

Teaching a Transformer to Think Like a Chemist: Predicting Nanocluster Stability

João Marcos T. Palheta¹, Octavio Rodrigues Filho¹, Mohammad Soleymanibrojeni², Alexandre Cavalheiro Dias³, Diego Guedes-Sobrinho⁴, Wolfgang Wenzel⁵, Roland Aydin², Celso R. C. Rêgo^{*5}, and Maurício Jeomar Piotrowski¹

¹Department of Physics, Federal University of Pelotas, 96010-900 Pelotas, Rio Grande do Sul, Brazil

²Hamburg University of Technology, 21073 Hamburg, Germany

³Institute of Physics and International Center of Physics, University of Brasília, 70919-970 Brasília, Distrito Federal, Brazil

⁴Quantum Chemistry and Thermodynamic Materials Group – Q²M, Department of Chemistry, Federal University of Paraná, 81531-980 Curitiba, Paraná, Brazil

⁵Institute of Nanotechnology Hermann-von-Helmholtz-Platz, Karlsruhe Institute of Technology, Karlsruhe, Baden-Württemberg, Germany

Atomically precise metal nanoclusters bridge the molecular and bulk regimes, but designing bimetallic motifs with targeted stability and reactivity remains challenging. Here we combine density functional theory (DFT) and physics-grounded predictive artificial intelligence to map the configurational landscape of 13-atom icosahedral nanoclusters $X_{12}TM$, with hosts $X=(Ti\ Zr\ Hf)$, and Fe and a single transition-metal dopant spanning the $3d-5d$ series. Spin-polarized DFT calculations on 240 bimetallic clusters reveal systematic trends in binding and formation energies, distortion penalties, effective coordination number, d-band centre, and HOMO–LUMO gap that govern the competition between core–shell (in) and surface-segregated (out) arrangements. We then pretrain a transformer architecture on a curated set of 2968 unary clusters from the Quantum Cluster Database and fine-tune it on bimetallic data to predict formation energies and in/out preference, achieving mean absolute errors of about 0.6–0.7 eV and calibrated uncertainty intervals. The resulting model rapidly adapts to an unseen Fe-host domain with only a handful of labelled examples. At the same time, attention patterns and Shapley attributions highlight size mismatch, d -electron count, and coordination environment as key descriptors. All data, code, and workflows follow FAIR/TRUE principles, enabling reproducible, interpretable screening of unexplored nanocluster chemistries for catalysis and energy conversion.

Keywords: 13-atoms, Nanoclusters, Stability, Transformer, AI

1 Introduction

Atomically precise metal nanoclusters connect the molecular and bulk worlds, especially those with a core-shell structure where one metal forms the core and another forms the shell.^[1–5] These clusters have become important

*Corresponding author: celso.rego@kit.edu

in catalysis and materials science because their properties can be adjusted, and their surfaces are highly reactive. Transition-metal (TM) clusters exhibit quantum confinement, *d*-band effects, and strong surface effects, making their electronic structures highly sensitive to size and shape.^[6,7] As a result, both single-metal and bimetallic clusters can display unique energy levels, unexpected magnetism, improved catalytic activity, and unusual optical properties. For example, small icosahedral clusters like Au₁₃ and Ag₁₃ can have magnetic moments that do not appear in their bulk forms.^[8,9]

When nanoclusters contain so-called “magic” numbers of atoms (like 13 or 55), they often form stable, closed-shell structures such as the icosahedron (ICO).^[10,11] This stability comes from their geometry, electronic shell closure, and efficient atomic packing. In addition to their robust structure, these highly symmetric clusters can be tailored to specific electronic and magnetic properties, making them promising for catalysis, sensing, and energy conversion applications.^[1,12–14] The possibilities expand even further with bimetallic nanoclusters, such as nano-alloys and core-shell structures, which offer an even wider range of design options. When combined with different metal species, they leverage the metals’ interactions to influence adsorption and electronic properties, potentially leading to improved stability, reactivity, and selectivity.^[1,3,15,16] Beyond that, hetero-metallic clusters can have better performance in CO oxidation and hydrogen-evolution reactions through cooperative effects.^[17–19]

Predictive artificial intelligence (PAI) now provides a systematic way to speed up discoveries in this area. High-throughput density functional theory (DFT), an approach that uses DFT calculations to predict the properties of a large number of materials in a short time, combined with algorithmic exploration, has mapped low-energy nanocluster structures at scale, as shown by the Quantum Cluster Database (QCD).^[20,21] Building on this, transformer models have shown they can combine different types of structural and electronic information using self-attention, leading to reliable and interpretable predictions of material properties.^[22]

In this paper, as illustrated in Fig. 1, we focus on 13-atom bimetallic icosahedral clusters composed of 12 early-row group-IV hosts (Ti, Zr, Hf) using a single dopant drawn from the 3*d*, 4*d*, and 5*d* series. Beyond that, we analyze core-shell (*in*) and surface-segregated (*out*) configurations using DFT to provide structural, vibrational, energetic, and electronic characterizations. To generalize these insights, we develop a PAI framework centered on a transformer architecture trained to internalize the physics and chemistry of pure metallic clusters using a curated subset of low-energy structures from QCD.^[20] After that, we fine-tuned this pretrained model on a small, high-quality DFT dataset of bimetallic clusters, enabling accurate predictions of binding and formation energies and the *in/out* configuration preference. The attention structure affords interpretability, exposing which atomic environments, shells, and electronic descriptors drive stability trends, while offering design rules aligned with known *d*-band and size/geometry effects.

Our workflow follows the FAIR (Findable, Accessible, Interoperable, Reusable) and TRUE (Transparent, Reproducible, Usable by Others, and Extensible) data principles.^[23–25] Following these ideas, we make sure that our data and methods are open, reproducible, and easy for others to use or build on. We share the workflow in a public repository so anyone can find the inputs, models, and outputs and work with the latest version. We also keep track of data sources and uncertainties and provide clear instructions for reuse.^[26–28] This setup offers a flexible, reproducible framework that combines advanced DFT methods with interpretable transformers. It can transfer knowledge from pre-trained to fine-tuned models, needing less data than a fully data-driven method. This approach allows us to better understand how stability, distortion, and reactivity interact in bimetallic nanoclusters, helping us design new catalysts and materials for energy conversion.

2 Results and Discussions

2.1 DFT-based trends

Structural stability: The vibrational frequency analysis, as shown in Supporting Information (Supporting Information Fig. S10), shows that all optimized X₁₂TM^{*in*} and X₁₂TM^{*out*} nanoclusters are stable, as they are local minima on the potential energy surface with only real, positive frequencies. This confirms their mechanical stability. Interestingly, the *out* configurations often have slightly higher vibrational frequencies for specific modes. This happens because these structures are more geometrically distorted than the more symmetric *in* (core-shell) arrangements, leading to stiffer local bonding in some less symmetric cases.

The computed binding energies shown in Fig. 2 confirm that all binary nanoclusters, regardless of the TM position (*in* or *out*), are thermodynamically stable, exhibiting negative values. This trend is consistent with the strong metallic bonding characteristic of group IV metals (Ti, Zr, Hf) and their interactions with other

TM.^[29] As shown, we observe a negative parabolic trend in binding energy modulus as a function of the TM atomic number, consistent with the classical *d*-band filling model, i.e., a behavior analogous to that observed in bulk TM systems and small metallic clusters.^[29–31] This arises from the systematic filling of the *d*-orbitals across the transition series. For early TM elements (Sc to Cr), the progressive filling of bonding-type *d*-states increases the binding energy modulus, enhancing the stability near the center of the series (Mn, Fe, Co), the occupation of non-bonding and anti-bonding states begins to counteract the bonding gains, leading to a peak in stability. Toward the end of the series (Zn, Cd, Hg), the full occupation of *d*-shells (*d*¹⁰ configuration) reduces the availability of bonding states, the interactions become predominantly repulsive due to filled anti-bonding character, sharply decreasing the binding energy modulus and contributing minimally to metallic bonding, as well as inducing anti-bonding interactions due to Pauli repulsion.

The decomposition of binding energy in Fig. 2 into its fundamental components, E_{int} and ΔE_{dist} , reveals a competitive interplay governing stability. The core-shell (*in*) configurations generally benefit from enhanced E_{int} compared to their *out* counterparts, which is attributed to the maximized coordination of the central TM atom, since it interacts with 12 nearest neighbors in the icosahedral cage. However, these configurations are penalized by geometric strain when there is a significant mismatch in size or electronic character between the dopant and host atoms. In contrast, the *out* configurations experience lower interaction energies because, when the dopant is surface-coordinated, the TM atom interacts with a reduced number of ligands, typically 6 to 7, resulting in lower interaction energy and greater exposure to surface-induced distortions.

Distortion energy plays a crucial role in modulating stability. The nanoclusters with lower binding energy modulus are more likely to have higher distortion energies, particularly the late TMs (Zn, Cd, Hg), where mismatch of electronic character and atomic sizes between dopant and host (Ti, Zr, Hf) causes geometric frustration. Interestingly, despite the larger distortion energies observed in *out* nanoclusters, the interaction energies are, in most cases, sufficient to compensate for them, preserving the overall thermodynamic stability. However, in specific systems where the distortion penalty becomes comparable to or exceeds the interaction gain, mainly with larger or electronically inert dopants, the *out* configuration becomes less favorable, as quantified by the relative total energy.

Thus, a more direct form to indicate the TM energetic preference in *in* and *out* configurations is obtained by subtracting the binding energies of the respective *in* and *out* configurations of $X_{12}TM$, resulting in the ΔE_{tot} , as shown in Fig. 3. The ΔE_{tot} profile quantitatively captures the energetic preference between *in* and *out* configurations, since negative values indicates a thermodynamic preference for the core-shell structure, predominantly observed for mid-row TMs (except for *4d* species in $Ti_{12}TM$), where the combination of bonding strength and favorable electronic hybridization maximizes stability. In contrast, positive values of ΔE_{tot} are characteristic of early and late TMs, where geometric strain or electronic inertness favors the dopant occupying surface (*out*) positions. The formation energies E_{form} as displayed in Supporting Information Fig. S11 reinforce these trends, confirming that the most stable configurations are those with minimized distortion and maximized bonding synergy between the host (Ti, Zr, Hf) and the dopant.

For $Fe_{12}TM$ nanoclusters, we observe that energetically they exhibit trends analogous to those observed for the group IV hosts (Ti, Zr, and Hf). The binding energies are consistently negative across the transition series, indicating thermodynamic stability regardless of the dopant position. A distinctive feature in Fe-based nanoclusters is the influence of Fe atomic radius and significant magnetic character. This effect leads to a larger relative distortion when dopants occupy the central site, particularly for heavier and larger TMs (e.g., Au, Hg). As a result, the distortion energy penalty increases for *in* configurations, making *out* arrangements energetically more competitive in selected cases, as clearly shown in the relative stability plots (Figs. S12 and S13). This is due to size mismatch and magnetic effects.

Geometric properties: The substitution of a TM atom as *in* or *out* configurations has a profound effect on the geometric and electronic structure of the nanoclusters. To capture this primarily from a geometric perspective, we have evaluated two key metrics, d_{av} and ECN, shown in Fig. 4. Core-shell structures always consist of smaller d_{av} values and higher ECN, signifying a closer-together bonding environment and a more symmetric arrangement inherent in the icosahedral cage. The high coordination of the central TM atom (ECN = 12) leads to stronger bonding, which shortens bond distances. On the other hand, *out* configurations exhibit d_{av} elongation and a significant reduction in ECN. This is attributed to two factors: surface localization of the dopant, which inherently reduces the number of nearest neighbors, and local relaxation effects, as the nanocluster adapts its

geometry to accommodate size and electronic mismatch. For instance, late TMs such as Cd, Hg, and Au exhibit the most pronounced deviations, characterized by bond length expansions exceeding 5 % relative to their *in* counterparts.

These structural metrics correlate directly with the energetic descriptors. For example, in *out* systems, structural flexibility is directly associated with weaker binding energies and higher distortion penalties. Systems with larger d_{av} and lower ECN tend to show higher ΔE_{tot} and less negative binding energies, particularly in the *out* configurations, supporting the classical principle that under-coordinated sites are structurally less stable but potentially more reactive. In this context, as shown in Fig. 4, the ECN is a quantitative descriptor of the local atomic environment between the two cases (*in* or *out*). In *in* configurations, the TM dopant achieves the maximum ECN, which is associated with more efficient bonding and higher stability. In *out* configurations, the ECN decreases significantly, typically below 7, depending on the specific TM and the host nanocluster, which indicates under-coordination and higher chemical reactivity. This behavior aligns with general trends in metallic bonding: atoms tend to maximize their coordination to lower the system’s energy. However, if the TM is placed outside the cage, steric hindrance and size mismatch can prevent ideal coordination, leading to a metastable yet locally stable structure.

For $Fe_{12}TM$ nanoclusters, we observe that, structurally, d_{av} and ECN (Fig. S14) reflect systematic differences between *in* and *out* configurations. The *in* systems possess shorter d_{av} and higher ECN values, reflecting the close coordination environment of the center TM atom. The *out* structures show a sharp reduction in ECN (typically below 7), which reflects the under-coordination and greater surface exposure (accessibility) trend.

Electronic properties: The energetic trends are intimately connected to the electronic structure of the binary nanoclusters, i.e., a detailed analysis of the electronic properties, as shown in Supporting Information Figs. S15 and S16 reveal a profound correlation between electronic structure and thermodynamic stability. In core-shell configurations, the different central TM atom experiences a ligand field analogous to a highly symmetric crystal field, splitting the *d*-orbitals into bonding and anti-bonding combinations with the surrounding shell. The degree of splitting and orbital hybridization depends on the *d*-electron count of the dopant, e.g., early TMs contribute to strong σ and π bonding interactions, enhancing stability; mid-series TMs reach maximum stabilization due to half-filled or nearly half-filled *d*-subshells, favoring exchange interactions; and late TMs exhibit filled *d*-states with minimal bonding contribution, resulting in weaker metal–metal interactions dominated by Pauli repulsion and size effects.

TMs-centered systems (*in*) register systematically lower ϵ_d values as displayed in Supporting Information Fig. S15. This shift down signals stronger hybridization between the TM *d*-orbitals and the metallic cage, stabilizing bonding states and lowering the energy. In *out* configurations, ϵ_d moves up, signaling weaker hybridization and a greater fraction of unsaturated or dangling electronic states. On the other hand, the HOMO–LUMO gap (Supporting Information Fig. S16) reflects electronic stability and potential chemical reactivity. Core-shell nanoclusters have larger gaps, consistent with an electronically more closed-shell, less reactive system. In contrast, *out* nanoclusters have systematically smaller gaps, particularly for mid-row and late-row TMs. This trend suggests that *out*-nanoclusters could exhibit enhanced chemical reactivity, making them promising candidates for catalytic applications where active sites at undercoordinated positions are desirable. We found that there is an anti-correlation between ϵ_d and the HOMO–LUMO gap, as ϵ_d increases (i.e., approaches the Fermi level), the HOMO–LUMO gap narrows, in agreement with the expected behavior from the *d*-band center theory,^[32,33] which links the position of ϵ_d to adsorption strength and catalytic activity.

