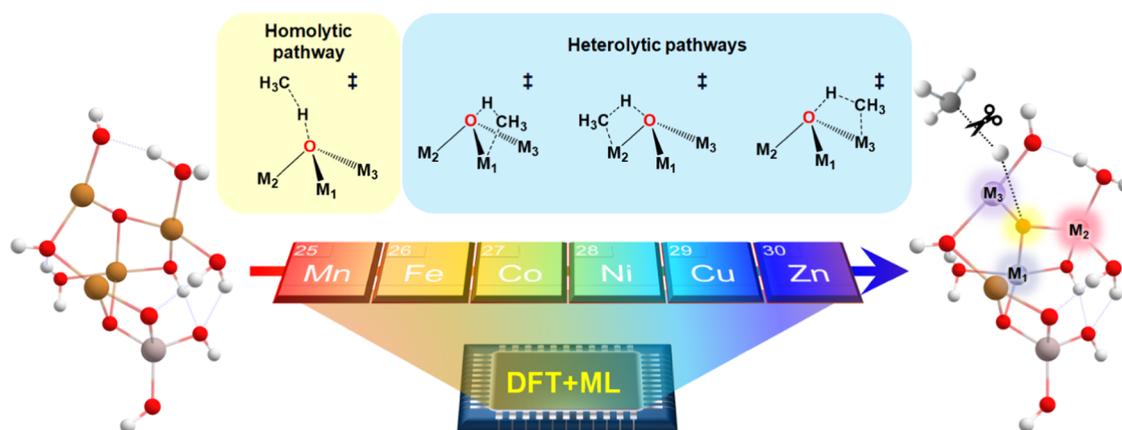


Figure 1. Schematic illustration of the substitution of the three Cu sites (labeled as M_1 , M_2 , and M_3) surrounding the tricoordinated active oxygen site (highlighted in yellow at right), aiming to enhance methane activation, with emphasis on the homolytic and heterolytic pathways. Note that the $-OH$ group at the bottom of the cluster represents the location where the cluster connects to the underlying support. Previous work²⁶ has validated that the catalytic site at the top of the cluster is sufficiently far from the support that the support may be neglected in the DFT calculations. White, red, golden, and light peach spheres represent H, O, Cu, and Al atoms, respectively.

compressed-sensing-based dimensionality reduction. Within the framework of supervised learning tasks in this study, SISSO operates in several key steps. First, the SISSO algorithm constructs a feature space by combining features using specific mathematical operators, namely, $\hat{H}^{(m)} \equiv \{I, +, -, \times, \div, \log, \exp, \exp^{-1}, \sqrt{\quad}, |\cdot|, \sin, \cos\}$. Only combinations that make physical sense, such as those involving features with compatible units, are retained and are indicated by the (m) notation. By employing these operators, a wide array of nonlinear expressions emerges from the amalgamation of features and mathematical operations, creating an expansive candidate space for subsequent analysis. In this study, the feature complexity parameter within SISSO is set to 3, which determines the size of the feature space selected by the algorithm. This setting aims to identify the most relevant features that correlate with the target property of interest. By setting it to 3, we achieve a balance between model complexity and computational efficiency, ensuring that the selected descriptors are optimally representative without excessively consuming computational resources. The descriptor dimension is also set to 3, allowing descriptors to use up to three terms. Under these settings, the average time to generate the one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) mapping equations in this study was notably fast, at 0.01, 0.13, and 76.75 s, respectively.

Next, the sure independence screening (SIS) technique, known for its effectiveness in feature selection, is applied. It ranks the descriptors by assessing their correlation with the target property and screens them based on their independence from the target variable. This process results in the selection of a subset of descriptors that demonstrate strong correlations with the target variable. Following the SIS stage, a sparsifying operator (SO) is introduced to further encourage sparsity in the descriptor space. The SO encourages most of the descriptor coefficients to approach zero or become zero, effectively reducing the dimensionality of the problem. This reduction in dimensionality helps streamline the analysis and enhances the efficiency of the model. The models' predictive accuracy and error margins were assessed via Pearson correlation coefficients (r) and root-mean-square errors (RMSE).

Unlike black-box machine learning methods like artificial neural networks, SISSO can reveal mathematical mappings that may provide human-understandable physical insights, making it valuable for finding meaningful descriptors in physical and chemical applications. It differs from other symbolic regression techniques like genetic algorithms and random search by conducting an exhaustive search of the solution space, reducing bias, and yielding low-complexity descriptors that are robust to data noise. More importantly, SISSO excels at handling smaller training data sets, effectively addressing the overfitting challenges often encountered with small-data-based ML predictions. This sets it apart from many ML algorithms that typically require extensive data to achieve reliable performance,^{42,43} making it a suitable choice for this study.

2.3. Principal Component Analysis (PCA). The visualization of the clustering outcomes employs the PCA method,⁴⁴ a well-established algorithm for reducing data dimensionality. It projects multidimensional input features onto a two-dimensional plane, thereby simplifying the depiction of the results. The essence of PCA involves converting features from an n -dimensional space to a reduced k -dimensional space, termed principal components. For instance, consider a data set $X = \{x_1, x_2, x_3, \dots, x_n\}$, which we aim to reduce to k dimensions. The process begins by centering the data, a step involving subtraction of the mean. Subsequently, we calculate the covariance matrix using the formula $\frac{1}{n}XX^T$. Following this, we determine the covariance matrix's eigenvalues and eigenvectors through eigen decomposition. The eigenvalues are then arranged in descending order, with the top k values being selected for further processing. Corresponding to these top k eigenvalues, an eigenvector matrix P is formed with these eigenvectors as its rows. Ultimately, the original data set is transposed into a new dimensionality defined by these chosen eigenvectors, expressed as $Y = PX$.

2.4. Balanced Iterative Reducing and Clustering Using Hierarchies (BIRCH) Method. In this work, we utilized the balanced iterative reducing and clustering using hierarchies (BIRCH) algorithm⁴⁵ for data clustering implemented using Scikit-learn.^{46,47} BIRCH, known for its effectiveness in clustering relatively small to medium-sized

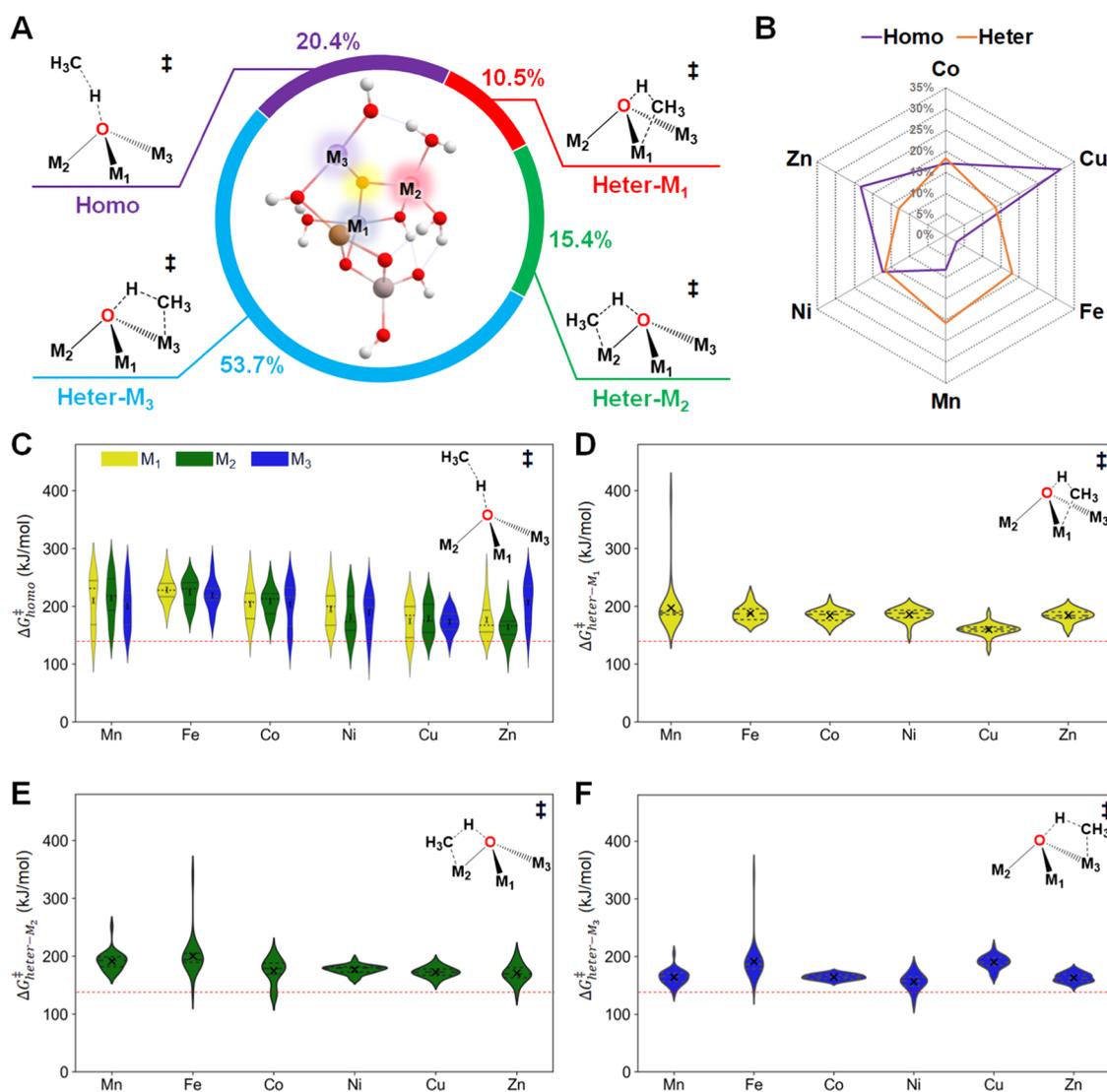


Figure 2. (A) Percentage of MeO-NCs favoring the homolytic pathway and the three heterolytic pathways. (B) Radar chart illustrating the percentage of each type of transition metal in the M_1 , M_2 , and M_3 sites, comparing those favoring homolytic versus heterolytic pathways. Violin plots illustrating the distribution of (C) $\Delta G_{\text{homo}}^{\ddagger}$, (D) $\Delta G_{\text{heter-M}_1}^{\ddagger}$, (E) $\Delta G_{\text{heter-M}_2}^{\ddagger}$ and (F) $\Delta G_{\text{heter-M}_3}^{\ddagger}$ on the tricoordinated oxygen site based on metal types at the M_1 (yellow), M_2 (green), and M_3 (blues) sites. In the violin plots, the mean values are denoted by \times , while the quartile lines are represented by black dashed lines. The $\Delta G_{\text{homo}}^{\ddagger}$ value for CuCuCu (139 kJ/mol) is marked with a red dashed line.

data sets, was chosen for its capability to handle noise in the data.