For $Fe_{12}TM$ nanoclusters, we observe that from the electronic perspective, the HOMO-LUMO gap (Supporting Information Figs. S17 and S18) is typically narrower for *out* configurations, suggesting higher chemical reactivity and catalytic activity. Variations in electronic gaps also support the established understanding that *in* configurations, with stronger ligand fields and higher symmetry, lead to electronic delocalization and increased stability. Therefore, it is important to note that these electronic descriptors – ϵ_d and the HOMO–LUMO gap – provide the physics-grounded signals from which we have used for our PAI supervision. The nearly 3000 set of relaxed unary nanoclusters from the QCD and the 240 bimetallics DFT computations form a corpus that serves as the basis for the pre-training and fine-tuning stages discussed next.

2.2 Physics-Grounded Signals to Transformers Performance

After the hyperparameter search described in Methods, we selected the final model by comparing its complexity and performance during finetuning. The chosen FTTransformer has 6 transformer blocks with 8 attention heads, an embedding dimension of 16, and a feed-forward width of 128. Transformer blocks use swish, and the MLP head uses gelu activation with hidden-size factors (2, 1) – the first hidden layer has twice the flattened transformer output, and the second equals it. During fine-tuning, the backbone transformer and tokenizer remain frozen. More information can be found in the SI and in the GitHub [repository](#).

The finetuning was performed using robust Ridge regression. The original data is highly non-linear and complex. The transformer’s intricate architecture (attention mechanisms, residual connections, multiple layers) learns this complexity (Fig. S19). It combines relevant features from the input in a way that makes the underlying patterns more explicit and easier to discern. In this new space of transformed representation (embeddings), the features are structured such that the target variable (E_{form}^{in} and E_{form}^{out}) can be approximated well by a linear combination of these features. Therefore, we leverage the transformer architecture to simplify the fine-tuning stage, making it much more robust, faster, and numerically stable.

In this regard, we created different splits of fine-tuning sets. The total 120 rows of fine-tuning set is made of 30 rows with Fe as host atom (X_{12}) and the rest of 90 have Ti, Zr, Hf as host. By incrementally introducing host Fe in the training set of the finetuning stage, we observed the behavior of errors and prediction intervals. Fig. 5 Shows when the fraction of host Fe in holdout set is full (1.0) meaning all host Fe entries are kept in holdout set and the fine-tuned model was not exposed to any of those entries, the coverage of prediction interval for **in** and **out** configurations are 0.0 and 0.4, respectively, the MAE 3.4 eV, and R^2 score (coefficient of determination) -2 . But after introducing only two rows of host atom Fe (fraction of Fe in holdout set = 0.95, 0.05 in training set, $[0.05 \times 30] = 2$) in training of the fine-tuned model, the coverage jumps to 0.8 and MAE reduced to 1.1 and R^2 score improved to 0.7. Several factors, including the size of the finetuning set, influence the mean width of prediction intervals. With a limited fine-tuning set, the prediction intervals are kept conservatively large to ensure coverage. The trend shows a clear, rapid stability of model performance with a small amount of fine-tuning data. We observe that when the fraction of Fe in the holdout set is $x = 0.65$ (model fine-tuned with 10 Fe host entries), the model’s performance reaches a level of stability that can be considered representative. We picked this model to show its response details during fine-tuning and evaluation with a holdout set in Fig. 6. The results for both targets of E_{form}^{in} and E_{form}^{out} respectively are MAE of 0.15 and 0.11 for training step, and MAE of 0.67 and 0.68 on holdout set, the average prediction intervals are 3.37 and 3.24 eV, and the coverage of 0.9 and 1.0.

To gain a clearer view of the model’s underlying behavior, we performed token importance analysis at both training and evaluation, using a holdout set for the fine-tuning stage (Fig. 7). The SHAP (Shapley Additive exPlanations^[34]) value of a feature (token) is the average marginal contribution of that feature to the model output. The training set analysis shows which features the model learned to prioritize based on the data it saw. The holdout set analysis shows whether those same priorities hold for new, unseen data. The features are sorted in descending order by importance. For both configurations **in** and **out**, we observe that the top-most essential features in the training step are identical to those in the holdout dataset. For configuration **in**, this is valid for top-10 features, and for configuration **out**, the top-12 features (limited to 12 out of a maximum of 66 for printing limitation only). Another valuable observation is that the distributions of importance across the top features are very similar in both the training/holdout datasets, and across both **in** and **out** configurations. The importance is distributed evenly across features and is not concentrated in a few. Another observation is that both categorical, numerical, and engineered features, across a wide range, such as electronic, energetic, and physicochemical properties, contribute in accordance with the model targets. These results show a robust, accurate, and reliable model.

Our analyses demonstrate that the fine-tuned transformer is more than a black box, since it internalizes chemically interpretable rules. The most influential tokens reliably represent features such as size mismatch, d -electron count, coordination environment, and d -band-related descriptors. Essentially, such factors closely resemble the heuristics chemists use to explain trends in core-shell stability and reactivity. In essence, the model systematically evaluates the same competing factors, strain versus coordination, and electronic filling versus hybridization, that underlie the DFT trends in Figs. 2–4. This indicates that the model not only extrapolates beyond the computed dataset, but does so in a manner analogous to expert chemical reasoning.

3 Methods

3.1 DFT Calculations and Computational Details

All DFT results reported here are obtained from spin-polarized scalar-relativistic calculations^[35,36] within the generalized gradient approximation for the exchange-correlation functional according to Perdew–Burke–Ernzerhof^[37] (PBE) parametrization. Those calculations are performed with the Vienna *Ab-initio* Simulation Package (VASP).^[38–40] The Kohn–Sham (KS) equations are solved with the all-electron projector augmented wave method,^[41,42] implemented in VASP,^[38–40] with KS orbitals expanded in plane-waves up to a cutoff energy of 500 eV, corresponding to a 20 % larger than the VASP recommended value (ENMAX).

Our nanocluster calculations are simulated in a cubic supercell of size 20 Å, which avoids interactions with periodic images, since the distance between them is at least 14 Å. On the other hand, the respective free-atoms are simulated in an orthorhombic supercell of size 19 Å×19.25 Å×19.5 Å. As nanoclusters and free-atoms, there is no electronic dispersion; the Brillouin zone (BZ) integration was performed with a single **k**-point (Γ-point). The nanocluster equilibrium geometries were obtained until atomic forces on every atom were smaller than 0.015 eV/Å and a total energy convergence of 1.0×10^{-6} eV was achieved. The main convergence tests for binary nanoclusters are presented in Supporting Information (Tables S1–S4). At the same time, specific reference results for TM bulk systems (cohesive energy, E_{coh} , and atomic radius, R_{TM}), are calculated and presented in Supporting Information Fig. S1, in agreement with our previous work.^[43]

3.2 Atomic Configurations

The *xyz* configurations for 13-atom unary nanoclusters of Group 4 (Ti, Zr, and Hf), which consist in the main structural set investigated, and Fe, complementary set used for a validation case, are obtained (and re-optimized) as the lowest energy structures from our previous works.^[30,31] For several TM_{13} nanoclusters, the ICO geometry is the most stable configuration, as expected; this is also the case for Ti_{13} , Zr_{13} , Hf_{13} , and Fe_{13} .^[30,31] This quasi-spherical geometry has a central atom and 12 (external) equivalent, equidistant atoms, with high-symmetry I_h and a closed-packed structure.

For the binary nanoclusters, we replaced one atom central or external from X_{13} , where X = Ti, Zr, Hf, or Fe by TM (30 elements from the Periodic Table). For the central atom, there is only one option to consider, while for the external case, any of the 12 outer atoms can be chosen for replacement, since they are all symmetrically equivalent. Thus, we have two possibilities for each system, with TM preferring to stay inside the 12-atom shell (X_{12}), called *in* ($\text{X}_{12}\text{TM}^{\text{in}}$), or TM outside, called *out* ($\text{X}_{12}\text{TM}^{\text{out}}$), as shown in Fig. 1. We have performed DFT-PBE calculations to test these two configurational sets: *in* and *out*, where the *in* configurations are also treated as core-shells. All unary and binary nanoclusters are relaxed without any symmetry or spin constraint to allow symmetry breaking and a complete potential energy surface exploration. The complete optimized set of structures for $\text{X}_{12}\text{TM}^{\text{in}}$, $\text{X}_{12}\text{TM}^{\text{out}}$, and the TM bulk properties are presented in the Supporting Information (Figs. S1–S9).

To characterize the $\text{X}_{12}\text{TM}^{\text{in}}$ and $\text{X}_{12}\text{TM}^{\text{out}}$ systems, we have studied the main energetic, structural, and electronic properties. For example, for elucidated the stability of the systems, in addition to the calculation of vibrational frequencies ($3N - 6$ vibrational modes, where $N = 13$) as an indication of whether the systems formed are in fact local minima, we have also calculated the binding energy per atom (E_b), which is giving by:

$$\begin{aligned} E_b &= (E_{\text{tot}}^{\text{X}_{12}\text{TM}} - 12E_{\text{tot}}^{\text{X f-a}} - E_{\text{tot}}^{\text{TM f-a}}) / 13 = \\ &= (12E_b^u + E_{\text{int}} + 12\Delta E_{\text{dist}}) / 13, \end{aligned} \quad (1)$$

where the first expression is composed by the total energies of the *in* or *out* X_{12}TM systems ($E_{\text{tot}}^{\text{X}_{12}\text{TM}}$) and the X ($E_{\text{tot}}^{\text{X f-a}}$) and TM free-atoms ($E_{\text{tot}}^{\text{TM f-a}}$), while the second expression is a decomposition of E_b into the binding energy of X_{12} ICO-derived structure (E_b^u), the interaction energy between X_{12} and TM in the equilibrium geometry of $\text{X}_{12}\text{TM}^{\text{in}}$ or $\text{X}_{12}\text{TM}^{\text{out}}$ (E_{int}), and the distortion energy occasioned by interaction between TM species and X_{12} structure (ΔE_{dist}). More details about the E_b equation are given in the Supporting Information and previous works.^[43–45]

By subtracting the binding energies of the respective *in* and *out* configurations of X_{12}TM , we can have a direct idea of the TM preference to stay inside the X_{12} cage, forming the core-shell (*in*), or to stay outside, forming the *out* configurations. Subtracting the binding energies is equivalent to subtracting the total energies of the

configurations of interest, according to the following ΔE_{tot} equation:

$$\Delta E_{tot} = E_{tot}^{in} - E_{tot}^{out}, \quad (2)$$

where E_{tot}^{in} and E_{tot}^{out} are the total energies of the $X_{12}TM^{in}$ and $X_{12}TM^{out}$ configurations, respectively. This expression indicates which configuration is more likely, giving preference to the formation of *in* (core-shell) configurations for negative relative energies and to the formation of *out* configurations for positive relative energies. Another energy check regarding the stability of binary systems *in* or *out* is the formation energy, E_{form} , obtained by the equation:

$$E_{form} = E_{tot}^{X_{12}TM} - (E_{tot}^{X_{12}} + E_{tot}^{TM f-a}), \quad (3)$$

where $E_{tot}^{X_{12}}$ is the total energy of the lowest energy X_{12} systems. More details about the energy decomposition equations are presented in Supporting Information.

Structurally, we used the effective coordination number concept^[46,47] to calculate the average bond lengths (d_{av}) and the effective coordination numbers (ECN). For more details, see the Supporting Information. This tool employs a self-consistent approach, utilizing an exponential decay function to derive the effective geometric parameters. Thus, it enables providing an adequate description of the d_{av} and ECN parameters of a nanocluster, accounting for possible distortions.^[30,31]

Electronically, we characterized the binary *in* and *out* nanoclusters by obtaining the highest occupied molecular orbital – lowest unoccupied molecular orbital gap (HOMO-LUMO gap) and the center of gravity of the occupied *d*-states (ϵ_d). While the first one is well-defined and consist of the energy difference between the highest energy electronic state that is occupied (HOMO) and the lowest energy electronic state that is unoccupied (LUMO) in a finite system (nanocluster), the second one is based on the *d*-band model proposed by Hammer and Nørskov,^[33] and can be correlated with the adsorption strength of an adsorbate on the surface of the nanocluster.

3.3 Prediction Model

The prediction of E_{form} and structural preferences (*in/out* configurations) was conducted through a feature-tokenizer transformer for tabular data (FTTransformer).^[48] This architecture is designed to handle mixed-type tabular inputs, mapping each categorical and continuous feature to a token in a shared embedding space. These tokens are then processed by stacked self-attention blocks and a multilayer perceptron (MLP) prediction head. We use 2968 entries of unary data (QCD) to learn generalizable atomic and feature representations for pretraining, and 120 entries of a binary dataset for finetuning and implementing the transfer learning strategy.

Categorical and numerical feature spaces: Our model uses two main types of features: (i) categorical variables, such as atomic numbers for the two sites (TM1, TM2) and discrete periodic-table descriptors for each site and (ii) continuous variables, including base structural/energetic terms and engineered physicochemical descriptors. To enrich the input, we augmented the base columns with per-site periodic-table numeric properties, electron-configuration descriptors, and pairwise features such as absolute differences and averages (e.g., $|\Delta|$, $\text{avg}(\cdot)$), as well as binary flags for when categorical properties are the same. Categorical features are consistently integer-encoded across both the monometal and bimetal datasets, ensuring fixed embedding sizes. Specifically, we use 10 categorical inputs: two atomic-number columns and, for each of the four categorical properties (OxidationStates, StandardState, GroupBlock, ElectronConfiguration), one column per site. Numerical features comprise 56 inputs: four base continuous variables (e.g., $n_{atoms,1}$, $n_{atoms,2}$, coe_1 , coe_2); for each of the eight numeric periodic properties (AtomicMass, AtomicRadius, Electronegativity, IonizationEnergy, ElectronAffinity, MeltingPoint, BoilingPoint, Density), we include the pairwise difference and average; for each categorical property, a same/different binary flag; and for the eight electron-configuration descriptors (core_n_electron, 3d, 4s, 4d, 4f, 5s, 5d, 6s), site-wise values for both sites and their pairwise difference and average. In total, this yields 66 input features (10 categorical and 56 continuous). All continuous inputs and targets are robustly scaled to reduce the influence of outliers and improve optimization stability.

Model architecture and training: We performed automated hyperparameter optimization with Optuna,^[49] exploring a range of architectural and training parameters (embedding dimension, number of attention heads, transformer blocks, feed-forward width, dropout rate, activation functions, learning rate, batch size, weight

decay, and loss function). Trials were run in parallel using Dask.^[50] The top model architectures were manually evaluated for best finetuning performance, and the final decision was made accordingly.

Categorical features are embedded and stacked into a token sequence. Continuous features are normalized and also projected into an embedding dimension, forming a token. The complete token sequence is processed by several transformer blocks (multi-head self-attention and feed-forward layers with dropout and normalization). The output is flattened and passed to an MLP head with dropout, producing two values per sample: E_{form}^{in} and E_{form}^{out} . Training uses the Adam optimizer^[51] with early stopping. Transfer learning is performed by first training the whole model, then freezing the transformer backbone, and fine-tuning using Ridge regression between the transformed representation (embeddings) and the fine-tuning targets.

To quantify uncertainty, we generated calibrated prediction intervals using 5-fold cross-validation (CV+) on the bimetal fine-tuning pool, with the transformer backbone frozen. For each fold, we collect validation residuals and compute holdout predictions, using the median across folds as the final prediction. Intervals are constructed using a fast CV+ mode with per-output residual scoring and a stratification scheme (“mixed Mondrian”)^[52] that groups samples by predicted value or outlier score. Within each group, we compute empirical quantities to set interval widths, ensuring coverage of prediction intervals.

Implementation: Implementation was structured in a modular configuration in a standalone repository (https://github.com/KIT-Workflows/Nanocluster_Transformers). The repository holds: (i) Jupyter Notebooks for model administration (*model_manager.ipynb*), end-to-end management (*manager.ipynb*), hyperparameter tuning using Optuna (*optimization.ipynb*), and visualization (*view.ipynb*). (ii) Python scripts for top-level model declarations (*models.py*), training procedures (*training.py*), and data preprocessing pipelines. (iii) Raw (*src/data_actual*) and processed data directories (*src/data_processed*), final models (*src/models_save*), and temporary models generated during optimization experiments (*src/tmp/models*). The codebase also includes dynamic visualization libraries and a Flask-based graphical user interface (GUI) for model interpretability and user interaction. Overall, this integrated DFT–ML framework provides a predictive, transferable tool for exploring the configurational landscape of bimetallic nanoclusters, enabling rapid screening and design guidance beyond the computed dataset.

Funding

The project (HGF ZT-I-PF-5-261 GENIUS) underlying this publication is/was funded by the Initiative and Networking Fund of the Helmholtz Association in the framework of the Helmholtz AI project call.

Acknowledgements

This project was in part supported by DFG project WE 1863/41-1 in SPP2363. Authors thank the Rio Grande do Sul Research Foundation (FAPERGS, grant 24/2551 – 0001551 – 5), the Federal District Research Support Foundation (FAPDF, grants 00193 – 00001817/2023 – 43 and 00193 – 00002073/2023 – 84), the National Council for Scientific and Technological Development – CNPq (303206/2025 – 0, 408144/2022 – 0, 305174/2023 – 1, 444431/2024 – 1, 141176/2024 – 5, and 444069/2024 – 0), and the Coordination for Improvement of Higher Level Education – CAPES (finance Code 001) for the financial support. Part of this work was performed on the HoreKa supercomputer funded by the Ministry of Science, Research, and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research. A.C.D. also acknowledges funding from PDPG-FAPDF-CAPES Centro-Oeste grant number 00193 – 00000867/2024 – 94.

Data Availability Statement

The data that supports the findings of this study are available at https://github.com/KIT-Workflows/Nanocluster_Transformers or from the corresponding author upon reasonable request.

Code Availability

The code supporting this study’s findings is available at https://github.com/KIT-Workflows/Nanocluster_Transformers.

References

- [1] R. Ferrando, J. Jellinek, and R. L. Johnston. Nanoalloys: From Theory to Applications of Alloy Clusters and Nanoparticles. *Chem. Rev.*, 108:845–910, 2008. doi: 10.1021/cr040090g.
- [2] L.-L. Wang and D. D. Johnson. Predict trends of core-shell preferences for 132 late transition-metal binary-alloy nanoparticles. *J. Am. Chem. Soc.*, 131:14023–14029, 2009. doi: 10.1021/ja903247x.
- [3] M. B. Gawande, A. Goswami, T. Asefa, H. Guo, A. V. Biradar, D.-L. Peng, R. Zboril, and R. S. Varma. Core-shell nanoparticles: synthesis and applications in catalysis and electrocatalysis. *Chem. Soc. Rev.*, 44: 7540–7590, 2015. doi: 10.1039/C5CS00343A.
- [4] P. C. D. Mendes, S. G. Justo, J. Mucelini, M. D. Soares, K. E. A. Batista, M. G. Quiles, M. J. Piotrowski, and J. L. F. Da Silva. Ab initio insights into the formation mechanisms of 55-atom pt-based core-shell nanoalloys. *J. Phys. Chem. C*, 124:1158–1164, 2020. doi: 10.1021/acs.jpcc.9b09561.
- [5] N. Eom, M. E. Messing, J. Johansson, and K. Deppert. General trends in core-shell preferences for bimetallic nanoparticles. *ACS Nano*, 15:8883–8895, 2021. doi: 10.1021/acsnano.1c01500.
- [6] I. Chakraborty and T. Pradeep. Atomically precise clusters of noble metals: Emerging link between atoms and nanoparticles. *Chem. Rev.*, 117:8208–8271, 2017. doi: 10.1021/acs.chemrev.6b00769.
- [7] A. W. Castleman, Jr. and S. N. Khanna. Clusters, superatoms, and building blocks of new materials. *J. Phys. Chem. C*, 113:2664–2675, 2009. doi: 10.1021/jp806850h.
- [8] Y. Yamamoto, T. Miura, M. Suzuki, N. Kawamura, H. Miyagawa, T. Nakamura, K. Kobayashi, T. Teranishi, and H. Hori. Direct observation of ferromagnetic spin polarization in gold nanoparticles. *Phys. Rev. Lett.*, 93:116801, 2006. doi: 10.1103/PhysRevLett.93.116801.
- [9] M. Pereiro, D. Baldomir, and J. E. Arias. Unexpected magnetism of small silver clusters. *Phys. Rev. A*, 75: 063204, 2007. doi: 10.1103/PhysRevA.75.063204.
- [10] A. L. Mackay. A Dense Non-Crystallographic Packing of Equal Spheres. *Acta Crystallogr.*, 15:916–918, 1962. doi: 10.1107/s0365110x6200239x.
- [11] J. D. Aiken III and R. G. Finke. A review of modern transition-metal nanoclusters: their synthesis, characterization, and applications in catalysis. *J. Mol. Catal. A: Chem.*, 145:1–44, 1999. doi: 10.1016/S1381-1169(99)00098-9.
- [12] J. A. Alonso. Electronic and Atomic Structure, and Magnetism of Transition-Metal Clusters. *Chem. Rev.*, 100:637–678, 2000. doi: 10.1021/cr980391o.
- [13] Francesca Baletto and Riccardo Ferrando. Structural Properties of Nanoclusters: Energetic, Thermodynamic, and Kinetic Effects. *Rev. Mod. Phys.*, 77:371–423, 2005. doi: 10.1103/revmodphys.77.371.
- [14] A. Fernando, K. L. Dimuthu, M. Weerawardene, N. V. Karimova, and C. M. Aikens. Quantum mechanical studies of large metal, metal oxide, and metal chalcogenide nanoparticles and clusters. *Chem. Rev.*, 115: 6112–6216, 2015. doi: 10.1021/cr500506r.
- [15] N. Toshima and T. Yonezawa. Bimetallic nanoparticles - novel materials for chemical and physical applications. *New. J. Chem.*, 22:1179–1201, 1998. doi: 10.1039/A805753B.
- [16] S. Ke, Y. Zhao, X. Min, X. Zhu, X. Li, B. Yang, F. Yang, X. Wu, R. Mi, Y. Liu Z. Huang, and M. Fang. Tailoring the d-band center in pt-based catalysts for hydrogen evolution via transition metals incorporation. *Int. J. Hydrogen Energy*, 105:806–816, 2025. doi: 10.1016/j.ijhydene.2025.01.303.
- [17] K. E. A. Batista, J. L. F. Da Silva, and M. J. Piotrowski. Adsorption of co, no, and h₂ on the pd_nau_{55-n} nanoclusters: A density functional theory investigation within the van der waals d3 corrections. *J. Phys. Chem. C*, 123:7431–7439, 2019. doi: 10.1021/acs.jpcc.8b12219.

- [18] C. H. Wu, C. Liu, D. Su, H. L. Xin, H.-T. Fang, B. Eren, S. Zhang, C. B. Murray, and M. B. Salmeron. Bimetallic synergy in cobalt-palladium nanocatalysts for co oxidation. *Nat. Catal.*, 2:78–85, 2019. doi: 10.1038/s41929-018-0190-6.
- [19] D. Kaya, I. Demiroglu, I. B. Isik, H. H. Isik, S. K. Çetin, C. Sevik, A. Ekicibil, and F. Karadag. Highly active bimetallic pt-cu nanoparticles for the electrocatalysis of hydrogen evolution reactions: Experimental and theoretical insight. *Int. J. Hydrogen Energy*, 48:37209–37223, 2023. doi: 10.1016/j.ijhydene.2023.06.100.
- [20] S. Manna, Y. Wang, A. Hernandez, P. Lile, S. Liu, and T. Mueller. A database of low-energy atomically precise nanoclusters. *Scientific Data*, 10:308, 2023. doi: 10.1038/s41597-023-02200-4.
- [21] Mohammad Soleymanibrojeni, Celso Ricardo Caldeira Rego, Meysam Esmailpour, and Wolfgang Wenzel. An active learning approach to model solid-electrolyte interphase formation in Li-ion batteries. *J. Mater. Chem. A*, 12(4):2249–2266, 2024.
- [22] Xin Huang, Ashish Khetan, Milan Cvitkovic, and Zohar Karnin. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*, 2020.
- [23] Celso R. C. Rêgo, Jörg Schaarschmidt, Tobias Schlöder, Montserrat Penaloza-Amion, Saientan Bag, Tobias Neumann, Timo Strunk, and Wolfgang Wenzel. SimStack: An intuitive workflow framework. 9, 2022.
- [24] Joerg Schaarschmidt, Jie Yuan, Timo Strunk, Ivan Kondov, Sebastiaan P. Huber, Giovanni Pizzi, Leonid Kahle, Felix T. Bölle, Ivano E. Castelli, Tejs Vegge, Felix Hanke, Tilmann Hickel, Jörg Neugebauer, Celso R. C. Rêgo, and Wolfgang Wenzel. Workflow engineering in materials design within the battery 2030+ project. *Adv. Energy Mater.*, 12(17):2102638, 2021.
- [25] Carlos Maciel de O. Bastos, Alexandre C. Dias, Ana Carolina Ferreira de Brito, Ingrid D. Barcelos, Andréia Luisa da Rosa, Danilo Neves Silveira, Maurício J. Piotrowski, Wolfgang Wenzel, Celso R. C. Rêgo, and Diego Guedes-Sobrinho. First-principles statistical investigation of thermodynamic behavior with excitonic effects in mo1-xwxse2 alloys through a data-driven workflow approach. *Journal of Materials Chemistry A*, 2025. ISSN 2050-7496. doi: 10.1039/d5ta02721g. URL <http://dx.doi.org/10.1039/D5TA02721G>.
- [26] Celso R. C. Rêgo, Wolfgang Wenzel, Maurício J. Piotrowski, Alexandre C. Dias, Carlos Maciel de Oliveira Bastos, Luis O. de Araujo, and Diego Guedes-Sobrinho. Digital workflow optimization of van der Waals methods for improved halide perovskite solar materials. 2025.
- [27] Simon Bekemeier, Celso Ricardo Caldeira Rêgo, Han Lin Mai, Ujjal Saikia, Osamu Waseda, Markus Apel, Felix Arendt, Alexander Aschemann, Bernd Bayerlein, Robert Courant, Gordian Dziwis, Florian Fuchs, Ulrich Giese, Kurt Junghanns, Mohamed Kamal, Lukas Koschmieder, Sebastian Leineweber, Marc Luger, Marco Lukas, Jürgen Maas, Jana Mertens, Björn Mieller, Ludger Overmeyer, Norbert Pirch, Jan Reimann, Sebastian Schröck, Philipp Schulze, Jörg Schuster, Alexander Seidel, Oleg Shchyglo, Marek Sierka, Frank Silze, Simon Stier, Marvin Tegeler, Jörg F. Unger, Matthias Weber, Tilmann Hickel, and Jörg Schaarschmidt. Advancing digital transformation in material science: The role of workflows within the materialdigital initiative. 2025.
- [28] J. F. Dalmedico, D. N. Silveira, L. O. de Araujo, W. Wenzel, C. R. C. Rêgo, A. C. Dias, D. Guedes-Sobrinho, and Maurício J. Piotrowski. Tuning electronic and structural properties of lead-free metal halide perovskites: A comparative study of 2D Ruddlesden-Popper and 3D compositions. *ChemPhysChem*, 25(16):e202400118, 2024.
- [29] W. A. Harrison. *Electronic structure and the properties of solids: the physics of the chemical bond*. Courier Corporation, 2012.
- [30] M. J. Piotrowski, P. Piquini, and J. L. F. Da Silva. Density functional theory investigation of 3d, 4d, and 5d 13-atom metal clusters. *Phys. Rev. B*, 81:155446, 2010. doi: 10.1103/PhysRevB.81.155446.

- [31] A. S. Chaves, M. J. Piotrowski, and J. L. F. Da Silva. Evolution of the structural, energetic, and electronic properties of the $3d$, $4d$, and $5d$ transition-metal clusters (30 TM_n systems for $n=2-15$): A density functional theory investigation. *Phys. Chem. Chem. Phys.*, 19(23):15484–15502, 2017. doi: 10.1039/C7CP02240A.
- [32] B. Hammer and J. K. Nørskov. Electronic factors determining the reactivity of metal surfaces. *Surf. Sci.*, 343:211–220, 1995. doi: 10.1016/0039-6028(96)80007-0.
- [33] B. Hammer and J. K. Nørskov. *Advances in Catalysis*. Academic Press Inc, San Diego, 2000.
- [34] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30:4765–4774, 2017.
- [35] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964. doi: 10.1103/PhysRev.136.B864.
- [36] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965. doi: 10.1103/PhysRev.140.A1133.
- [37] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996. doi: 10.1103/PhysRevLett.77.3865.
- [38] G. Kresse and J. Hafner. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B*, 48:13115–13126, 1993. doi: 10.1103/PhysRevB.48.13115.
- [39] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, 1996. doi: 10.1103/PhysRevB.54.11169.
- [40] J. Hafner. *Ab – initio* simulations of materials using VASP: Density-functional theory and beyond. *J. Comput. Chem.*, 29:2044–2078, 2008. doi: 10.1002/jcc.21057.
- [41] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, 1994. doi: 10.1103/PhysRevB.50.17953.
- [42] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, 1999. doi: 10.1103/PhysRevB.59.1758.
- [43] M. J. Piotrowski, J. M. T. Palheta, and R. Fournier. Cage doping of ti, zr, and hf-based 13-atom nanoclusters: two sides of the same coin. *Phys. Chem. Chem. Phys.*, 26:13172–13181, 2024. doi: 10.1039/d4cp00518j.
- [44] A. F. Yonezawa, G. R. Nagurniak, R. P. Orenha, E. H. da Silva, R. L. T. Parreira, and M. J. Piotrowski. Stability changes in iridium nanoclusters via monoxide adsorption: A dft study within the van der waals corrections. *J. Phys. Chem. A*, 125:4805–4818, 2021. doi: 10.1021/acs.jpca.1c02694.
- [45] J. P. C. S. Felix, K. E. A. Batista, W. O. Morais, G. R. Nagurniak, R. P. Orenha, C. R. C. Rêgo, D. Guedes-Sobrinho, R. L. T. Parreira, M. M. Ferrer, and M. J. Piotrowski. Molecular adsorption on coinage metal subnanoclusters: A DFT+D3 investigation. *J. Comput. Chem.*, 44:1040–1051, 2023. doi: 10.1002/jcc.27063.
- [46] R. Hoppe. The coordination number – an ”inorganic chameleon”. *Angew. Chem. Int. Ed.*, 9:25–34, 1970. doi: 10.1002/anie.197000251.
- [47] R. Hoppe. Effective Coordination Numbers (ECoN) and Mean Active Fictive Ionic Radii (MEFIR). *Z. Kristallogr.*, 150:23–52, 1979. doi: 10.1524/zkri.1979.150.1-4.23.
- [48] Yury Gorishniy, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. Revisiting deep learning models for tabular data. *Advances in neural information processing systems*, 34:18932–18943, 2021.