We applied the BIRCH algorithm as implemented in Scikit-learn, adhering to the default settings. These settings include a threshold value of 0.5, which determines the maximum radius of subclusters in the CF (Clustering Feature) tree, and a branching factor of 50, which dictates the number of child nodes in the tree. The algorithm operates in a two-step process: initially constructing a CF tree by dynamically adjusting clusters and then performing a global clustering phase on the leaf nodes of the CF tree.

3. RESULTS AND DISCUSSION

3.1. Catalytic Activity Evaluation and Mechanism Preference. For each of the 216 MeO-NCs, we identified its ground spin state by evaluating the energies across possible ferromagnetic spin multiplicities and opting for the one with the lowest energy, as listed in Table S3. The catalytic activity of each MeO-NC was characterized by its free energy barrier for

activating the C–H bond along two widely accepted mechanisms:^{48–50} homolytic and heterolytic. As depicted in Figure 1, in the homolytic pathway, also known as hydrogen atom transfer (HAT), a C–H bond in methane dissociates homolytically, releasing a hydrogen atom to the active oxygen site and leaving a CH_3 radical with an unpaired electron. Alternatively, the heterolytic pathway involves a proton-coupled electron transfer (PCET) process, cleaving the C–H bond heterolytically. This process transfers a proton to the active oxygen site, simultaneously directing the negatively charged methyl species toward bonding on one of the three adjacent metal sites, which acts as a Lewis acid center. Our previous studies,^{27,28} combined with other studies,^{51–53} highlight the need to consider both mechanisms when evaluating methane activation on MeO-NCs.

For each of the clusters, we computed the free energy barriers for methane activation along the homolytic pathway, represented as $\Delta G_{\text{homo}}^{\ddagger}$, and the three heterolytic pathways, with CH_3 bonding to M_1 ($\Delta G_{\text{heter-M}_1}^{\ddagger}$), M_2 ($\Delta G_{\text{heter-M}_2}^{\ddagger}$), and

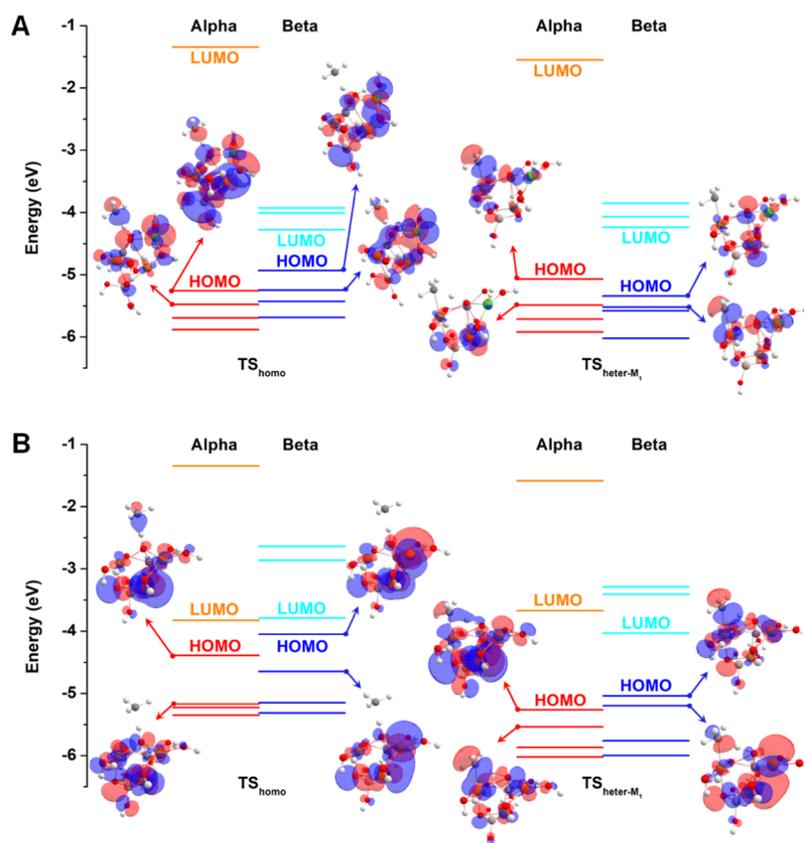


Figure 3. Energy levels and spatial distributions of the frontier molecular orbitals in the TS structures associated with the homolytic (left) and heterolytic (right) pathways for (A) CuZnNi and (B) CuZnFe. Refer to Figure S7 for larger images depicting the spatial distributions of the frontier molecular orbitals.

M_3 ($\Delta G_{\text{heter}-M_3}^\ddagger$) sites. These calculations were based on TS calculations at 200 °C, a standard temperature for experimental methane activation.⁵⁴ Progressing to the evaluation of each nanocluster, we determined the minimum of the four free energy barriers ($\Delta G_{\text{min}}^\ddagger$) to assess the catalytic activity. Of the 216 MeO-NCs considered, 54 were found to be unsuitable for further analysis due to significant structural changes when exposed to methane. As shown in the percentage distribution of transition metals at the M_1 , M_2 , and M_3 sites (Figure S3), these unstable structures are slightly more likely to contain Fe or Zn. The observed structural changes include proton transfer among peripheral water or OH groups and the breaking of one of the three bonds connecting the active oxygen site to adjacent metal sites upon interaction with methane. These changes indicate that such nanoclusters cannot be easily stabilized at their initial configuration in a methane-abundant environment, making it challenging to reliably estimate their methane activation barriers. Consequently, our focus shifted to the remaining 162 MeO-NCs (Table S4 and Figures S4–S6). Among them, 12 stood out for having a lower $\Delta G_{\text{min}}^\ddagger$ than the base material CuCuCu, which has a value of 139 kJ/mol. Remarkably, the best performer, CuZnNi, showed a $\Delta G_{\text{min}}^\ddagger$ of 113 kJ/mol. Delving deeper into the atomic makeup of these 12 high-performing clusters, we noticed a recurring presence of Cu atoms: 9 of the 12 nanoclusters contained at least one Cu atom at the M_1 , M_2 , or M_3 sites. This pattern supports previous interest in using copper-based nanoclusters for methane activation. Interestingly, a notable synergy was observed in the nanoclusters combining Cu and Zn, as half of these top

clusters, including the three most efficient ones, CuZnNi, CuZnFe, and CuCoZn, comprised CuZn combinations.

Building on these insights, we then turned our focus to examine the clusters' preference for homolytic versus heterolytic pathways in methane activation. As depicted in Figure 2A, our analysis revealed that 20.4% (33 of 162) of these MeO-NCs preferred the homolytic pathway. In comparison, 10.5% (17 of 162) favored the heterolytic pathway with CH_3 bonding on the M_1 site, 15.4% (25 of 162) on the M_2 site, and a substantial 53.7% (87 of 162) on the M_3 site. This marked preference for the M_3 site can be attributed to its tricoordinated nature, in contrast to the tetra-coordinated M_1 and M_2 sites. As one might expect, less coordinatively saturated metal sites are more inclined to bond with CH_3 . Interestingly, among the 33 MeO-NCs that favor the homolytic pathway, the proportion of Cu atoms at the M_1 , M_2 , and M_3 sites of these MeO-NCs is 31%, which is higher than that of Zn (23%) and significantly exceeds the fractions of the other transition metals (Figure 2B). Also, when M_1 , M_2 , and M_3 contain two or three Cu atoms, the reaction tends to favor the homolytic pathway. These statistical results imply that Cu-containing nanoclusters have a predisposition for the homolytic pathway, which is consistent with the observations made for various mono- and dicopper nanoclusters.^{26–28} In comparison, when focusing on MeO-NCs that favor the heterolytic pathways, the distribution among Mn, Fe, Co, and Ni showed no distinct bias (Figure 2B).

Figure 2C showcases the distribution of $\Delta G_{\text{homo}}^\ddagger$ values across the 33 MeO-NCs favoring the homolytic pathway, segmented by the metal types at the M_1 , M_2 , and M_3 sites.

(Since most nanoclusters contain multiple types of metal atoms, each value of $\Delta G_{\text{hom}}^{\ddagger}$ shows up in multiple violins.) The results show that MeO-NCs with Mn and Cu on M_1 , Ni, and Zn on M_2 , and Co and Ni on M_3 presented a slightly higher likelihood of possessing a $\Delta G_{\text{hom}}^{\ddagger}$ lower than the value for the base material **CuCuCu** of 139 kJ/mol (red dashed line). However, we note that the $\Delta G_{\text{hom}}^{\ddagger}$ value is collectively determined by the three metal sites rather than by a single metal site. This is evident from the wide range of $\Delta G_{\text{hom}}^{\ddagger}$ values, which span from ~ 100 to ~ 300 kJ/mol for each metal on the M_1 , M_2 , or M_3 sites. In sharp contrast, the violin plots in Figure 2D–F show more localized $\Delta G_{\text{heter}}^{\ddagger}$ values. For example, nearly all of the MeO-NCs favoring the heterolytic pathway at M_1 exhibit $\Delta G_{\text{heter}-M_1}^{\ddagger}$ values close to 200 kJ/mol when M_1 is Co (Figure 2D). Such narrow distributions emphasize the pivotal role of the metal site interacting with CH_3 in determining the $\Delta G_{\text{heter}}^{\ddagger}$ values, aligning well with our chemical intuition.

To elucidate the preference for the homolytic versus heterolytic pathways, we delved further into two exemplary systems. The first, **CuZnNi**, stands out as the most active among the 162 MeO-NCs. It favors the homolytic mechanism with a $\Delta G_{\text{hom}}^{\ddagger}$ value of 113 kJ/mol, which is lower than its $\Delta G_{\text{heter}-M_1}^{\ddagger}$ value of 140 kJ/mol. The second, **CuZnFe**, is also notably active but favors the heterolytic pathway, with values of $\Delta G_{\text{hom}}^{\ddagger} = 174$ kJ/mol and $\Delta G_{\text{heter}-M_1}^{\ddagger} = 126$ kJ/mol. To shed light on the pathway preferences, we investigated the frontier molecular orbitals for their TS structures, which are crucial for understanding bonding strength and reactivity. For **CuZnNi**, as illustrated in Figure 3A, we observed pronounced σ bonding orbitals in the spin-up (α) HOMO and HOMO–1, with complementary spin-down (β) HOMO–1 in the homolytic TS structure. However, the heterolytic TS counterpart paints a different picture: although the spin-up HOMO facilitates proton transfer, the Cu– CH_3 bonding is weakened by its σ^* antibonding nature. Such σ^* antibonding between Cu and CH_3 was also found in the spin-down HOMO–1, in tandem with the nonbonding orbitals, such as the spin-up HOMO–1 and spin-down HOMO. This electronic structure renders the heterolytic pathway less energetically favorable. For comparison, Figure 3B illustrates the electronic structures of the homolytic and heterolytic TS structures of **CuZnFe**. For the homolytic TS, the spin-up HOMO–1 and the spin-down HOMO and HOMO–1 are characterized as nonbonding orbitals. Conversely, in the heterolytic TS, the spin-up HOMO–1 and the spin-down HOMO and HOMO–1 exhibit a strengthened bonding interaction between Cu and CH_3 , thus, making the heterolytic pathway more energetically favorable. These two exemplary systems show how the tuning of electronic structures significantly influences the preference for either homolytic or heterolytic pathways in methane activation.