- [49] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2623–2631, 2019. doi: 10.1145/3292500.3330701.
- [50] Matthew Rocklin. Dask: Parallel computation with blocked algorithms and task scheduling. pages 130 – 136, 2015.
- [51] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [52] Alexander Timans, Christoph-Nikolas Straehle, Kaspar Sakmann, and Eric Nalisnick. Adaptive bounding box uncertainties via two-step conformal prediction. pages 363–398, 2024.

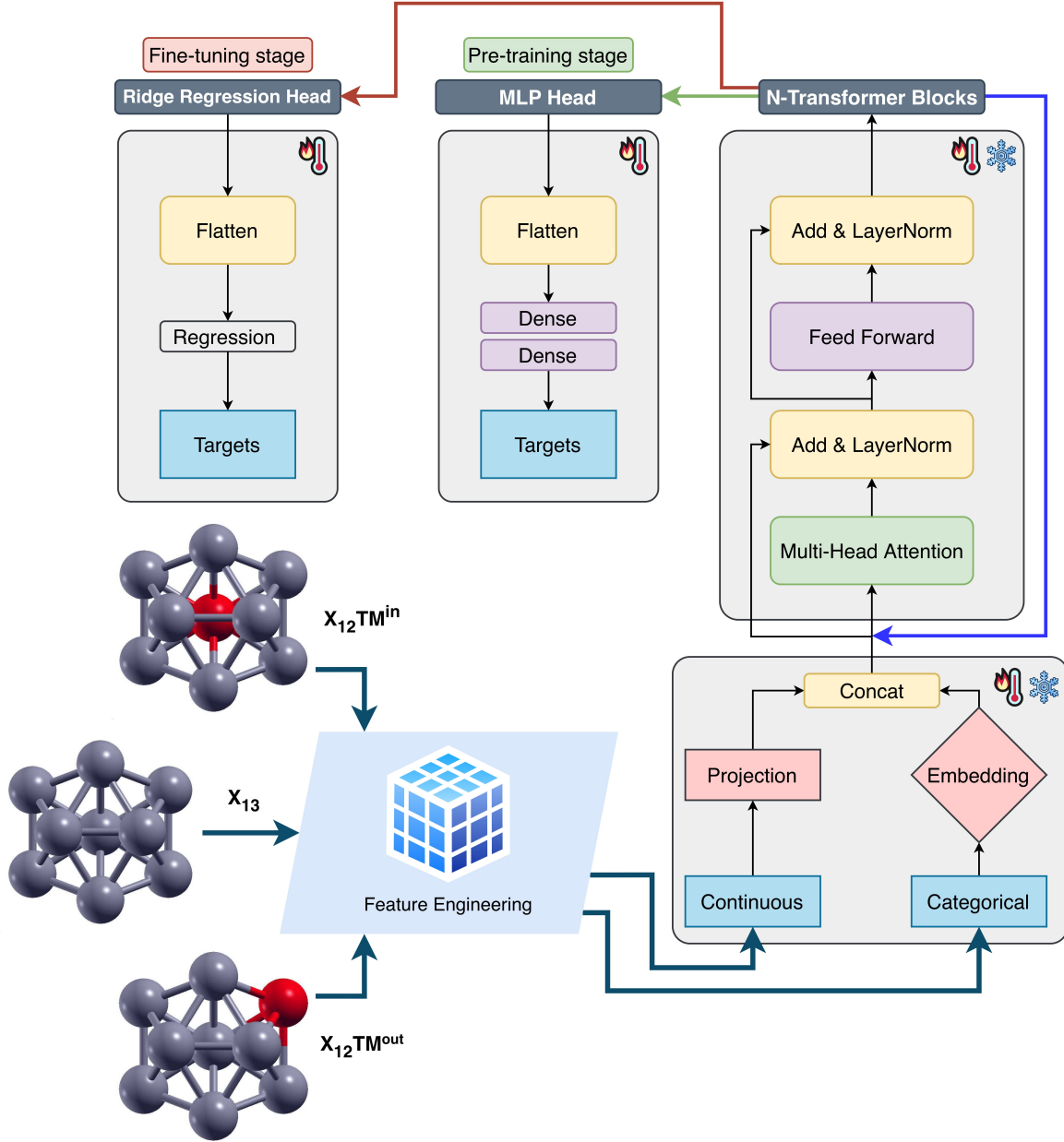


Figure 1. Schematic of the dataset and model pipeline. A 13-atom icosahedral cluster X_{13} is considered with two single-atom substitution topologies: (i) inner-site (core) replacement $X_{12}TM^{in}$ and (ii) outer/surface-site replacement $X_{12}TM^{out}$. Here $X \in \{Ti, Zr, Hf\}$ and TM spans the $3d-5d$ series; *in* and *out* denote substitution at the central atom and at one of the 12 surface vertices, respectively. Feature engineering produces continuous and categorical descriptors that are projected/embedded and concatenated, then passed through N Transformer blocks (frozen during head-only fine-tuning) to capture cross-feature interactions. A lightweight MLP head (trainable) flattens the representation and predicts the targets, e.g., E_{form} and ΔE_{tot} for *in* and *out* cases.

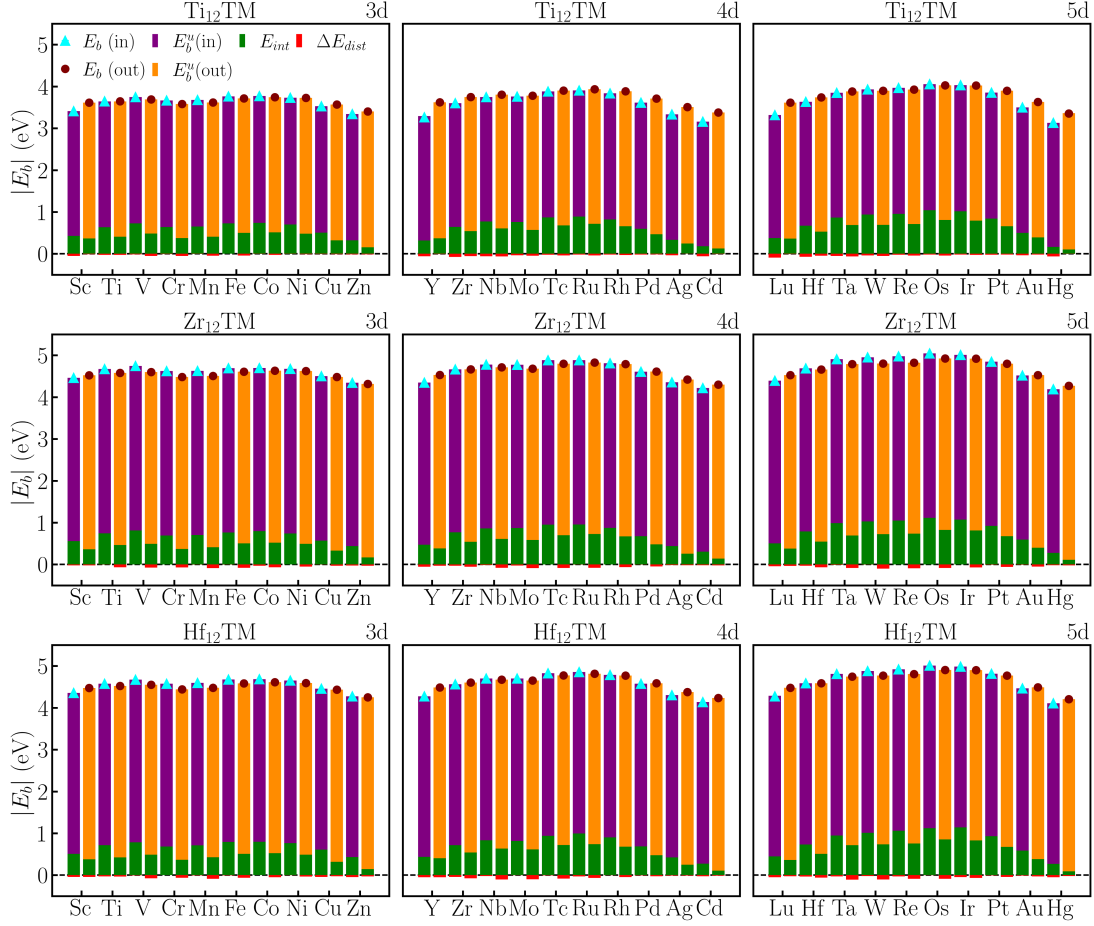


Figure 2. The binding energy modulus is shown as function of the atomic number for $X_{12}TM^{in}$ (triangle) and $X_{12}TM^{out}$ (circle) ($X = \text{Ti, Zr, Hf}$) nanoclusters. The energy decomposition presents the binding energy of X_{12} ICO-derived structure (E_b^u), the interaction energy between X_{12} and TM in the equilibrium geometry of $X_{12}TM^{in}$ or $X_{12}TM^{out}$ (E_{int}), and the distortion energy occasioned by interaction between TM species and X_{12} structure (ΔE_{dist}).

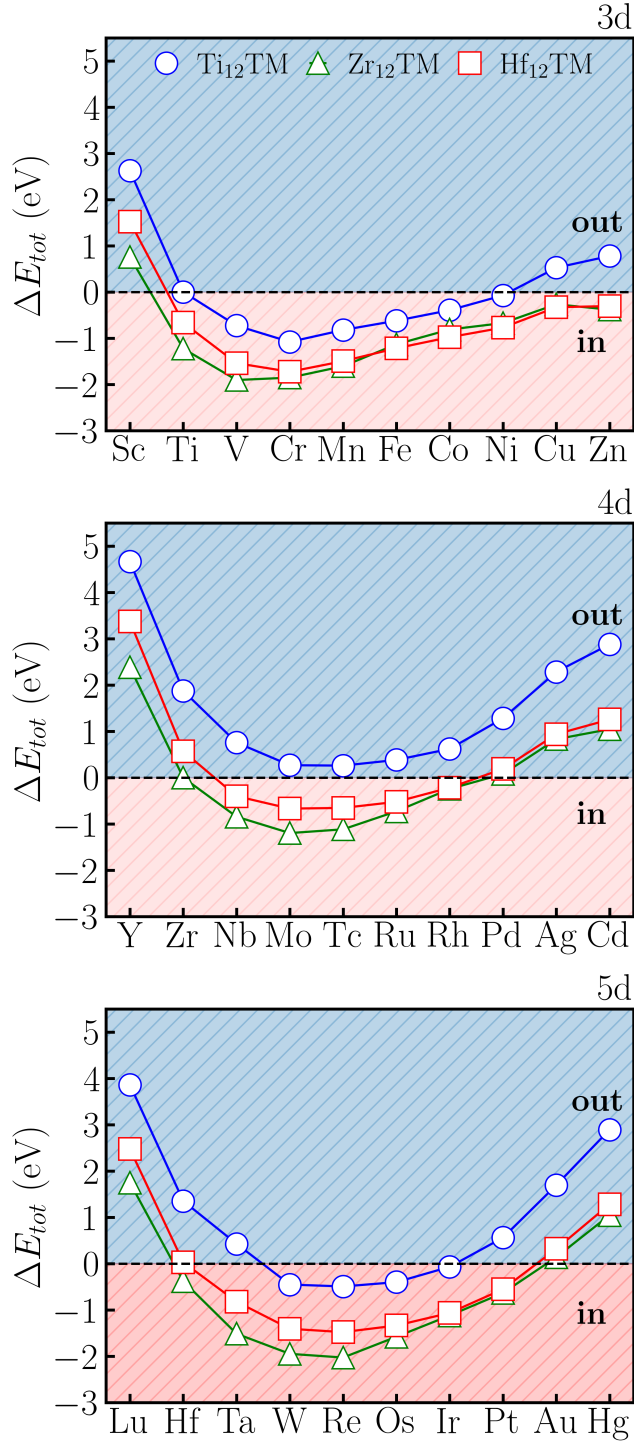


Figure 3. Relative total energy (ΔE_{tot}) as a function of atomic number for $Ti_{12}TM$ (blue circles), $Zr_{12}TM$ (green triangles), and $Hf_{12}TM$ (red squares) nanoclusters. Negative values indicate energetic stabilization of *in* configurations, while positive values indicate stabilization of *out* configurations.

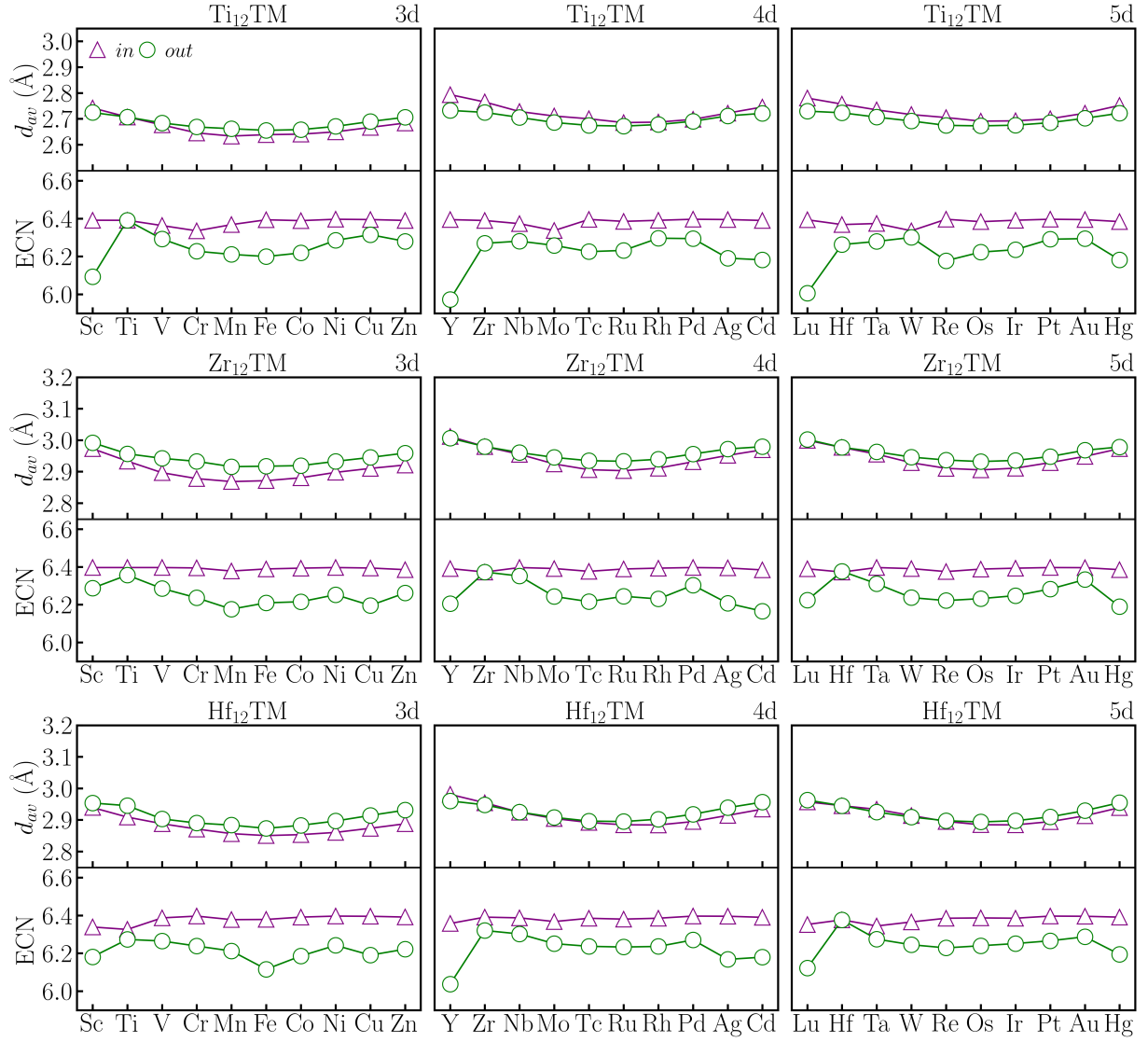


Figure 4. Structural properties: the average bond lengths (d_{av}) and the effective coordination number (ECN) as function of the TM atomic number for $X_{12}TM^{in}$ (triangle) and $X_{12}TM^{out}$ (circle) ($X = Ti, Zr, Hf$) nanoclusters.

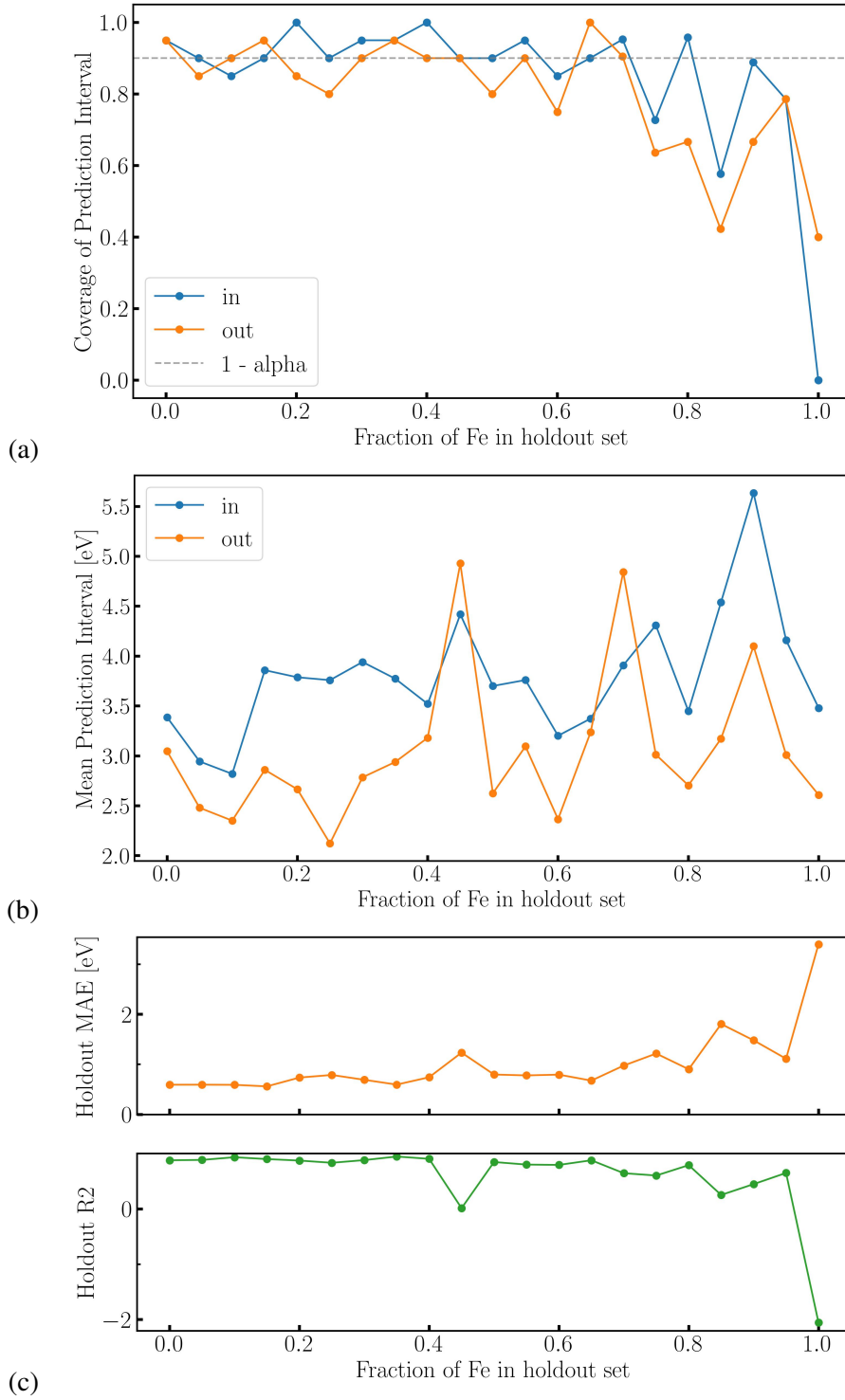


Figure 5. Fraction of the total Fe host atoms (X_{12}) in holdout set compared against the coverage and width of prediction intervals, as well as MAE and R^2 score. This fraction changes between 0.0 and 1.0. When 0.0 that means all Fe host atoms are in the training set and the holdout set does not include any entries with Fe as host atom instead with other elements, when 1.0 the opposite holds. The minimum size of the holdout set is 20 and the maximum is 30. At $x = 0.95$ which is insertion of only 2 entries containing Fe host atoms in training set, all metrics improve, at $x = 0.65$ and lower x , we can see a stable trend starts. (a) The coverage of prediction intervals, $\alpha = 0.1$, (b) Mean prediction intervals for both *in* and *out* configurations, (c) MAE for both targets in the holdout set, (d) R^2 score is the regression score for both targets.

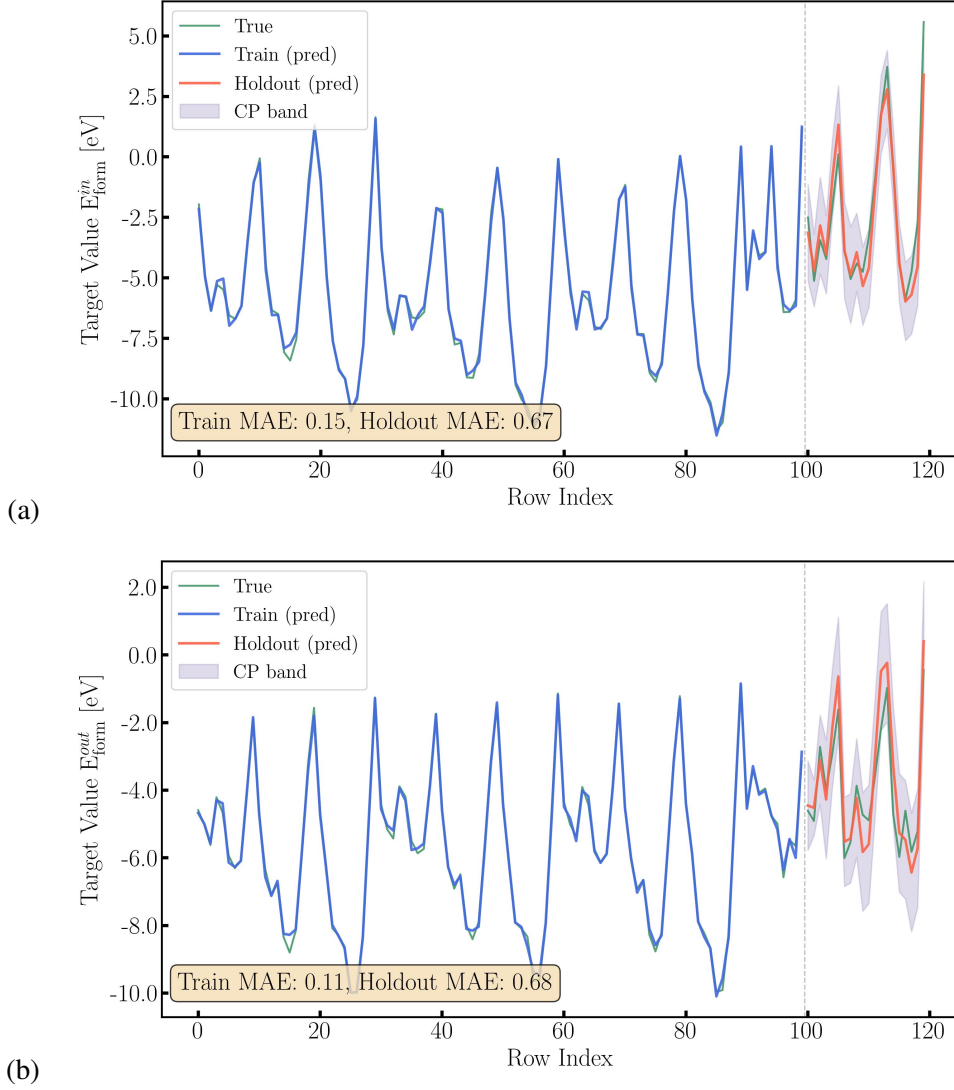


Figure 6. Fine-tuning stage with 5-fold CV+, and conformal prediction intervals with $\alpha = 0.1$. Bands denote calibrated prediction intervals. For (a) *in* configuration the mean width of prediction intervals is 3.37 eV, coverage 0.9, and holdout MAE 0.67 eV, and for (b) *out* the mean width of prediction intervals is 3.24 eV, coverages 1.0, and holdout MAE 0.68 eV. These results belong to the model trained at $x = 0.65$ "Fraction of Fe in holdout set" meaning with 10 Fe host atoms in the training of the fine-tuned model, and the rest of 20 Fe host atoms in the holdout set.

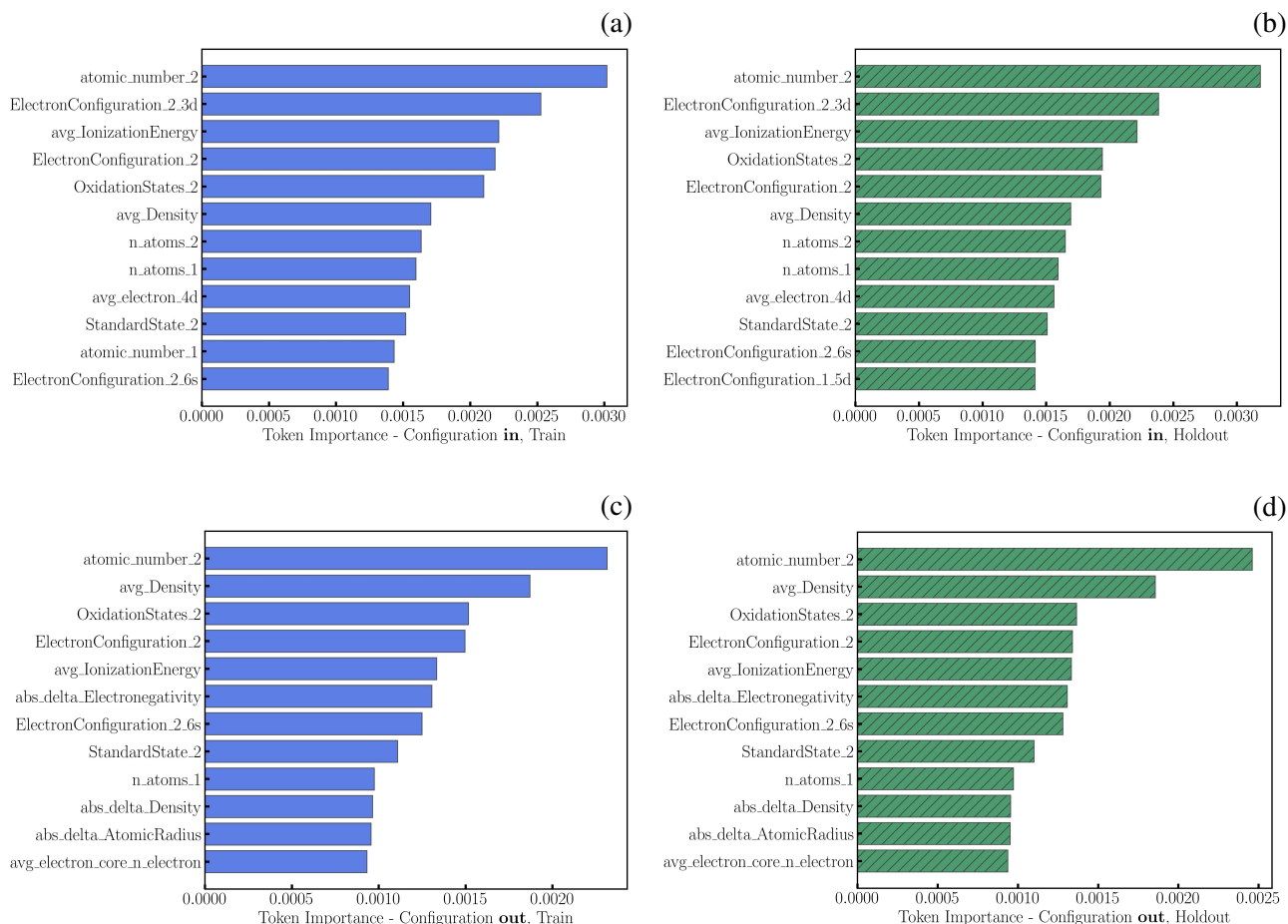


Figure 7. Token importance for both configurations, comparing fine-tune and holdout. Dominant contributors include electron-configuration descriptors and periodic-table properties, as well as engineered features (absolute differences and averages). The stability of ranking across splits supports physically meaningful representations and transferred learning. (a,b) Training and holdout sets for *in* configuration, and (c,d) Training and holdout sets for *out* configuration. Similar token importance ranking in both the training and holdout sets, as well as distributed importance across different features, are indications of a reliably fine-tuned model.