3.2. Supervised Machine Learning for Probing Structure–Property Relationships. Next, we employed machine learning techniques to establish a connection between the inherent characteristics of the active oxygen site and the four transition metal sites and their critical role in catalytic hydrogen abstraction. Our approach leveraged various input features derived from the DFT-optimized initial structure of the MeO-NCs. These features encompassed the partial charge (q) and spin density (ρ) associated with the tricoordinated oxygen and the four transition metals, in addition to the p-

band center of the tricoordinated oxygen (ϵ_p), the d-band center of the transition metal atoms (ϵ_d), the HOMO–LUMO gap of the entire nanocluster (E_g), the number of d-electrons (N_d) for the metal elements on the M_1 , M_2 , and M_3 sites, as well as crucial bond lengths (d_{M_1-O} , d_{M_2-O} , d_{M_3-O}), bond angles ($\angle_{M_1OM_2}$, $\angle_{M_1OM_3}$, $\angle_{M_2OM_3}$), and the dihedral angle $\theta_{M_1M_2M_3O}$ (see Table S5 for detailed data on these parameters). The objective of this study was to correlate these electronic and structural parameters with the free energy barriers associated with both homolytic and heterolytic mechanisms, thus shedding light on the catalytic process at the active oxygen sites.

For this task, we divided the 162 MeO-NCs into 80% training and 20% test sets and employed a widely used supervised ML method, the extra-trees regression (ETR),⁵⁵ which is a tree-based method recognized for its effectiveness in predicting material properties. However, the results showed persistent overfitting that could not be resolved by modulating the complexity of the models (refer to Section 2 and Figures S8–S10 in the Supporting Information for more details). This overfitting likely arises from the limited training data available, a constraint imposed by the computationally intensive nature of the DFT.

To overcome this limitation, we adopted the SISO method,⁴¹ a state-of-the-art technique noted for its effectiveness in handling small data sets^{42,43} and commonly used in the context of material science.^{56,57} A key advantage of SISO is its ability to derive explicit mathematical expressions connecting input features to target properties. In our implementation of SISO, we evaluated descriptors in 1D, 2D, and 3D forms (i.e., descriptors employing a single term, two terms, and three terms), each representing an increasing level of complexity in the mathematical fitting. The effectiveness of this approach is demonstrated in Figure S11, which shows that overfitting was significantly reduced. This suggests improved generalization and robustness of the SISO-based predictions compared with those obtained from ETR. Despite these improvements, the predictive performance on the test set remains suboptimal, with r values still below 0.8.

A potential factor impacting the accuracy of our predictions is the inherent structural variability in amorphous nanoclusters. This is particularly evident in the atoms peripheral to the active oxygen site, including the dangling OH groups, H_2O groups, and minor, yet nonnegligible, structural disturbances upon methane approaching. Such structural flexibility can lead to substantial energy fluctuations in the MeO-NCs, making it challenging to predict the free energy barriers accurately using ML.

To mitigate the impact of the structural flexibility of peripheral atoms and primarily focus on the structure–property relationships at the active oxygen center, a tractable strategy is to concentrate on structures exhibiting minimal structural changes from their initial states to their TS. For the next stage of our research, we decided to discard any TS structure where the distance between any O–H pair changed by more than 0.5 Å from the initial configuration. The SISO prediction results for these discarded structures are expectedly poor (refer to Figure S12 for more details). As a result, we retained a refined set of 47, 66, 29, and 35 TS structures (see Table S6 for the data), totaling 177, for the homolytic pathway and the three heterolytic pathways, respectively, and retrained the model with this subset.

Table 1. SISSO-Suggested Best-Performing Mathematical Expressions Connecting the Electronic and Structural Features and the Target Free Energy Barriers for the 177 TS Structures with Minimal Structural Changes during Methane Activation

| target | dimension of descriptor | mathematical expression |
|--|-------------------------|--|
| $\Delta G_{\text{homo}}^{\ddagger}$ | 1D | $\Delta G_{\text{homo}}^{\ddagger} = 5.304 \frac{\exp(-\rho_{M_3})}{q_{\text{O}} + q_{\text{Cu}}} + 214.842$ |
| | 2D | $\Delta G_{\text{homo}}^{\ddagger} = 5.003 \frac{\exp(-\rho_{M_3})}{q_{\text{O}} + q_{\text{Cu}}} + 8.013 \frac{q_{M_3}}{\epsilon_{dM_2} \rho_{\text{Cu}} - \rho_{M_1}} + 220.973$ |
| | 3D | $\Delta G_{\text{homo}}^{\ddagger} = 4.074 \frac{\exp(-\rho_{M_3})}{q_{\text{O}} + q_{\text{Cu}}} - 2.056 \frac{q_{M_3}^3}{\rho_{\text{Cu}} - \rho_{M_1}} - 0.027 \frac{\angle_{M_1\text{OM}_2} \angle_{M_1\text{OM}_3}}{q_{\text{O}} \cdot \epsilon_{\text{PO}}} + 311.263$ |
| $\Delta G_{\text{heter-M}_1}^{\ddagger}$ | 1D | $\Delta G_{\text{heter-M}_1}^{\ddagger} = -370.236 \frac{N_{dM_1} + N_{dM_2}}{q_{M_1} \angle_{M_1\text{OM}_2}} + 264.802$ |
| | 2D | $\Delta G_{\text{heter-M}_1}^{\ddagger} = -0.225 \frac{\rho_{M_1}^2}{\cos(N_{dM_1})} - 37534.484 \frac{\exp(\epsilon_{\text{PO}})}{\angle_{M_1\text{OM}_2} E_{\text{g}}} + 205.149$ |
| | 3D | $\Delta G_{\text{heter-M}_1}^{\ddagger} = -380.509 \frac{\exp(\epsilon_{\text{PO}})}{q_{M_1} E_{\text{g}}} - 0.188 \frac{\rho_{M_1}^2}{\cos(N_{dM_1})} + 304.315 \ln(q_{M_2}) \cdot \rho_{\text{O}} \cdot \rho_{\text{Cu}}$ $+ 201.268$ |
| $\Delta G_{\text{heter-M}_2}^{\ddagger}$ | 1D | $\Delta G_{\text{heter-M}_2}^{\ddagger} = 63.363 \frac{\angle_{M_1\text{OM}_3} q_{M_2}}{\angle_{M_1\text{OM}_2} q_{\text{O}}} + 79.754$ |
| | 2D | $\Delta G_{\text{heter-M}_2}^{\ddagger} = -0.837 \left(\frac{\angle_{M_1\text{OM}_3}}{q_{\text{O}}} + \frac{\angle_{M_1\text{OM}_2}}{q_{M_2}} \right) - 0.00133 d_{M_1-\text{O}}^6 \cdot \angle_{M_2\text{OM}_3}$ $\cdot \epsilon_{dM_2} + 115.685$ |
| | 3D | $\Delta G_{\text{heter-M}_2}^{\ddagger} = 0.879 (q_{\text{O}} \cdot \angle_{M_1\text{OM}_2} + q_{M_2} \cdot \angle_{M_1\text{OM}_3}) - 0.065 d_{M_1-\text{O}}^7 \cdot \epsilon_{\text{PO}}$ $+ 21.642 \left \frac{\rho_{\text{O}} - \rho_{M_3}}{\exp(\rho_{M_3})} \right + 111.843$ |
| $\Delta G_{\text{heter-M}_3}^{\ddagger}$ | 1D | $\Delta G_{\text{heter-M}_3}^{\ddagger} = 28.333 \sin^6 N_{dM_3} + 162.699$ |
| | 2D | $\Delta G_{\text{heter-M}_3}^{\ddagger} = 33.536 \sin^6 N_{dM_3} - 3.023 \frac{(d_{M_2-\text{O}} - d_{M_3-\text{O}}) N_{dM_2}}{E_{\text{g}}} + 164.421$ |
| | 3D | $\Delta G_{\text{heter-M}_3}^{\ddagger} = 90.166 \cos(\cos(N_{dM_3})) + 877.857 \frac{(d_{M_2-\text{O}} - d_{M_3-\text{O}})}{\theta_{M_1M_2M_3\text{O}} E_{\text{g}}}$ $- 5.729 \left(\cos(\rho_{M_2}) + \frac{\epsilon_{dM_1}}{\epsilon_{dM_3}} \right) + 109.373$ |

The SISSO prediction results, illustrated in Table 1 and Figure 4, demonstrated a promising performance. Initially, the use of 1D descriptors was too simplistic to capture the complicated interactions among the various chemical factors, leading to underfitting in our models (left panels). Transitioning to 2D descriptors resulted in a significant improvement in prediction performances (middle panels). The r values for all training data sets were close to or exceeded 0.9, with the highest RMSE reaching just 10.6 kJ/mol. For the test data sets, r values remained above 0.85, with RMSEs for heterolytic pathways ($\Delta G_{\text{heter-M}_1}^{\ddagger}$, $\Delta G_{\text{heter-M}_2}^{\ddagger}$, $\Delta G_{\text{heter-M}_3}^{\ddagger}$) staying below 13 kJ/mol, while $\Delta G_{\text{homo}}^{\ddagger}$ reached 19.8 kJ/mol. Advancing to models with 3D descriptors (right panels), we observed further improvements: r values for both training and test data sets were close to or above 0.9, and all RMSE values were below 10 kJ/mol, except for $\Delta G_{\text{homo}}^{\ddagger}$, which stood at 14.7 kJ/mol. Notably, the increased complexity with 2D and 3D descriptors effectively minimized overfitting, evidenced by the small difference in r values (less than 0.08) between training and test data sets across all predictions. Although SISSO yields satisfactory predictions for the 177 TSs with minimal structural variations, making predictions for nanoclusters with significant structural changes (O–H pair distance changes exceeding 0.5 Å during methane activation) remains challenging (although the results in Figure S11 might be considered sufficiently accurate for initial screening). This is due to the substantial randomness introduced during the structural relaxation of

these systems, resulting in r values generally lower than 0.5 (Figure S13).