Author information

Contributions

JMTP and MJP performed the DFT calculations, analyzed the data, and prepared the figures. MS, ORF, and CRCR curated the data and developed the PAI algorithm code. CRCR and MS supervised and refined the ML coding for the final version. JMTP, MJP, MS, and CRCR drafted the initial version of the manuscript. ACD, DGS, WW, and RA contributed to data curation and critically reviewed the manuscript findings. MJP and CRCR jointly supervised the project and led manuscript preparation. All authors contributed to the writing and revision of the manuscript.

Ethics declarations

Competing Interests

All authors declare no competing financial or non-financial interests.

Supporting Information: Teaching a Transformer to Think Like a Chemist: Predicting Nanocluster Stability

João Marcos T. Palheta,[†] Octavio Rodrigues Filho,[†] Mohammad Soleymanibrojeni,[‡]
Alexandre Cavaleiro Dias,[¶] Diego Guedes-Sobrinho,[§] Wolfgang Wenzel,^{||} Roland
Aydin,[‡] Celso R. C. Rêgo,^{*,||} and Maurício Jeomar Piotrowski[†]

[†]*Department of Physics, Federal University of Pelotas, PO Box 354, Pelotas, RS, 96010-900,
Brazil*

[‡]*Hamburg University of Technology, Hamburg, Germany*

[¶]*Institute of Physics and International Center of Physics, University of Brasília, Brasília
70919-970, DF, Brazil*

[§]*Quantum Chemistry and Thermodynamic Materials Group – Q²M, Chemistry Department,
Federal University of Paraná, Curitiba, PR, 81531-980, Brazil*

^{||}*Karlsruhe Institute of Technology (KIT), Institute of Nanotechnology,
Eggenstein-Leopoldshafen, Germany*

E-mail: celso.rego@kit.edu

Contents

1	Convergence Tests	S3
2	Proofs for the Energetic Equations	S5
3	Properties of TM Bulks	S7
4	Properties of Nanoclusters $X_{12}TM$	S8
4.1	Complete Set of Optimized Nanoclusters	S8
4.2	Effective Coordination Number and Average Bond Length	S12
4.3	Vibrational Frequencies	S14
4.4	Formation Energy	S15
4.5	Binding and Formation Energies for Fe-based Systems	S16
4.6	Geometric Properties for Fe-based Systems	S17
4.7	Center of Gravity of the Occupied d -states	S18
4.8	HOMO-LUMO Gap	S19
4.9	Validation for Fe-based Systems	S21
5	Predictive Artificial Intelligence	S23
	References	S24

1 Convergence Tests

We performed methodological tests to obtain the main parameters used in our calculations.

In Tables S1, S2, S3, and S4 we show some properties (the relative total energy – $\Delta E_{\text{tot}} = E_{\text{tot}}^{\text{A}} - E_{\text{tot}}^{\text{B}}$, average bond length – d_{av}^{A} and d_{av}^{B} , effective coordination number – ECN^{A} and ECN^{B} , and total magnetic moment – $m_{\text{tot}}^{\text{A}}$ and $m_{\text{tot}}^{\text{B}}$) for the arbitrary cases of the Hf_{13} (named A) and Hf_{12}Hg (named B) as a function of the parameters.

Table S1: *Box size*: the relative total energy (ΔE_{tot}), average bond length (d_{av}^{A} and d_{av}^{B}), effective coordination number (ECN^{A} and ECN^{B}), and total magnetic moment ($m_{\text{tot}}^{\text{A}}$ and $m_{\text{tot}}^{\text{B}}$).

Box Size (Å)	ΔE_{tot} (eV)	ECN^{A}	ECN^{B}	d_{av}^{A} (Å)	d_{av}^{B} (Å)	$m_{\text{tot}}^{\text{A}}$ (μ_{A})	$m_{\text{tot}}^{\text{B}}$ (μ_{B})
12	-12.4168	6.3730	6.3914	2.2093	2.9398	6.0000	2.0000
14	-12.4939	6.3722	6.3941	2.9461	2.9382	6.0000	2.0000
16	-12.4236	6.3730	6.3913	2.9457	2.9389	6.0000	2.0000
18	-12.4608	6.3731	6.3945	2.9457	2.9396	6.0000	2.0000
20	-12.4599	6.3729	6.3945	2.9450	2.9398	6.0000	2.0000
22	-12.4337	6.3702	6.3946	2.9441	2.9398	6.0000	2.0000
24	-12.4271	6.3724	6.3913	2.9456	2.9389	6.0000	2.0000

Table S2: *ENCUT = energy-cut-off: $X \cdot \text{ENMAX}$* , where the $\text{ENMAX} = 282.964$ eV, the relative total energy (ΔE_{tot}), average bond length (d_{av}^{A} and d_{av}^{B}), effective coordination number (ECN^{A} and ECN^{B}), and total magnetic moment ($m_{\text{tot}}^{\text{A}}$ and $m_{\text{tot}}^{\text{B}}$).

$X \cdot \text{ENMAX}$ (eV)	ΔE_{tot} (eV)	ECN^{A}	ECN^{B}	d_{av}^{A} (Å)	d_{av}^{B} (Å)	$m_{\text{tot}}^{\text{A}}$ (μ_{A})	$m_{\text{tot}}^{\text{B}}$ (μ_{B})
0.500	-10.4053	6.1700	6.3569	2.8254	2.7650	6.0000	2.0000
0.750	-12.3137	6.3542	6.3587	2.9876	2.9451	6.0000	2.0000
1.000	-12.3937	6.3776	6.3901	2.9423	2.9341	6.0000	2.0000
1.125	-12.4290	6.3733	6.3923	2.9461	2.9402	6.0000	2.0000
1.250	-12.4594	6.3731	6.3945	2.9459	2.9398	6.0000	2.0000
1.500	-12.4279	6.3730	6.3920	2.9456	2.9388	6.0000	2.0000
2.000	-12.4362	6.3729	6.3913	2.9450	2.9383	6.0000	2.0000

Table S3: EDIFF = energy difference: electronic convergence, the relative total energy (ΔE_{tot}), average bond length (d_{av}^A and d_{av}^B), effective coordination number (ECN^A and ECN^B), and total magnetic moment (m_{tot}^A and m_{tot}^B).

EDIFF (eV)	ΔE_{tot} (eV)	ECN^A	ECN^B	d_{av}^A (Å)	d_{av}^B (Å)	m_{tot}^A (μ_A)	m_{tot}^B (μ_B)
10^{-2}	-12.4600	6.3729	6.3946	2.9454	2.9386	6.0000	2.0000
10^{-3}	-12.4598	6.3740	6.3943	2.9457	2.9385	6.0000	2.0000
10^{-4}	-12.4600	6.3726	6.3939	2.9459	2.9394	6.0000	2.0000
10^{-5}	-12.4600	6.3731	6.3946	2.9457	2.9391	6.0000	2.0000
10^{-6}	-12.4599	6.3729	6.3945	2.9450	2.9398	6.0000	2.0000
10^{-7}	-12.4600	6.3731	6.3947	2.9456	2.9398	6.0000	2.0000
10^{-8}	-12.4599	6.3729	6.3946	2.9450	2.9398	6.0000	2.0000

Table S4: EDIFFG = energy difference gradient: ionic convergence, the relative total energy (ΔE_{tot}), average bond length (d_{av}^A and d_{av}^B), effective coordination number (ECN^A and ECN^B), and total magnetic moment (m_{tot}^A and m_{tot}^B).

EDIFFG (eV/Å)	ΔE_{tot} (eV)	ECN^A	ECN^B	d_{av}^A (Å)	d_{av}^B (Å)	m_{tot}^A (μ_A)	m_{tot}^B (μ_B)
-0,100	-12.4265	6.3729	6.3918	2.9450	2.9389	6.0000	2.0000
-0,050	-12.4265	6.3729	6.3918	2.9450	2.9389	6.0000	2.0000
-0,025	-12.4265	6.3729	6.3918	2.9450	2.9389	6.0000	2.0000
-0,020	-12.4265	6.3729	6.3918	2.9450	2.9389	6.0000	2.0000
-0,015	-12.4266	6.3731	6.3918	2.9457	2.9389	6.0000	2.0000
-0,010	-12.4266	6.3732	6.3918	2.9461	2.9389	6.0000	2.0000
-0,005	-12.4265	6.3731	6.3913	2.9460	2.9389	6.0000	2.0000

2 Proofs for the Energetic Equations

Below, we have provided the equation definitions used in our energetic analysis for the $X_{12}TM^{in}$ and $X_{12}TM^{out}$ systems, where X are the Ti, Zr, Hf, and Fe atoms and TM the transition metal atoms, as well as, the proofs that demonstrate the equivalence of the equations for the alternative and usual binding.

Binding Energy

The usual equation of the binding energy per atom for all systems is given by,

$$E_b = \frac{E_{tot}^{X_{12}TM} - 12E_{tot}^{X-free} - E_{tot}^{TM-free}}{13}, \quad (1)$$

where $E_{tot}^{X_{12}TM}$ is the total energy of the systems, and E_{tot}^{X-free} is total energy of the Ti, Zr and Hf free atoms, $E_{tot}^{TM-free}$ is total energy of the TM free atom.

Interaction Energy

The equation gives the interaction energy,

$$E_{int} = E_{tot}^{X_{12}TM} - E_{tot}^{X_{12}-frozen} - E_{tot}^{TM-free}, \quad (2)$$

where $E_{tot}^{X_{12}-frozen}$ is the total energy of the frozen system, X_{12} , without TM atoms and $E_{tot}^{TM-free}$ is the total energy of the TM free atom.

Relative Distortion Energy

The relative distortion energy is the energy differences per atom between the frozen and relaxed systems. The distortion energy between the frozen $E_{tot}^{X_{12}-frozen}$ system, without TM atom, and the optimized $E_{tot}^{X_{12}}$ nanocluster is given by,

$$\Delta E_{dist} = \frac{E_{tot}^{X_{12}-frozen} - E_{tot}^{X_{12}}}{12}, \quad (3)$$

Alternative Equation for the Binding Energy

Is it possible to write an alternative equation for the binding energy of the nanocluster systems? It is composed of the binding energies of the unprotected E_b^u nanoclusters, without TM atom, given by,

$$E_b^u = \frac{E_{tot}^{X_{12}} - 12E_{tot}^{X-livre}}{12}, \quad (4)$$

with the interaction energy (2) and the relative distortion energy (3) of systems, we can write an alternative equation for binding energy as,

$$E_b = \frac{12E_b^u + E_{int} + 12\Delta E_{dist}}{13}, \quad (5)$$

To prove the equivalence of both equations, (1) and (5), we replace equations (4), (2) and (3) in equation (5), consequently,

$$E_b = \frac{12 \left(\frac{E_{tot}^{X_{12}} - 12E_{tot}^{X-free}}{12} \right) + E_{tot}^{X_{12}TM} - E_{tot}^{X_{12}-frozen} - E_{tot}^{TM-free} + 12 \left(\frac{E_{tot}^{X_{12}-frozen} - E_{tot}^{X_{12}}}{12} \right)}{13}, \quad (6)$$

after simplifications,

$$E_b = \frac{E_{tot}^{X_{12}} - 12E_{tot}^{X-free} + E_{tot}^{X_{12}TM} - E_{tot}^{X_{12}-frozen} - E_{tot}^{TM-free} + E_{tot}^{X_{12}-frozen} - E_{tot}^{X_{12}}}{13}, \quad (7)$$

which results in equation,

$$E_b = \frac{E_{tot}^{X_{12}TM} - 12E_{tot}^{X-free} - E_{tot}^{TM-free}}{13}, \quad (8)$$

Therefore, both equations are equivalent and must produce the same results.

3 Properties of TM Bulks

To limit the scope, we also calculate bulk properties of the Transition Metals. The Properties such as cohesive energy (E_{coh}) and atomic radius (R_{TM}) for the lowest-energy unit cell are shown in the graphic below, compared with experimental data.

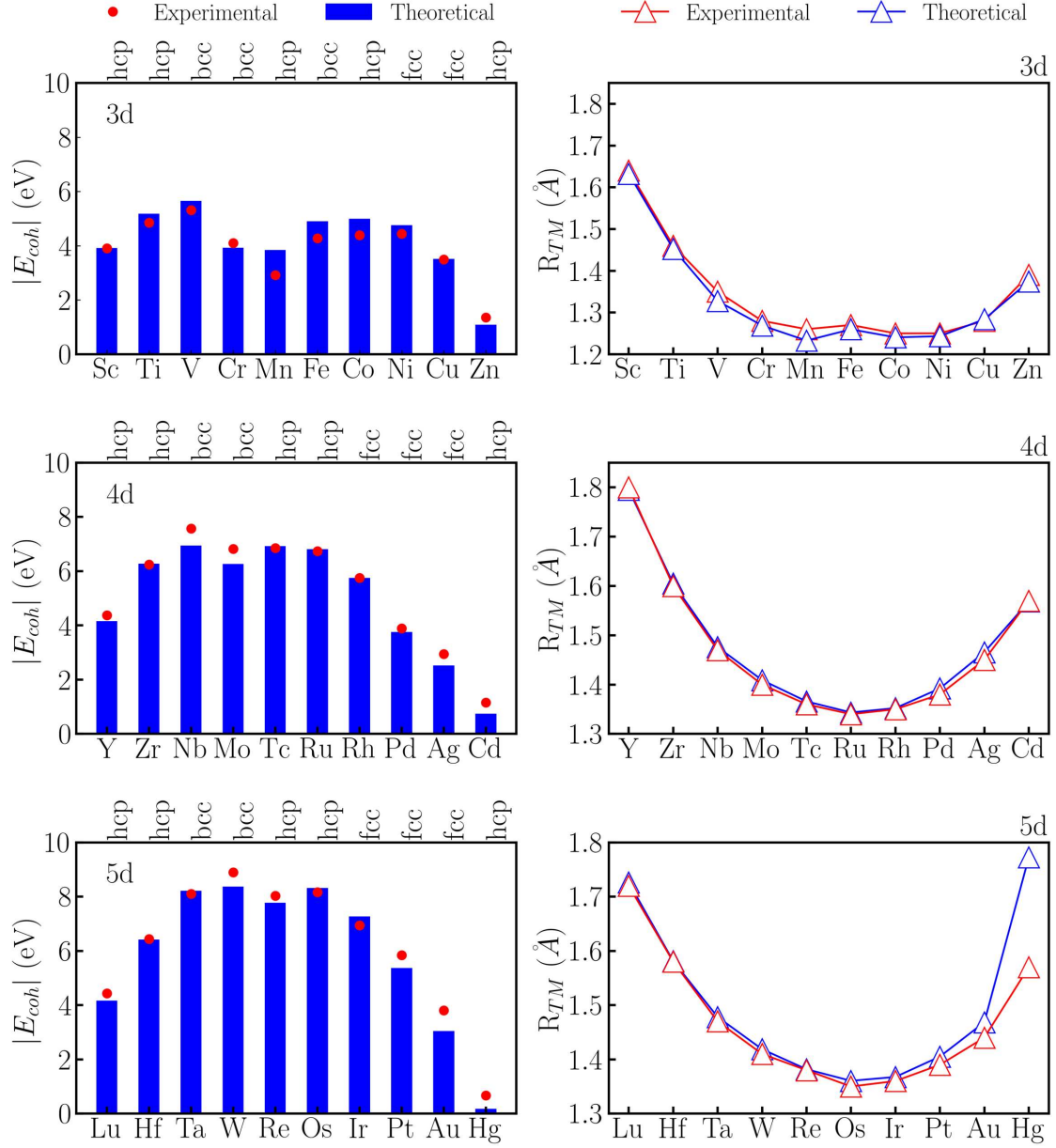


Figure S1: The bulks properties, E_{coh} and R_{TM} , for TMs as function of the atomic number. In color blue, the theoretical data, and in color red, the experimental data.

4 Properties of Nanoclusters $X_{12}TM$

4.1 Complete Set of Optimized Nanoclusters

Below, the optimized 13-atom nanoclusters are shown in Figures:

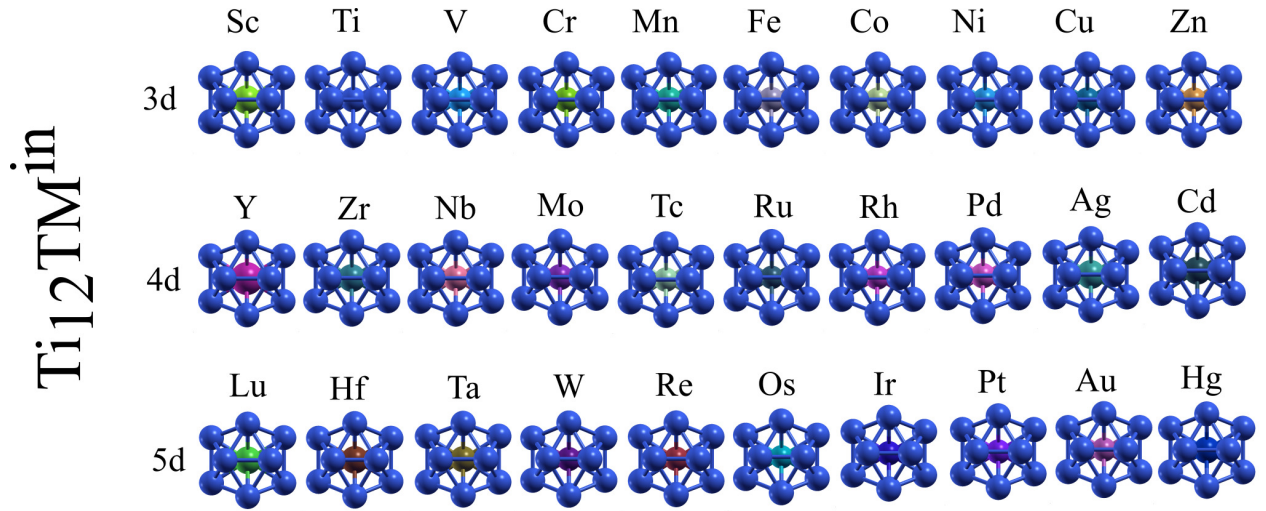


Figure S2: Optimized 13-atom nanoclusters $Ti_{12}TM^{in}$ in geometric structure icosahedral (ICO) as in Periodic Table.

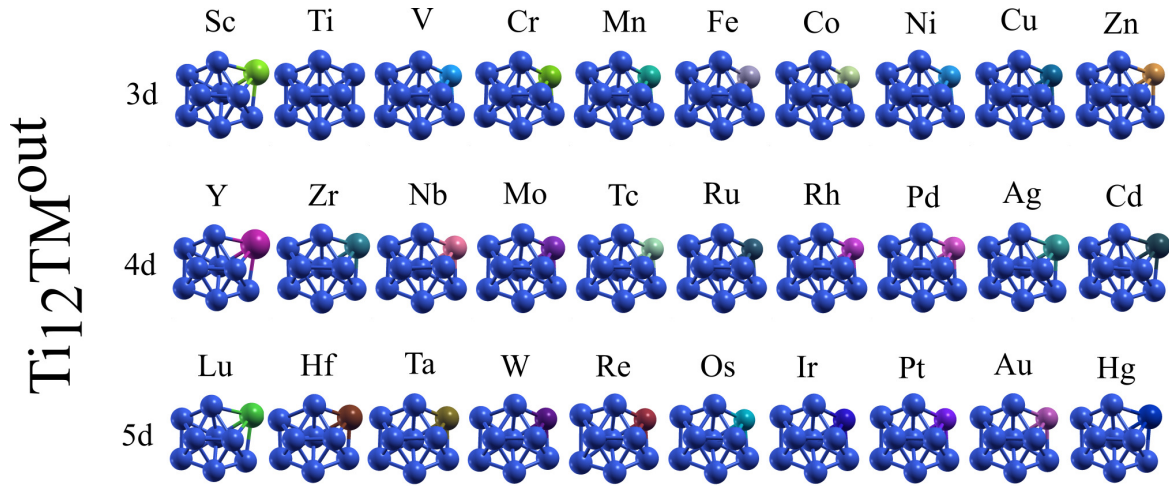


Figure S3: Optimized 13-atom nanoclusters $Ti_{12}TM^{out}$ in geometric structure icosahedral (ICO) as in Periodic Table.

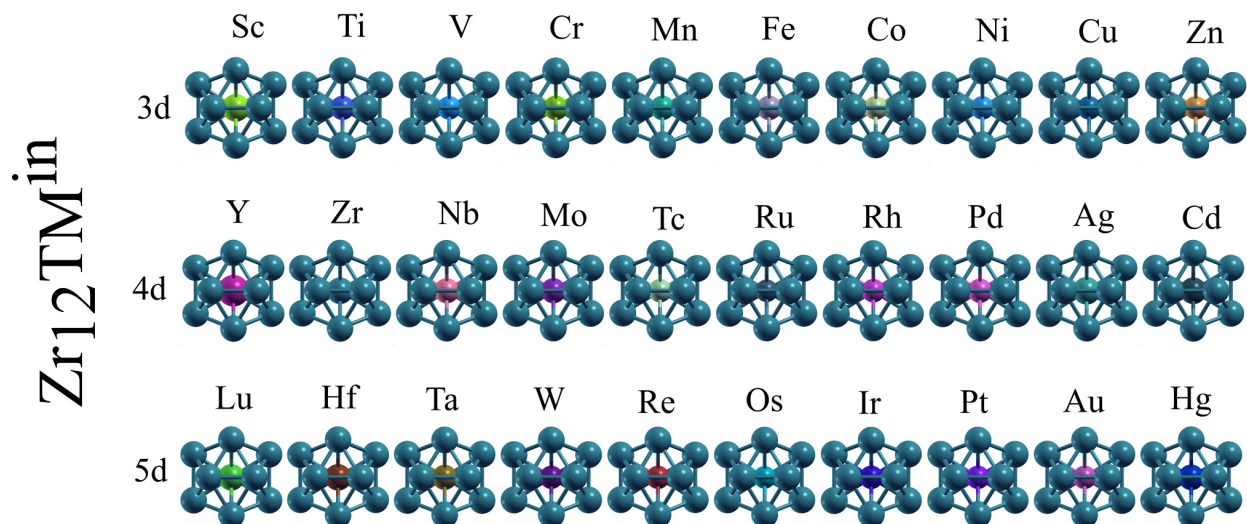


Figure S4: Optimized 13-atom nanoclusters $\text{Zr}_{12}\text{TM}^{\text{in}}$ in geometric structure icosahedral (ICO) as in Periodic Table.

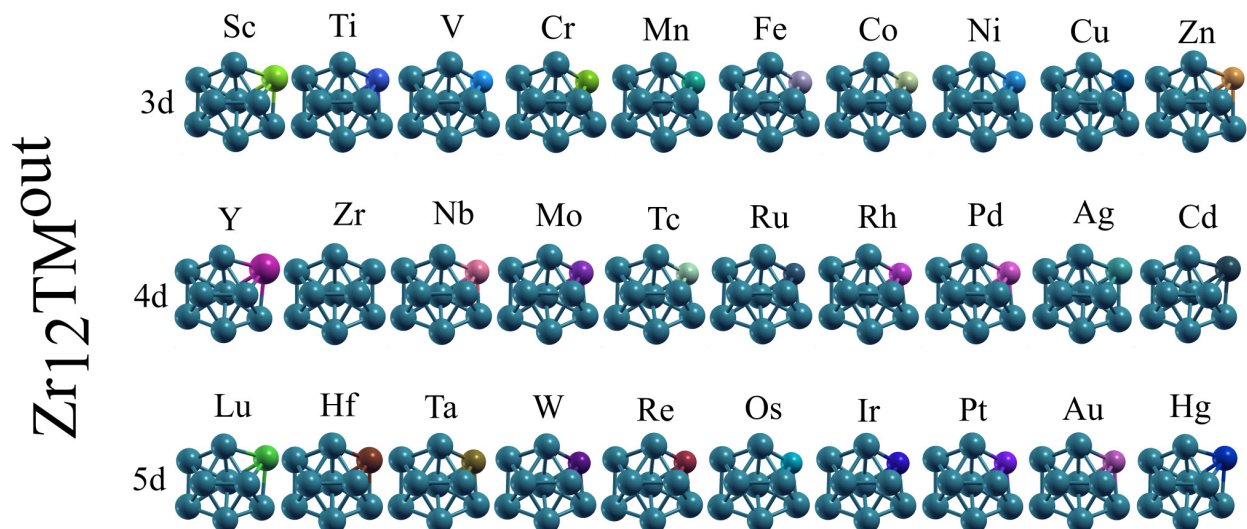


Figure S5: Optimized 13-atom nanoclusters $\text{Zr}_{12}\text{TM}^{\text{out}}$ in geometric structure icosahedral (ICO) as in Periodic Table.

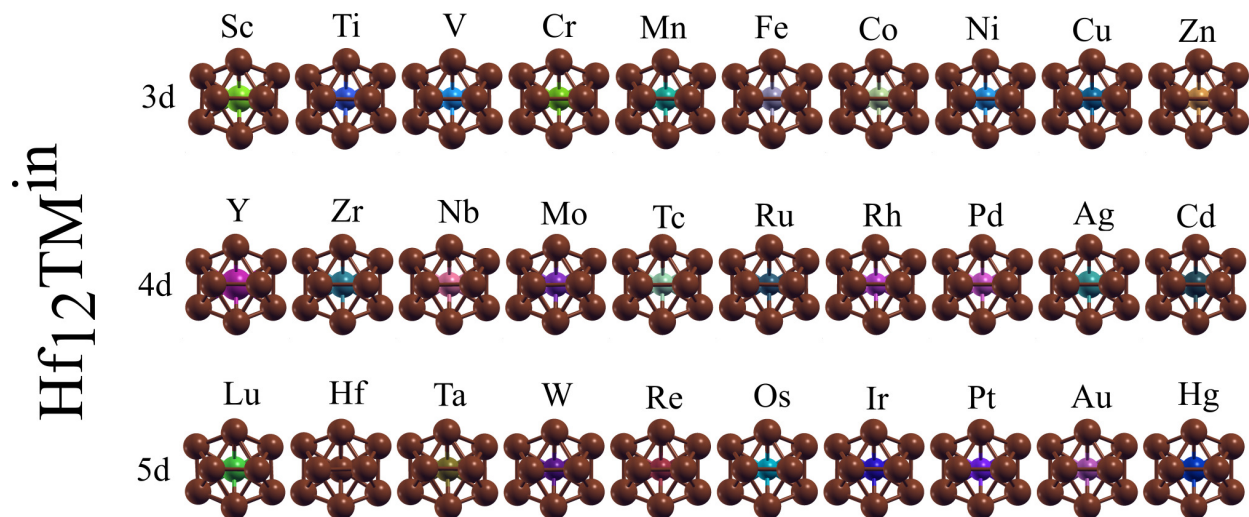


Figure S6: Optimized 13-atom nanoclusters $\text{Hf}_{12}\text{TM}^{\text{in}}$ in geometric structure icosahedral (ICO) as in Periodic Table.

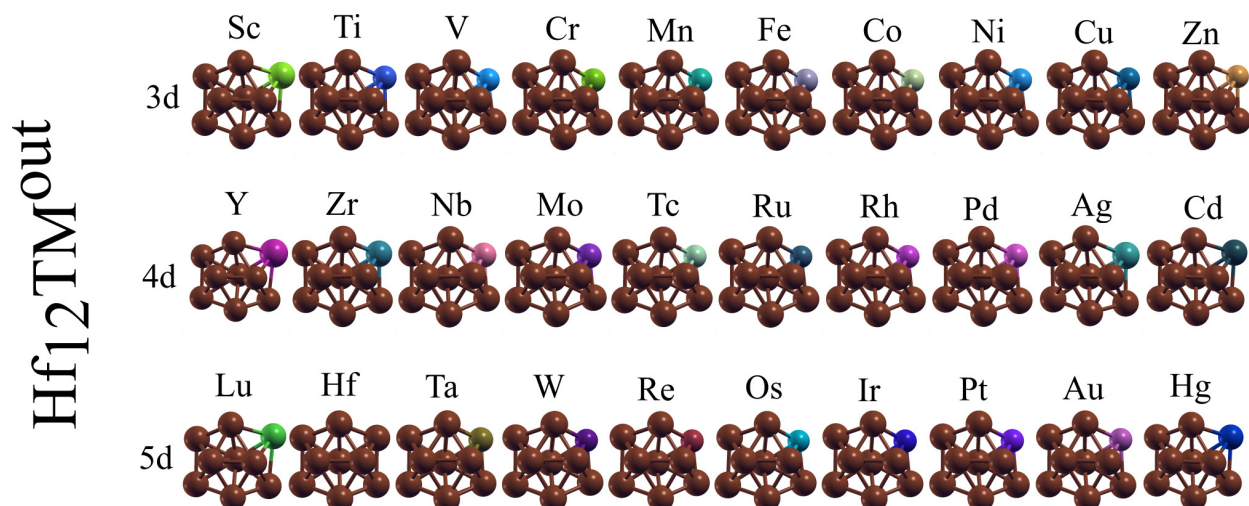


Figure S7: Optimized 13-atom nanoclusters $\text{Hf}_{12}\text{TM}^{\text{out}}$ in geometric structure icosahedral (ICO) as in Periodic Table.