Furthermore, the features selected by SISSO in these mathematical expressions (Table 1) align well with our physical intuition, reinforcing the model's validity and usefulness. For example, in the heterolytic pathways of methane activation on metal M_1 , the descriptors predominantly included properties related to M_1 and the active oxygen—such as the spin density and number of d-electrons on M_1 , the p-band center of oxygen, and the M_1 –O– M_2 bond angle. A noteworthy parameter is the p-band center of the asymmetrical structure of O (ϵ_{PO}) in the 2D and 3D equations for $\Delta G_{\text{heter-M}_1}^{\ddagger}$, exhibiting a negative correlation between ϵ_{PO} and $\Delta G_{\text{heter-M}_1}^{\ddagger}$. This can be rationalized by the existing p-band theory:^{58,59} a more positive ϵ_{PO} can lead to lower occupancy of the antibonding orbitals when binding with adsorbates, thereby potentially lowering the energy barrier due to the Brønsted–Evans–Polanyi (BEP) relation.⁶⁰ Similarly, we observed a negative correlation between the d-band center of M_2 (ϵ_{dM_2}) and $\Delta G_{\text{heter-M}_2}^{\ddagger}$, consistent with the trend suggested by the d-band center theory.^{61,62} Additional insights were gained from observing the influence of the spin density of the fourth Cu atom located further from the active site, as reflected in the 3D expression for $\Delta G_{\text{homo}}^{\ddagger}$. This suggests that reducing the spin density disparity between the Cu and M_1 sites could potentially lower $\Delta G_{\text{homo}}^{\ddagger}$. This provides a strategic direction for enhancing

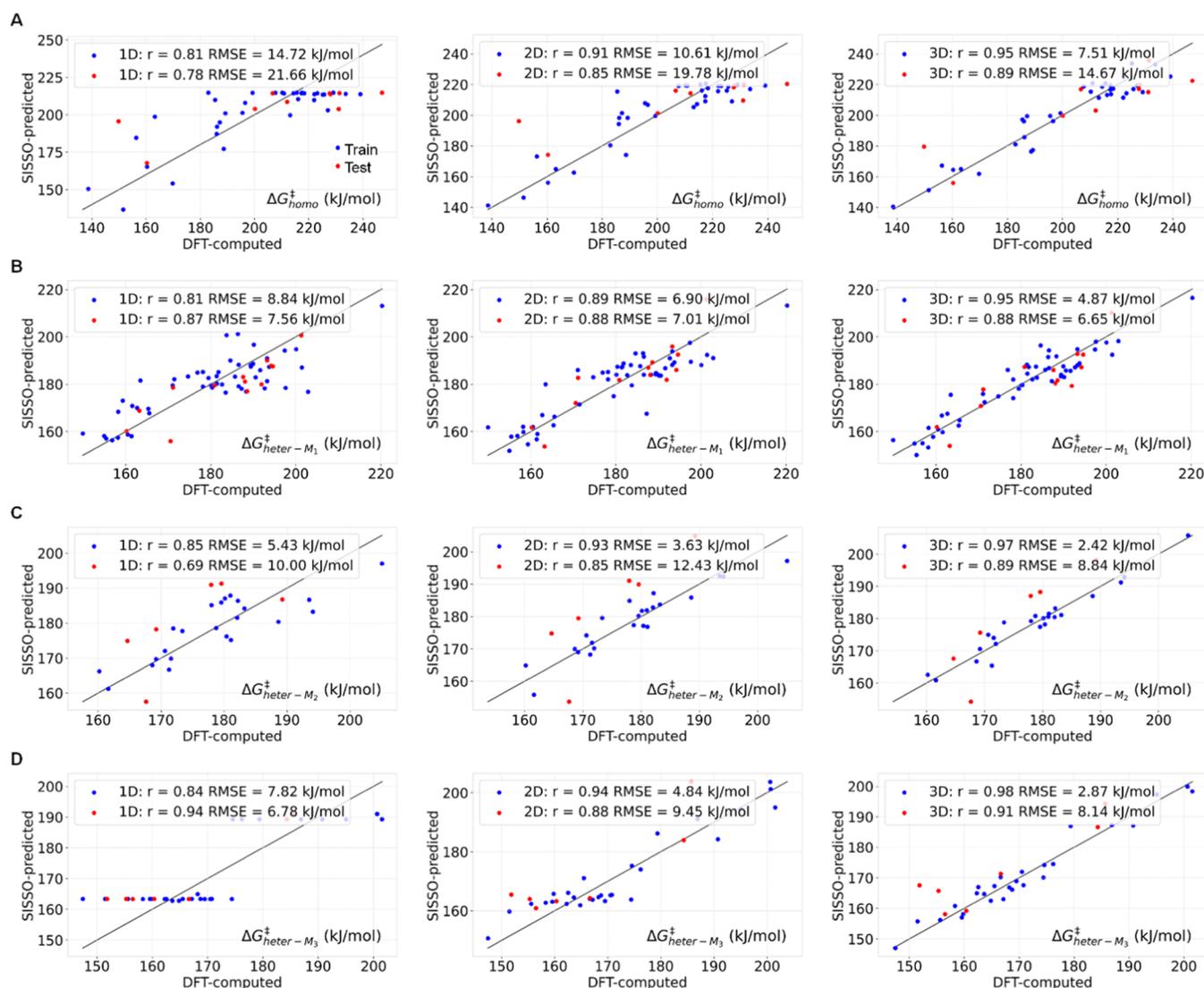


Figure 4. Comparison of DFT-computed and SISSO-predicted (A) $\Delta G_{\text{homo}}^{\ddagger}$, (B) $\Delta G_{\text{heter}-M_1}^{\ddagger}$, (C) $\Delta G_{\text{heter}-M_2}^{\ddagger}$, and (D) $\Delta G_{\text{heter}-M_3}^{\ddagger}$ for the 177 TS structures with minimal structural changes during methane activation. The data sets were divided into an 80% training set (blue) and a 20% test set (red). The left, middle, and right panels of parts A, B, C, and D are derived from the SISSO 1D, 2D, and 3D descriptors, respectively.

the catalytic efficiency of the MeO-NCs. Notably, the top seven species, ranked by $\Delta G_{\text{homo}}^{\ddagger}$, all have Cu as their M_1 site metal, including the best-performing species CuZnNi. Additionally, some species with Zn as the M_1 site (Zn has a spin density similar to that of Cu), such as ZnCuCu, also rank highly when ranked by $\Delta G_{\text{homo}}^{\ddagger}$.

We also implemented another strategy to exclude the impact of structural variability in MeO-NCs on the structure–property relationships. Specifically, we took inspiration from the activation-strain model,^{63–67} decomposing the overall free energy barriers associated with methane activation for all 162 MeO-NCs into two components, as depicted in Figure 5: (1) the free energy required to strain the MeO-NCs (ΔG_{strain}) during methane approach, taking structures post-strain by removing methane from the TS, referred to as IS_{strain} , and (2) the energy needed to activate the C–H bond in methane ($\Delta G_{\text{act}}^{\ddagger}$) on the strained structure (see Tables S7 and S8 for the data), which partially reflects the ability of active oxygen to activate methane and may serve as guidance for material design. We then utilized SISSO to correlate the electronic and

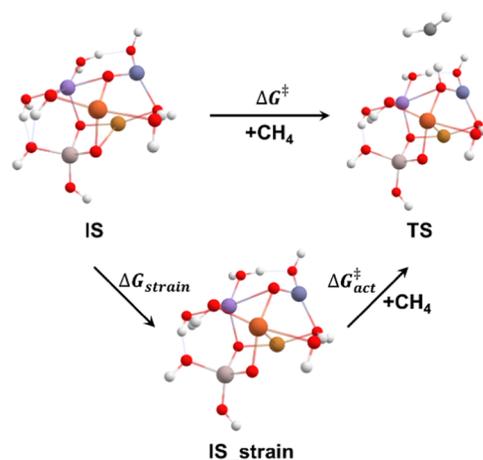


Figure 5. Illustration of contributions from the structural variation of MeO-NCs (ΔG_{strain}) and methane activation ($\Delta G_{\text{act}}^{\ddagger}$) to the total barrier in the methane activation process (ΔG^{\ddagger}).

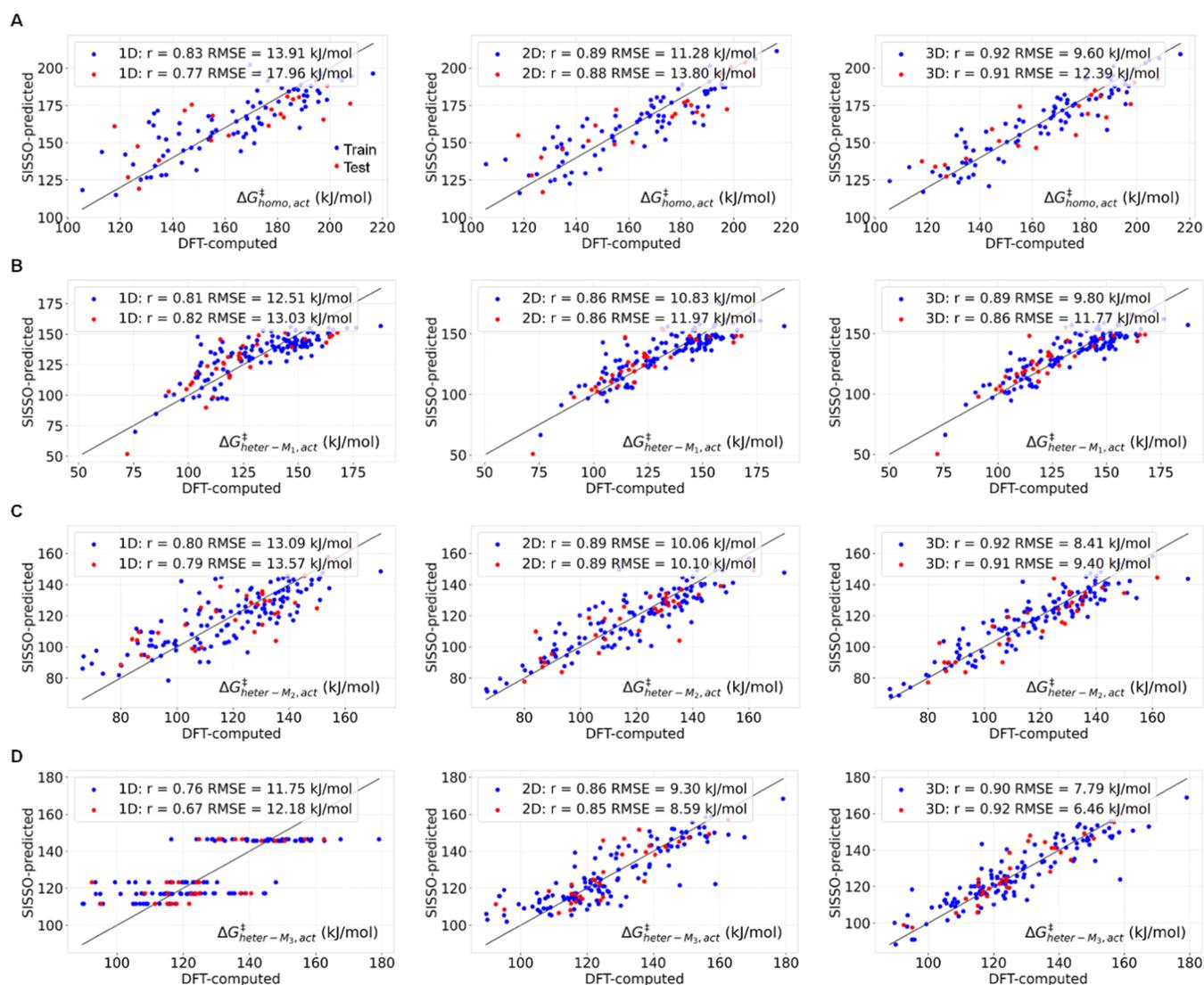


Figure 6. Comparison of DFT-computed and SISSO-predicted (A) $\Delta G_{\text{homo,act}}^{\ddagger}$ (B) $\Delta G_{\text{heter-M}_1,\text{act}}^{\ddagger}$ (C) $\Delta G_{\text{heter-M}_2,\text{act}}^{\ddagger}$ and (D) $\Delta G_{\text{heter-M}_3,\text{act}}^{\ddagger}$ for the 162 MeO-NCs. The data sets were divided into an 80% training set (blue) and a 20% test set (red). The left, middle, and right panels of figures (A–D) are derived from the SISSO 1D, 2D, and 3D descriptors, respectively.

structural parameters of IS_strain with $\Delta G_{\text{act}}^{\ddagger}$ aiming to gain insights into the ease of methane activation relative to the intrinsic properties of the active oxygen site.

The SISSO prediction results, as displayed in Figure 6, demonstrate excellent performance for both the training and test data sets, with acceptable predictive accuracy even for the model with 1D descriptors in predicting $\Delta G_{\text{homo,act}}^{\ddagger}$, $\Delta G_{\text{heter-M}_1,\text{act}}^{\ddagger}$, $\Delta G_{\text{heter-M}_2,\text{act}}^{\ddagger}$ and $\Delta G_{\text{heter-M}_3,\text{act}}^{\ddagger}$ with r values exceeding 0.8 for the training sets and r values of 0.77, 0.82, and 0.79, respectively, for the test sets (left panels). Moreover, models utilizing 2D (middle panel) and 3D descriptors (right panel) exhibit superior predictive performances, with all r values surpassing 0.85 and the RMSEs remaining below 13 kJ/mol for both training and test data sets. These favorable results highlight a strong correlation between the methane activation component ($\Delta G_{\text{act}}^{\ddagger}$) and the nanoclusters' electronic and structural features, as detailed in the mathematical expressions listed in Table 2. A particularly noteworthy finding is the negative correlation between $(N_{\text{d}_{\text{M}_1}} + N_{\text{d}_{\text{M}_2}})$ and $G_{\text{homo,act}}^{\ddagger}$ within the 1D expression of $\Delta G_{\text{homo,act}}^{\ddagger}$ (

$G_{\text{homo,act}}^{\ddagger} = -1.798d_{\text{M}_3-\text{O}}^2 \cdot (N_{\text{d}_{\text{M}_1}} + N_{\text{d}_{\text{M}_2}}) + 304.081$). Since the bond length $d_{\text{M}_3-\text{O}}$ remains relatively constant across all 162 structures (approximately 2.0 to 2.3 Å), $\Delta G_{\text{homo,act}}^{\ddagger}$ can be reduced by increasing $(N_{\text{d}_{\text{M}_1}} + N_{\text{d}_{\text{M}_2}})$, i.e., selecting metals with a higher number of d-electrons at the M_1 and M_2 sites. Similarly, the 1D expression for $\Delta G_{\text{heter-M}_1,\text{act}}^{\ddagger}$ given by $\Delta G_{\text{heter-M}_1,\text{act}}^{\ddagger} = 0.00235d_{\text{M}_1-\text{O}}^6 \cdot q_{\text{O}} \cdot \angle_{\text{M}_1\text{OM}_3} + 173.568$, indicates that decreasing the charge density on oxygen could potentially lower $\Delta G_{\text{heter-M}_1,\text{act}}^{\ddagger}$. Compared to more opaque “black-box-like” ML models, these expressions provide a more interpretable perspective that aids in the metal selection and optimization of the MeO-NCs.

3.3. Unsupervised Machine Learning for the Quick Distinction of Promising MeO-NCs. Above, we have predicted the energy barriers for methane activation directly from easily computable intrinsic material properties for a subset of nanoclusters with minimal structural change during the transition from the initial state to TS. However, it is still

Table 2. SISSO-Suggested Best-Performing Mathematical Expressions Connecting the Electronic and Structural Features of IS_Strain and the Target $\Delta G_{\text{act}}^{\ddagger}$ for 162 MeO-NCs

| target | dimension of descriptor | mathematical expression |
|---|-------------------------|---|
| $\Delta G_{\text{homo,act}}^{\ddagger}$ | 1D | $\Delta G_{\text{homo,act}}^{\ddagger} = -1.798d_{\text{M}_3\text{-O}}^2 \cdot (N_{\text{d}_{\text{M}_1}} + N_{\text{d}_{\text{M}_2}}) + 304.081$ |
| | 2D | $\Delta G_{\text{homo,act}}^{\ddagger} = -0.00718d_{\text{M}_3\text{-O}}^3 \cdot \angle_{\text{M}_1\text{M}_2\text{M}_3} \cdot N_{\text{d}_{\text{M}_1}} - 10.385(\cos(\rho_{\text{M}_2}) - \sin(N_{\text{d}_{\text{M}_3}})) + 231.280$ |
| | 3D | $\Delta G_{\text{homo,act}}^{\ddagger} = -0.0965d_{\text{M}_3\text{-O}}^3 \cdot N_{\text{d}_{\text{M}_1}} \cdot N_{\text{d}_{\text{M}_2}} + 68.776 \frac{N_{\text{d}_{\text{M}_3}}}{\theta_{\text{M}_1\text{M}_2\text{M}_3\text{O}} \cdot \sqrt{E_g}} - 10707.896 \frac{q_{\text{M}_2}}{\theta_{\text{M}_1\text{M}_2\text{M}_3\text{O}} \cdot q_{\text{O}} \cdot N_{\text{d}_{\text{M}_2}}} + 288.882$ |
| $\Delta G_{\text{heter-M}_1\text{,act}}^{\ddagger}$ | 1D | $\Delta G_{\text{heter-M}_1\text{,act}}^{\ddagger} = 0.00235d_{\text{M}_1\text{-O}}^6 \cdot q_{\text{O}} \cdot \angle_{\text{M}_1\text{OM}_3} + 173.568$ |
| | 2D | $\Delta G_{\text{heter-M}_1\text{,act}}^{\ddagger} = -0.118 \frac{d_{\text{M}_1\text{-O}}^3 \cdot \angle_{\text{M}_1\text{OM}_3}}{d_{\text{M}_3\text{-O}}} + 13.994\rho_{\text{Cu}} \cdot \rho_{\text{M}_3} \cdot (\varepsilon_{\text{d}_{\text{Cu}}} - \varepsilon_{\text{d}_{\text{M}_3}}) + 218.908$ |
| | 3D | $\Delta G_{\text{heter-M}_1\text{,act}}^{\ddagger} = -0.120 \frac{d_{\text{M}_1\text{-O}}^3 \cdot \angle_{\text{M}_1\text{OM}_3}}{d_{\text{M}_3\text{-O}}} + 14.011\rho_{\text{Cu}} \cdot \rho_{\text{M}_3} \cdot (\varepsilon_{\text{d}_{\text{Cu}}} - \varepsilon_{\text{d}_{\text{M}_3}}) - 0.0337 \frac{\cos(\angle_{\text{M}_2\text{OM}_3}) \cdot \rho_{\text{M}_1}}{\rho_{\text{M}_2}} + 220.879$ |
| $\Delta G_{\text{heter-M}_2\text{,act}}^{\ddagger}$ | 1D | $\Delta G_{\text{heter-M}_2\text{,act}}^{\ddagger} = 161.148 \frac{\cos(d_{\text{M}_2\text{-O}}) \cdot \angle_{\text{M}_2\text{OM}_3}}{\angle_{\text{M}_1\text{OM}_3}} + 217.535$ |
| | 2D | $\Delta G_{\text{heter-M}_2\text{,act}}^{\ddagger} = 101.405q_{\text{Cu}} \cdot q_{\text{M}_2} \cdot (d_{\text{M}_1\text{-O}} - d_{\text{M}_2\text{-O}}) - 0.0421 \frac{\angle_{\text{M}_1\text{OM}_2} \cdot \theta_{\text{M}_1\text{M}_2\text{M}_3\text{O}} \cdot q_{\text{O}}}{d_{\text{M}_1\text{-O}}} + 213.879$ |
| | 3D | $\Delta G_{\text{heter-M}_2\text{,act}}^{\ddagger} = 55.360(q_{\text{Cu}} + q_{\text{M}_2}) \cdot \cos(d_{\text{M}_2\text{-O}}) - 11.557 \frac{q_{\text{O}} \cdot \theta_{\text{M}_1\text{M}_2\text{M}_3\text{O}}}{d_{\text{M}_1\text{-O}}^3} + 1059.177 \frac{\rho_{\text{M}_2} \cdot \rho_{\text{O}} \cdot E_g}{\angle_{\text{M}_2\text{OM}_3}} + 238.610$ |
| $\Delta G_{\text{heter-M}_3\text{,act}}^{\ddagger}$ | 1D | $\Delta G_{\text{heter-M}_3\text{,act}}^{\ddagger} = -92.743 \cos(\sin(N_{\text{d}_{\text{M}_3}})) + 196.623$ |
| | 2D | $\Delta G_{\text{heter-M}_3\text{,act}}^{\ddagger} = -33.840 \cos(\rho_{\text{M}_3}) - \sin(\rho_{\text{M}_3}) - 9.193 \frac{(q_{\text{M}_2} - q_{\text{O}}) \cdot \varepsilon_{\text{PO}}}{q_{\text{M}_3}} + 109.666$ |
| | 3D | $\Delta G_{\text{heter-M}_3\text{,act}}^{\ddagger} = -32.406 \cos(\rho_{\text{M}_3}) - \sin(\rho_{\text{M}_3}) - 0.00130 \frac{\angle_{\text{M}_1\text{OM}_2} \cdot \angle_{\text{M}_1\text{OM}_3} \cdot \varepsilon_{\text{PO}}}{q_{\text{M}_3}} + 179.667 \frac{\rho_{\text{Cu}}^6}{q_{\text{Cu}} - q_{\text{M}_2}} + 122.949$ |

highly desirable to quickly and effectively distinguish between high- and low-performance nanoclusters while more robustly accounting for structural flexibility. This is especially important when considering the catalytic activity of amorphous nanoclusters under experimental conditions. To this end, we utilized an unsupervised machine learning model to cluster all 162 MeO-NCs, including all transition states, even those that showed significant structural changes. We hypothesized that MeO-NCs with high versus low performance levels might exhibit unique electronic and geometric properties, which can be identified via clustering algorithms. For this task, we employed the BIRCH algorithm,⁴⁵ known for its resilience against noisy and unevenly distributed data. Figure 7 visually represents the results on a 2D map, with axes derived from the input features using PCA.⁴⁴ Notably, the BIRCH algorithm effectively segregated the MeO-NCs into two distinct regions: the high-performing species (surpassing the CuCuCu base material with a $\Delta G_{\text{min}}^{\ddagger}$ of 139 kJ/mol) predominantly in the cyan region, and the low-performing ones ($\Delta G_{\text{min}}^{\ddagger} > 180$ kJ/mol) in the royal blue region. These results indicate that the unsupervised model can distinguish the vast majority of high- and low-performance MeO-NCs, even when including structures that exhibit substantial structural flexibility.

To test the predictive ability of the resultant 2D map, we introduced 32 new structures by substituting the distant (fourth) Cu site with other transition metals, referred to as $\text{M}_1\text{M}_2\text{M}_3\text{-M}_4$. The new data points are plotted in Figure 7 and marked as stars. Encouragingly, the high- and low-performing species, indicated by green and red boxed texts, align well with the cyan and royal blue regions, respectively. This alignment validates the predictive power of our 2D map for efficiently screening out less promising nanoclusters. For comparison, we also employed the Brønsted–Evans–Polanyi (BEP) relation,⁶⁰ using the hydrogen affinity energy ($\Delta G_{\text{H}}^{\ddagger}$) to predict $\Delta G_{\text{min}}^{\ddagger}$ (refer to Figure S14 for more details). However, no clear pattern emerged. This highlights the advantage of unsupervised learning in handling complex, amorphous systems by providing a valuable initial screening tool, significantly reducing the reliance on computationally expensive TS calculations and accelerating the design and discovery of highly active MeO-NCs.

3.4. Limitations and Considerations for Future Research. In this work, we attempted to conduct a systematic investigation of certain aspects of methane activation on MeO-NCs, with a focus on exploring a range of different transition metal compositions. It is important to acknowledge that our calculations rely on certain necessary approximations and

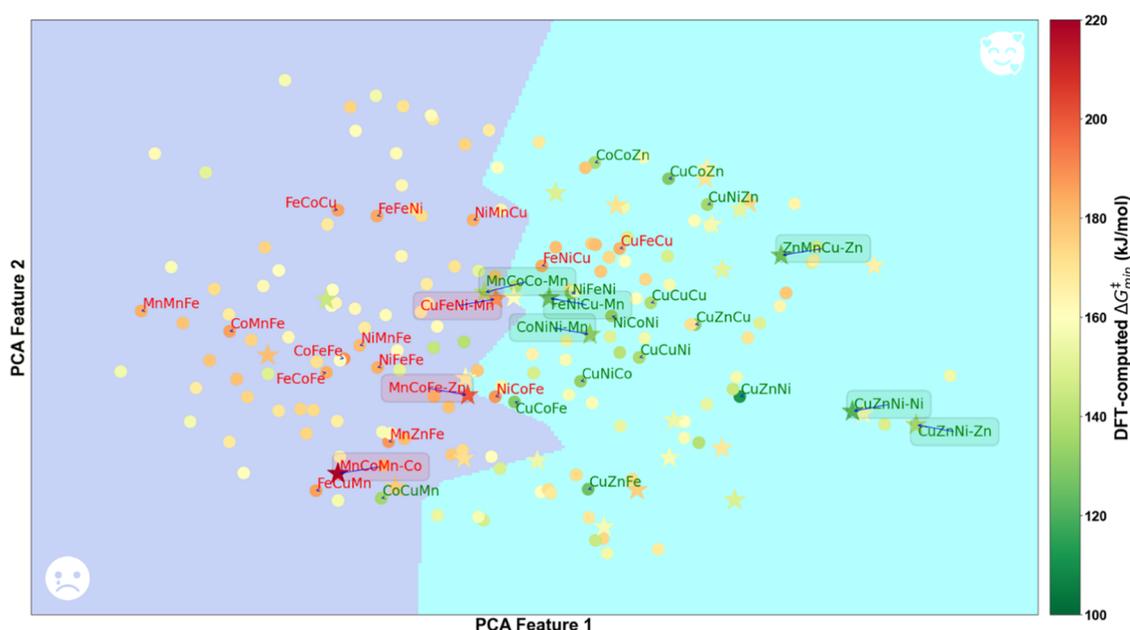


Figure 7. Application of unsupervised machine learning to cluster the MeO-NCs. The horizontal and vertical axes are the dimensionality reduced features using PCA, based on the electronic and geometric features. The color bar represents the $\Delta G_{\min}^{\ddagger}$ values for each MeO-NC. MeO-NCs performing well ($\Delta G_{\min}^{\ddagger} < 139$ kJ/mol) and poorly ($\Delta G_{\min}^{\ddagger} > 180$ kJ/mol) are labeled in green and red texts, respectively. The original 162 data points are represented by circles, and the newly added 32 validation data points (with high- and low-performing species marked in boxed green and red text) are depicted by stars.

assumptions to make this study tractable. First, all of the MeO-NC models examined in this study were built based on the substitution of the CuCuCu cluster from our previous work as a representative, interesting structure. This model simplification, which was made due to computational limits, does not fully account for variations in the coordination environment for each nanocluster, such as the number of surrounding H₂O and OH groups, which may vary depending on the ALD synthesis conditions and the experimental reaction conditions. However, it does reveal insights into how the composition affects methane activation. Future research could benefit from a more comprehensive examination using techniques like ab initio thermodynamics and AIMD free energy calculations, which allow for the inclusion of varying numbers of water and OH groups,⁶⁸ or more sophisticated global optimization methods, such as Basin Hopping (BH)⁶⁹ and grand canonical genetic algorithm (GCGA).⁷⁰

Additionally, although our analyses presumed a ferromagnetic state, which is the most stable spin state for CuCuCu, we recognize that other spin states may exist in different nanoclusters. The presence of multiple ferro/antiferromagnetic interactions could influence their electronic energy and structural configurations. For example, elements like Fe or Mn might exhibit preferences for different geometries, such as near-trigonal pyramidal versus near-tetrahedral or near-square planar configurations upon decoordination of a neutral H₂O.^{71,72} This diversity opens pathways for future studies to explore various spin states and their effects on structural configurations and reactivities. Investigating these aspects may require more complex assessments beyond standard DFT calculations, such as high-level multireference methods.⁷³

Moreover, this work focused only on the methane activation step of the catalytic cycle. While this is a critical and challenging step in methane utilization, the development of improved catalyst materials requires investigation of the full

catalytic cycle, including the formation of the metal-oxo site, in our case, the μ 3-O site, particularly from a kinetic perspective. Although our calculations on the base material CuCuCu and two high-performing nanoclusters, CuZnFe and CuZnNi, showed highly exothermic μ 3-O formation with no apparent energy barrier (Figure S15), it is possible that other nanoclusters may exhibit significant kinetic barriers for μ 3-O formation. Furthermore, factors such as product (or intermediate) desorption, side reactions, and cluster reoxidation could play essential roles in determining the overall catalytic performance. The various limitations discussed here highlight the challenges in modeling amorphous catalytic materials as well as nanoclusters containing multiple transition metal atoms and may point the way toward potential enhancements in future studies.

4. CONCLUSIONS

An integration of DFT calculations with advanced machine learning techniques has led to significant insights into a family of MeO-NCs as catalysts for efficient methane activation. Inspired by a previously identified tetra-copper oxide nanocluster,²⁸ our systematic substitution strategy uncovered 12 novel compositions that are predicted to have enhanced catalytic activity. We also analyzed the electronic structures of these systems to enhance our understanding of the competition between homolytic and heterolytic methane activation mechanisms. Further, we leveraged sophisticated machine learning algorithms to refine our understanding of complex structure–reactivity relationships, resulting in mathematical models that bridge electronic and structural features with the catalytic activity at the active oxygen center. Finally, we developed an unsupervised clustering model capable of distinguishing high-performing nanoclusters from their less effective counterparts while accommodating the inherent flexibility of peripheral atoms in amorphous structures.

Looking forward, this work focusing on C–H bond activation and the complexities of amorphous nanoclusters opens up new avenues for understanding and manipulating these challenging systems. The detailed examination of the relationship between the intrinsic properties of tricoordinated oxygen sites and adjacent metal sites, and their catalytic activity toward methane activation, lays a promising foundation for exploring larger nanoclusters with comparable active oxygen sites. For future exploration of new tetrametal nanoclusters for methane activation, the SISSO-derived equations and unsupervised clustering methods are readily applicable for material screening. Given a number of new nanoclusters, we recommend first optimizing their geometry, followed by single-point calculations using DFT to obtain the intrinsic features as detailed in Section 3.2. Subsequently, since the BIRCH model is more robust and demonstrated to work well even for nanoclusters showing changes in geometry between the reactant and transition state, we recommend feeding these features into the BIRCH model (refer to our GitHub repository⁷⁴ for the code and instructions) for preliminary screening. For species falling into high-performance regions, we can then input the features into our SISSO-derived equations to quickly estimate the methane activation energy barrier, screening the most promising candidate materials for more accurate DFT studies. This approach can bypass the expensive TS calculations for the most poorly performing nanoclusters.

The integration of computational chemistry and machine learning methods, as showcased in our research, is poised to play a crucial role in accelerating the pace of material discovery and can be readily adapted to other catalytic systems, potentially revolutionizing the way we approach catalyst design and optimization in various chemical transformations.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.4c04021>.

Introduction of base copper oxide nanocluster; spin density distribution for CuCuCu; electronic and free energy barrier plots; frontier molecular orbitals for CuZnNi and CuZnFe; ETR results for 162 MeO-NCs; SISSO results for 162 MeO-NCs and for the nanoclusters with significant variations during TS formation; relative free energies for each MeO-NC at relevant spin states; energy barrier values for each MeO-NC; and strain and activation components of free energy barriers for each MeO-NC (PDF)

DFT-optimized key structure coordinates in xyz format (Str.zip) (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Randall Q. Snurr – Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States; orcid.org/0000-0003-2925-9246; Email: snurr@northwestern.edu

Authors

Xijun Wang – Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States; orcid.org/0000-0001-9155-7653

Kaihang Shi – Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States; Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260, United States; orcid.org/0000-0002-0297-1746

Anyang Peng – Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States; orcid.org/0000-0002-0630-2187

Complete contact information is available at: <https://pubs.acs.org/10.1021/acscatal.4c04021>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Institute for Catalysis in Energy Processes (ICEP) via the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DOE DE-FG02-03ER15457. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP0023154. This research was also supported in part through the computational resources and staff contributions provided for the Quest high performance computing facility at Northwestern University which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. This work was also supported by the San Diego Supercomputer Center at University of California San Diego through allocation no. CHM230037 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program,⁷⁵ which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603 and #2138296.

■ REFERENCES

- (1) C. World Energy Council. *World Energy Resources 2013 Survey: Summary*; World Energy Council: London, UK, 2013; p 29.
- (2) Sushkevich, V. L.; Palagin, D.; Ranocchiaro, M.; Van Bokhoven, J. A. Selective anaerobic oxidation of methane enables direct synthesis of methanol. *Science* **2017**, *356* (6337), 523–527.
- (3) Ogura, K.; Takamagari, K. Direct conversion of methane to methanol, chloromethane and dichloromethane at room temperature. *Nature* **1986**, *319* (6051), No. 308.
- (4) Wang, S.; Fung, V.; Hülsey, M. J.; Liang, X.; Yu, Z.; Chang, J.; Folli, A.; Lewis, R. J.; Hutchings, G. J.; He, Q.; Yan, N. H₂-reduced phosphomolybdate promotes room-temperature aerobic oxidation of methane to methanol. *Nat. Catal.* **2023**, *6* (10), 895–905.
- (5) Wen, W.; Che, J.-W.; Wu, J.-M.; Kobayashi, H.; Pan, Y.; Wen, W.; Dai, Y.-H.; Huang, W.; Fu, C.; Zhou, Q.; Lu, G.-L.; Tian, H.; Liu, J.; Yang, P.; Chen, X.; Sun, T.-L.; Fan, J. Co³⁺–O bond elongation unlocks Co₃O₄ for methane activation under ambient conditions. *ACS Catal.* **2022**, *12* (12), 7037–7045.
- (6) Zheng, J.; Ye, J.; Ortuño, M. A.; Fulton, J. L.; Gutiérrez, O. Y.; Camaioni, D. M.; Motkuri, R. K.; Li, Z.; Webber, T. E.; Mehdi, B. L.; Browning, N. D.; Penn, R. L.; Farha, O. K.; Hupp, J. T.; Truhlar, D. G.; Cramer, C. J.; Lercher, J. A. Selective methane oxidation to methanol on Cu-oxo dimers stabilized by zirconia nodes of an NU-1000 metal–organic framework. *J. Am. Chem. Soc.* **2019**, *141* (23), 9292–9304.
- (7) Ikuno, T.; Zheng, J.; Vjunov, A.; Sanchez-Sanchez, M.; Ortuño, M. A.; Pahls, D. R.; Fulton, J. L.; Camaioni, D. M.; Li, Z.; Ray, D.;

- Mehdi, B. L.; Browning, N. D.; Farha, O. K.; Hupp, K. T.; Cramer, C. J.; Gagliardi, L.; Lercher, J. A. Methane oxidation to methanol catalyzed by Cu-oxo clusters stabilized in NU-1000 metal–organic framework. *J. Am. Chem. Soc.* **2017**, *139* (30), 10294–10301.
- (8) Lee, I.; Lee, M.-S.; Tao, L.; Ikuno, T.; Khare, R.; Jentys, A.; Huthwelker, T.; Borca, C. N.; Kalinko, A.; Gutiérrez, O. Y.; Govind, N.; Fulton, J. L.; Hu, J. Z.; Glezakou, V.-A.; Rousseau, R.; Sanchez-Sanchez, M.; Lercher, J. A. Activity of Cu–Al–Oxo extra-framework clusters for selective methane oxidation on Cu-exchanged zeolites. *JACS Au* **2021**, *1* (9), 1412–1421.
- (9) Zheng, J.; Lee, I.; Khramenkova, E.; Wang, M.; Peng, B.; Gutiérrez, O. Y.; Fulton, J. L.; Camaioni, D. M.; Khare, R.; Jentys, A.; Haller, G. L.; Pidko, E. A.; Sanchez-Sanchez, M.; Lercher, J. A. Importance of methane chemical potential for its conversion to methanol on Cu-exchanged mordenite. *Chem. - Eur. J.* **2020**, *26* (34), 7563–7567.
- (10) Ikuno, T.; Grundner, S.; Jentys, A.; Li, G.; Pidko, E.; Fulton, J.; Sanchez-Sanchez, M.; Lercher, J. A. Formation of active Cu-oxo clusters for methane oxidation in Cu-exchanged mordenite. *J. Phys. Chem. C* **2019**, *123* (14), 8759–8769.
- (11) Dinh, K. T.; Sullivan, M. M.; Narsimhan, K.; Serna, P.; Meyer, R. J.; Dinca, M.; Román-Leshkov, Y. Continuous partial oxidation of methane to methanol catalyzed by diffusion-paired copper dimers in copper-exchanged zeolites. *J. Am. Chem. Soc.* **2019**, *141* (29), 11641–11650.
- (12) Mahyuddin, M. H.; Tanaka, T.; Shiota, Y.; Staykov, A.; Yoshizawa, K. Methane partial oxidation over [Cu₂(μ-O)]²⁺ and [Cu₃(μ-O)₃]²⁺ active species in large-pore zeolites. *ACS Catal.* **2018**, *8* (2), 1500–1509.
- (13) Zhao, Z. J.; Kulkarni, A.; Vilella, L.; Nørskov, J. K.; Studt, F. Theoretical insights into the selective oxidation of methane to methanol in copper-exchanged mordenite. *ACS Catal.* **2016**, *6* (6), 3760–3766.
- (14) Yu, X.; Zhong, L.; Li, S. Catalytic cycle of the partial oxidation of methane to methanol over Cu-ZSM-5 revealed using DFT calculations. *Phys. Chem. Chem. Phys.* **2021**, *23* (8), 4963–4974.
- (15) Tao, L.; Khramenkova, E.; Lee, I.; Ikuno, T.; Khare, R.; Jentys, A.; Fulton, J. L.; Kolganov, A. A.; Pidko, E. A.; Sanchez-Sanchez, M.; Lercher, J. A. Speciation and reactivity control of Cu-Oxo clusters via extraframework Al in mordenite for methane oxidation. *J. Am. Chem. Soc.* **2023**, *145* (32), 17710–17719.
- (16) Zhao, Z.-J.; Kulkarni, A.; Vilella, L.; Nørskov, J. K.; Studt, F. Theoretical insights into the selective oxidation of methane to methanol in copper-exchanged mordenite. *ACS Catal.* **2016**, *6* (6), 3760–3766.
- (17) Mahyuddin, M. H.; Staykov, A.; Shiota, Y.; Miyayoshi, M.; Yoshizawa, K. Roles of zeolite confinement and Cu–O–Cu angle on the direct conversion of methane to methanol by [Cu₂(μ-O)]²⁺-exchanged AEI, CHA, AFX, and MFI zeolites. *ACS Catal.* **2017**, *7* (6), 3741–3751.
- (18) Falcaro, P.; Ricco, R.; Yazdi, A.; Imaz, I.; Furukawa, S.; Maspooh, D.; Ameloot, R.; Evans, J. D.; Doonan, C. J. Application of metal and metal oxide nanoparticles@ MOFs. *Coord. Chem. Rev.* **2016**, *307*, 237–254.
- (19) Farooq, U.; Ahmad, T.; Naaz, F.; ul Islam, S. Review on metals and metal oxides in sustainable energy production: progress and perspectives. *Energy Fuels* **2023**, *37* (3), 1577–1632.
- (20) Matus, M. F.; Häkkinen, H. Understanding ligand-protected noble metal nanoclusters at work. *Nat. Rev. Mater.* **2023**, *8*, 372–389.
- (21) Formalik, F.; Shi, K.; Joodaki, F.; Wang, X.; Snurr, R. Q. Exploring the structural, dynamic, and functional properties of metal–organic frameworks through molecular modeling. *Adv. Funct. Mater.* **2023**, No. 2308130.
- (22) Wang, X.; Jiang, S.; Hu, W.; Ye, S.; Wang, T.; Wu, F.; Yang, L.; Li, X.; Zhang, G.; Chen, X.; Jiang, J.; Luo, Y. Quantitatively determining surface–adsorbate properties from vibrational spectroscopy with interpretable machine learning. *J. Am. Chem. Soc.* **2022**, *144* (35), 16069–16076.
- (23) Mou, T.; Pillai, H. S.; Wang, S.; Wan, M.; Han, X.; Schweitzer, N. M.; Che, F.; Xin, H. Bridging the complexity gap in computational heterogeneous catalysis with machine learning. *Nat. Catal.* **2023**, *6* (2), 122–136.
- (24) Zhang, S.; Lu, S.; Zhang, P.; Tian, J.; Shi, L.; Ling, C.; Zhou, Q.; Wang, J. Accelerated discovery of single-atom catalysts for nitrogen fixation via machine learning. *Energy Environ. Mater.* **2023**, *6* (1), No. e12304.
- (25) Li, H.; Jiao, Y.; Davey, K.; Qiao, S.-Z. Data-driven machine learning for understanding surface structures of heterogeneous catalysts. *Angew. Chem., Int. Ed.* **2023**, *135* (9), No. e202216383.
- (26) Doan, H. A.; Li, Z.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q. Theoretical insights into direct methane to methanol conversion over supported dicopper oxo nanoclusters. *Catal. Today* **2018**, *312*, 2–9.
- (27) Doan, H. A.; Wang, X.; Snurr, R. Q. Computational screening of supported metal oxide nanoclusters for methane activation: Insights into homolytic versus heterolytic C–H bond dissociation. *J. Phys. Chem. Lett.* **2023**, *14*, 5018–5024.
- (28) Wang, X.; Shi, K.; Peng, A.; Snurr, R. Q. Probing the structure–property relationships of supported copper oxide nanoclusters for methane activation. *EES Catal.* **2024**, *2*, 351–364.
- (29) Avila, J. R.; Emery, J. D.; Pellin, M. J.; Martinson, A. B. F.; Farha, O. K.; Hupp, J. T. Porphyrins as templates for site-selective atomic layer deposition: vapor metalation and in situ monitoring of island growth. *ACS Appl. Mater. Interfaces.* **2016**, *8* (31), 19853–19859.
- (30) Gaussian 16, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*; Gaussian, Inc.: Wallingford CT, 2016.
- (31) Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125* (19), No. 194101.
- (32) 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. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305.
- (33) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. auxiliary basis sets to approximate coulomb potentials. *Chem. Phys. Lett.* **1995**, *240* (4), 283–290.
- (34) Barona, M.; Gaggioli, C. A.; Gagliardi, L.; Snurr, R. Q. DFT study on the catalytic activity of ALD-grown diiron oxide nanoclusters for partial oxidation of methane to methanol. *J. Phys. Chem. A* **2020**, *124* (8), 1580–1592.
- (35) 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), No. 154104.
- (36) Batsaikhan, E.; Lee, C. H.; Hsu, H.; Wu, C. M.; Peng, J. C.; Ma, M. H.; Deleg, S.; Li, W. H. Largely enhanced ferromagnetism in Bare CuO nanoparticles by a small size effect. *ACS Omega* **2020**, *5* (8), 3849–3856.
- (37) Dolai, S.; Dey, R.; Sarangi, S. N.; Hussain, S.; Bhar, R.; Pal, A. K. Ferromagnetic properties in CuO-nanocrystals embedded in PVDF matrix. *J. Magn. Mater.* **2020**, *495*, No. 165903.

- (38) Rehman, S.; Mumtaz, A.; Hasanain, S. K. Size effects on the magnetic and optical properties of CuO nanoparticles. *J. Nanopart. Res.* **2011**, *13*, 2497–2507.
- (39) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of solution-phase vibrational frequencies in continuum models for the free energy of solvation. *J. Phys. Chem. B* **2011**, *115* (49), 14556–14562.
- (40) Ertem, M. Z.; Gagliardi, L.; Cramer, C. J. Quantum chemical characterization of the mechanism of an iron-based water oxidation catalyst. *Chem. Sci.* **2012**, *3* (4), 1293–1299.
- (41) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, *2* (8), No. 083802.
- (42) Sun, S.; Ouyang, R.; Zhang, B.; Zhang, T. Y. Data-driven discovery of formulas by symbolic regression. *MRS Bull.* **2019**, *44* (7), 559–564.
- (43) Xie, S. R.; Kotlarz, P.; Hennig, R. G.; Nino, J. C. Machine learning of octahedral tilting in oxide perovskites by symbolic classification with compressed sensing. *Comput. Mater. Sci.* **2020**, *180*, No. 109690.
- (44) Labrín, C.; Urdinez, F. Principal Component Analysis. In *R for Political Data Science*; Chapman and Hall/CRC, 2020; pp 375–393.
- (45) Zhang, T.; Ramakrishnan, R.; Livny, M. BIRCH: an efficient data clustering method for very large databases. *ACM SIGMOD Record* **1996**, *25* (2), 103–114.
- (46) Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J. API Design for Machine Learning Software: Experiences from the Scikit-learn Project. 2013, arXiv:1309.0238v1. arXiv.org e-Print archive. <https://doi.org/10.48550/arXiv.1309.0238>.
- (47) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (48) Schwarz, H.; Shaik, S.; Li, J. Electronic effects on room-temperature, gas-phase C–H bond activations by cluster oxides and metal carbides: the methane challenge. *J. Am. Chem. Soc.* **2017**, *139* (48), 17201–17212.
- (49) Gaggioli, C. A.; Sauer, J.; Gagliardi, L. Hydrogen atom or proton coupled electron transfer? C–H bond activation by transition-metal oxides. *J. Am. Chem. Soc.* **2019**, *141* (37), 14603–14611.
- (50) Geng, C.; Li, J.; Weiske, T.; Schlangen, M.; Shaik, S.; Schwarz, H. Electrostatic and charge-induced methane activation by a concerted double C–H bond insertion. *J. Am. Chem. Soc.* **2017**, *139* (4), 1684–1689.
- (51) Nandy, A.; Adamji, H.; Kastner, D. W.; Vennelakanti, V.; Nazemi, A.; Liu, M.; Kulik, H. J. Using computational chemistry to reveal nature's blueprints for single-site catalysis of C–H activation. *ACS Catal.* **2022**, *12* (15), 9281–9306.
- (52) Kumar, M.; Dar, M. A.; Katiyar, A.; Agrawal, R.; Shenai, P. M.; Srinivasan, V. Role of magnetization on catalytic pathways of non-oxidative methane activation on neutral iron carbide clusters. *Phys. Chem. Chem. Phys.* **2022**, *24* (19), 11668–11679.
- (53) Fujisaki, H.; Ishizuka, T.; Kotani, H.; Shiota, Y.; Yoshizawa, K.; Kojima, T. Selective methane oxidation by molecular iron catalysts in aqueous medium. *Nature* **2023**, *616* (7957), 476–481.
- (54) Sun, L.; Wang, Y.; Guan, N.; Li, L. Methane activation and utilization: current status and future challenges. *Energy Technol.* **2020**, *8* (8), No. 1900826.
- (55) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42.
- (56) Shu, W.; Li, J.; Liu, J. X.; Zhu, C.; Wang, T.; Feng, L.; Ouyang, R.; Li, W. X. Structure Sensitivity of Metal Catalysts Revealed by Interpretable Machine Learning and First-Principles Calculations. *J. Am. Chem. Soc.* **2024**, *146* (12), 8737–8745.
- (57) Chen, F.; Guo, Z.; Wang, J.; Ouyang, R.; Ma, D.; Gao, P.; Pan, F.; Ding, P. Accelerated feasible screening of flame-retardant polymeric composites using data-driven multi-objective optimization. *Comput. Mater. Sci.* **2023**, *230*, No. 112479.
- (58) Meng, C.; Gao, Y.; Zhou, Y.; Sun, K.; Wang, Y.; Han, Y.; Zhao, Q.; Chen, X.; Hu, H.; Wu, M. P-band center theory guided activation of MoS₂ basal S sites for pH-universal hydrogen evolution. *Nano Res.* **2023**, *16*, 6228–6236.
- (59) Wang, X.; Krzystowczyk, E.; Dou, J.; Li, F. Net electronic charge as an effective electronic descriptor for oxygen release and transport properties of SrFeO₃-based oxygen sorbents. *Chem. Mater.* **2021**, *33*, 2446–2456.
- (60) Bligaard, T.; Nørskov, J. K.; Dahl, S.; Matthiesen, J.; Christensen, C. H.; Sehested, J. The Bronsted–Evans–Polanyi relation and the volcano curve in heterogeneous catalysis. *J. Catal.* **2004**, *224*, 206–217.
- (61) Hammer, B.; Nørskov, J. K. Why gold is the noblest of all the metals. *Nature* **1995**, *376*, 238–240.
- (62) Hammer, B.; Nørskov, J. K. Electronic factors determining the reactivity of metal surfaces. *Surf. Sci.* **1995**, *343*, 211–220.
- (63) Fernández, I. Combined activation strain model and energy decomposition analysis methods: a new way to understand pericyclic reactions. *Phys. Chem. Chem. Phys.* **2014**, *16*, 7662–7671.
- (64) Bickelhaupt, F. M. Understanding reactivity with Kohn–Sham molecular orbital theory: E2–SN₂ mechanistic spectrum and other concepts. *J. Comput. Chem.* **1999**, *20*, 114–128.
- (65) Yue, S.; Nandy, A.; Kulik, H. J. Discovering Molecular Coordination Environment Trends for Selective Ion Binding to Molecular Complexes Using Machine Learning. *J. Phys. Chem. B* **2023**, *127*, 10592–10600.
- (66) Gani, T. Z. H.; Kulik, H. J. Unifying exchange sensitivity in transition-metal spin-state ordering and catalysis through bond valence metrics. *J. Chem. Theory Comput.* **2017**, *13*, 5443–5457.
- (67) Kwon, S.; Liao, P.; Stair, P. C.; Snurr, R. Q. Alkaline-earth metal-oxide overlayers on TiO₂: application toward CO₂ photo-reduction. *Catal. Sci. Technol.* **2016**, *6*, 7885–7895.
- (68) Chen, T. H.; Vlachos, D. G.; Caratzoulas, S. Ab Initio Molecular Dynamics Study of Pt Clustering on γ -Al₂O₃ and Sn-Modified γ -Al₂O₃. *J. Phys. Chem. C* **2023**, *127* (39), 19778–19787.
- (69) Kumari, S.; Masubuchi, T.; White, H. S.; Alexandrova, A.; Anderson, S. L.; Sautet, P. Electrocatalytic hydrogen evolution at full atomic utilization over ITO-supported sub-nano-Pt n clusters: High, size-dependent activity controlled by fluxional Pt hydride species. *J. Am. Chem. Soc.* **2023**, *145* (10), 5834–5845.
- (70) Zhang, Z.; Masubuchi, T.; Sautet, P.; Anderson, S. L.; Alexandrova, A. N. Hydrogen Evolution on Electrode-Supported Ptn Clusters: Ensemble of Hydride States Governs the Size Dependent Reactivity. *Angew. Chem.* **2023**, *135* (20), No. e202218210.
- (71) Pantazis, D. A.; Krewald, V.; Orto, M.; Neese, F. Theoretical magnetochemistry of dinuclear manganese complexes: broken symmetry density functional theory investigation on the influence of bridging motifs on structure and magnetism. *Dalton Trans.* **2010**, *39* (20), 4959–4967.
- (72) Krewald, V.; Neese, F.; Pantazis, D. A. On the magnetic and spectroscopic properties of high-valent Mn₃CaO₄ cubanes as structural units of natural and artificial water-oxidizing catalysts. *J. Am. Chem. Soc.* **2013**, *135* (15), 5726–5739.
- (73) Dawes, R.; Ndengué, S. A. Single- and multireference electronic structure calculations for constructing potential energy surfaces. *Int. Rev. Phys. Chem.* **2016**, *35* (3), 441–478.
- (74) https://github.com/xwsci/BIRCH_plot_Gmin/ (accessed October 14, 2024).
- (75) Boerner, T. J.; Deems, S.; Furlani, T. R.; Knuth, S. L.; Towns, J. In *Access: Advancing Innovation: Nsf's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support*, Practice and Experience in Advanced Research Computing; Association for Computing Machinery: Portland, OR, USA, 2023; pp 173–176.