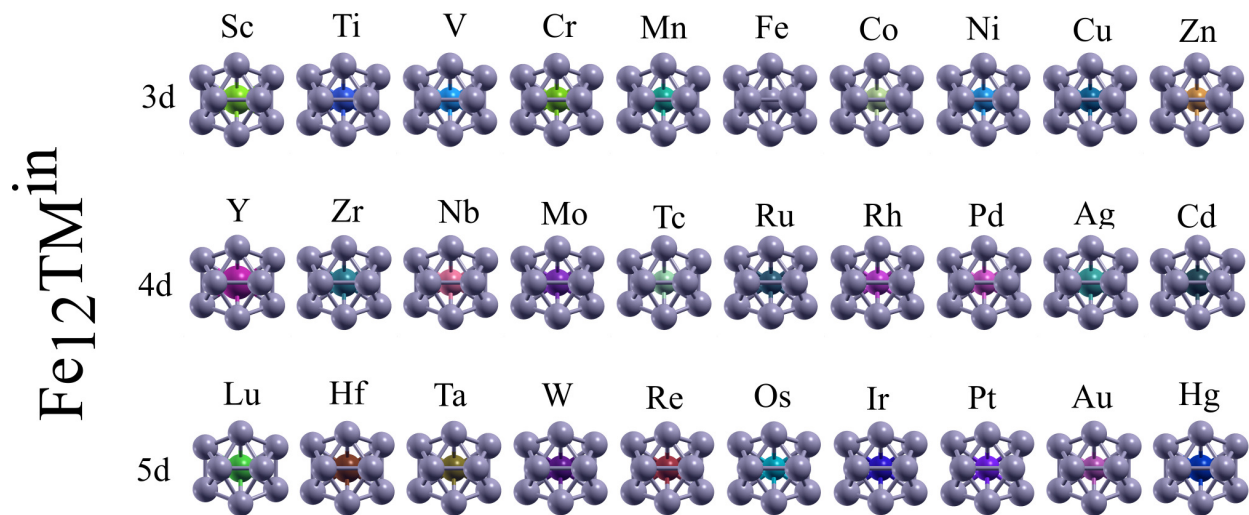


Figure S8: Optimized 13-atom nanoclusters $\text{Fe}_{12}\text{TM}^{\text{in}}$ in geometric structure icosahedral (ICO) as in Periodic Table.

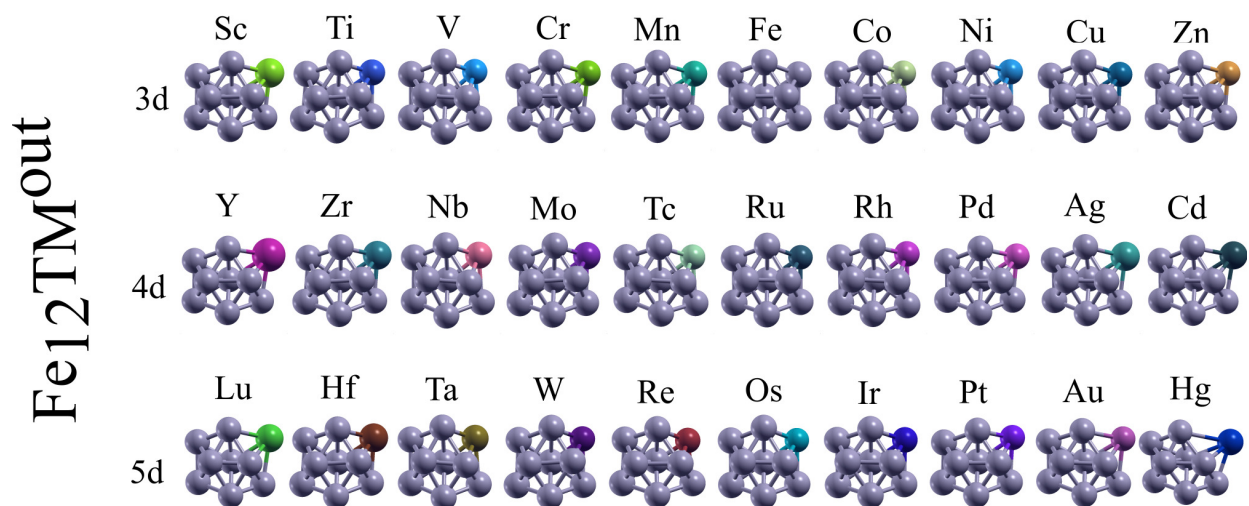


Figure S9: Optimized 13-atom nanoclusters $\text{Fe}_{12}\text{TM}^{\text{out}}$ in geometric structure icosahedral (ICO) as in Periodic Table.

4.2 Effective Coordination Number and Average Bond Length

The effective coordination number (ECN) and average bond length (d_{av}) were employed to analyze the X_{12} TM ICO-like nanoclusters.^{1,2} The usual coordination number (CN) defines the number of neighboring atoms to the atom j and attributes the same weight for all atoms i closer to the coordination sphere. This method does not account for bonds that are longer than the cutting radius; therefore, it does not accurately represent distorted structures. In the ECN, the cutting radius is not necessary, since the weight is attributed to every atom i according to the bond length. In this sense, the atom j is surrounded by atoms i with different distances and different weights. Consequently, the weight depends directly on the d_{av} , with values smaller (larger) than d_{av} contributing with a weight larger (smaller). Thus, ECN can represent distorted structures while still yielding the correct value for bulk systems. The ECN_i is obtained by

$$ECN_i = \sum_j \exp \left[1 - \left(\frac{d_{ij}}{d_{av}^i} \right)^6 \right], \quad (9)$$

where d_{ij} is the atomic distance between i and j , d_{av} is defined as

$$d_{av}^i = \frac{\sum_j d_{ij} \exp \left[1 - \left(\frac{d_{ij}}{d_{av}^i} \right)^6 \right]}{\sum_j \exp \left[1 - \left(\frac{d_{ij}}{d_{av}^i} \right)^6 \right]}. \quad (10)$$

The d_{av}^i is obtained self-consistently, where $|d_{av}^i(new) - d_{av}^i(old)| < 0.0001$ is the convergence parameter. The initial value is defined as the smallest bond length between the atom i and all j atoms (d_{min}^i). The final value of d_{av}^i is obtained after a few interactions. The average ECN and d_{av} values are obtained by

$$ECN = \frac{1}{N} \sum_{i=1}^N ECN_i, \quad (11)$$

$$d_{av} = \frac{1}{N} \sum_{i=1}^N d_{av}^i, \quad (12)$$

where N is the total number of atoms in the nanocluster.

4.3 Vibrational Frequencies

Vibrational frequencies were calculated for systems Ti, Zr, and Hf. This result shows that systems are local minima with honest and positive frequencies.

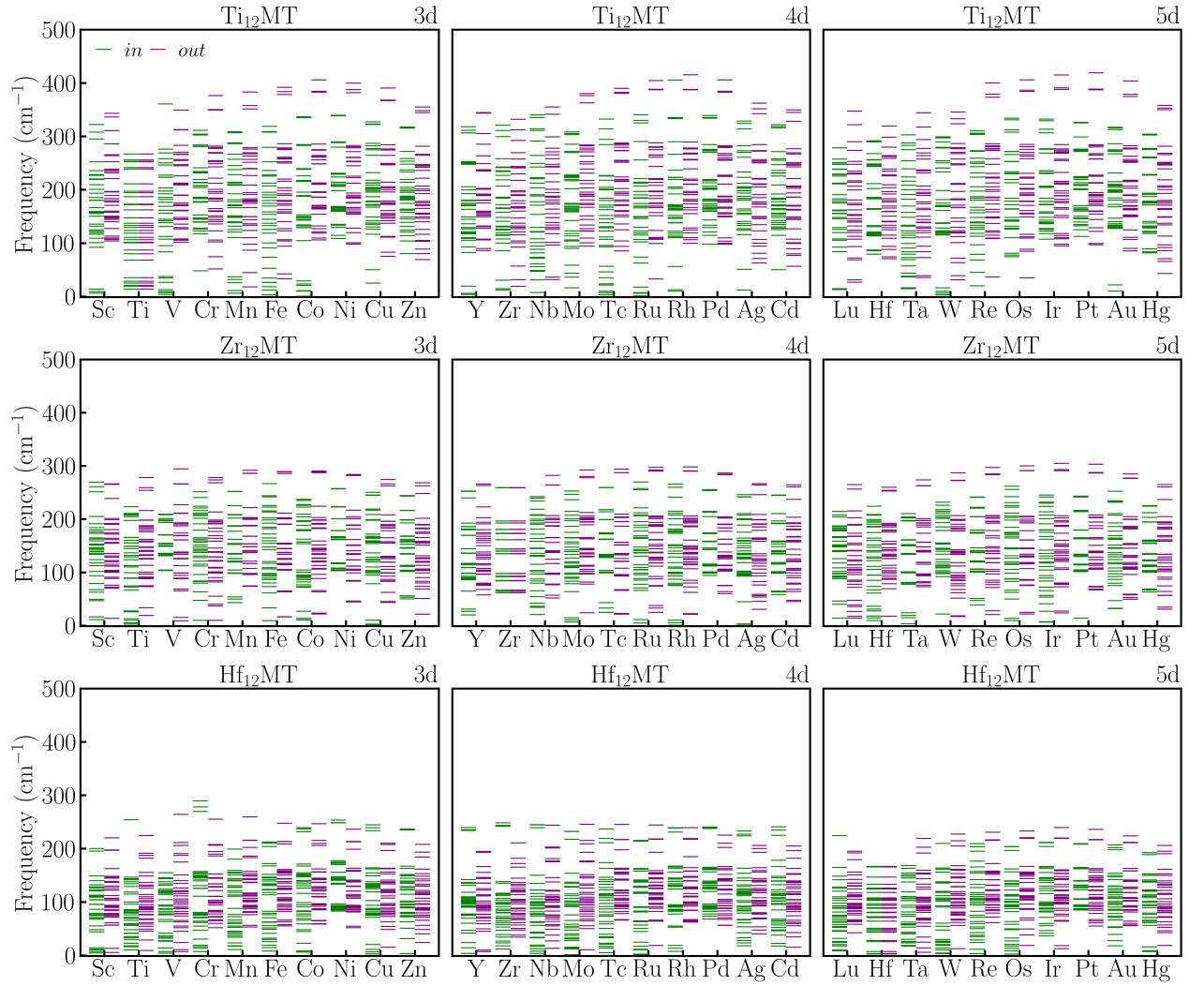


Figure S10: Vibrational frequencies for nanoclusters of Ti, Zr, and Hf as a function of the atomic number. In purple are systems *in* and green are systems *out*.

4.4 Formation Energy

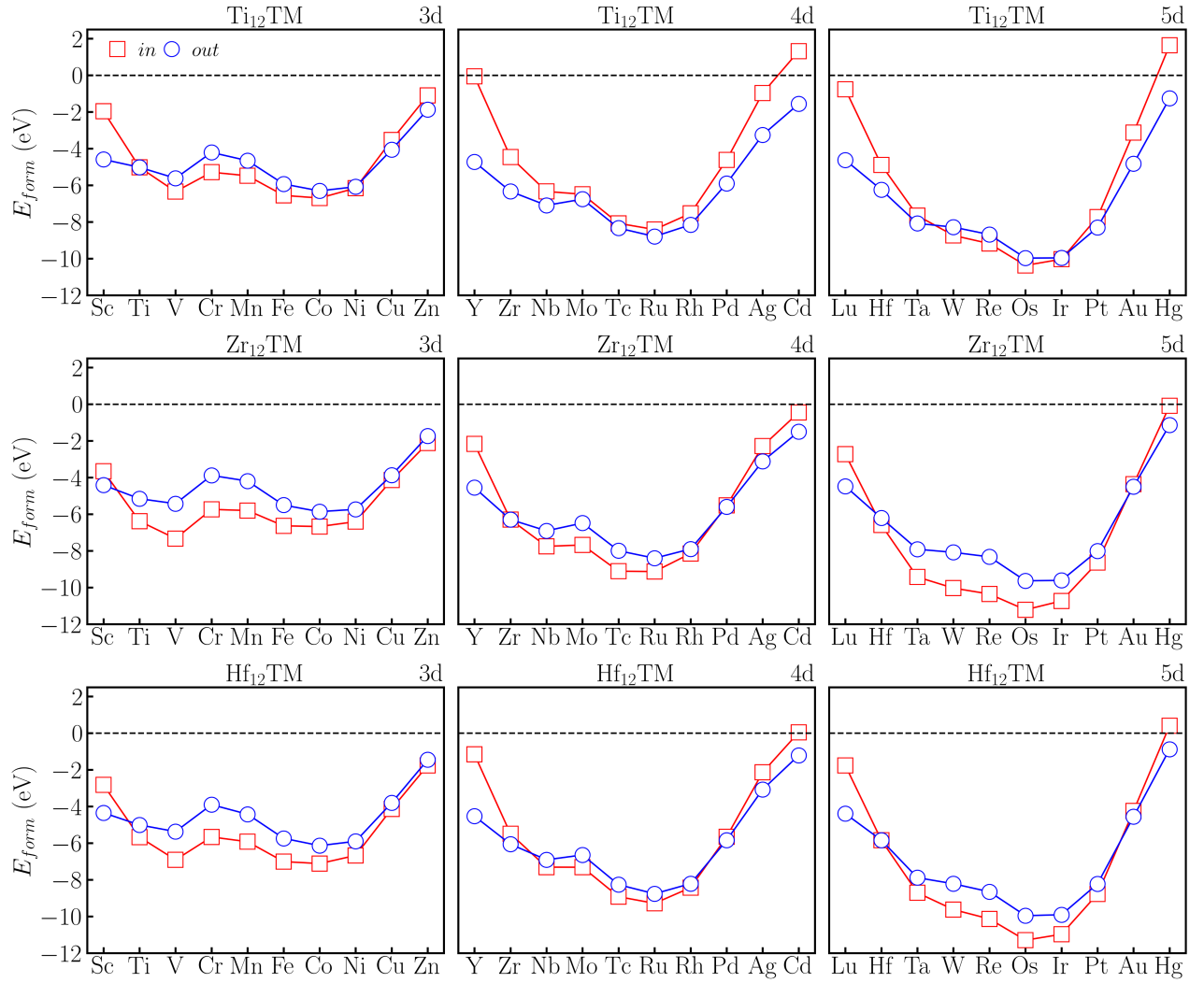


Figure S11: Formation Energy for nanoclusters of Ti, Zr, and Hf as a function of the atomic number. In color, blues are the *in* systems, and blue is the *out* system.

4.5 Binding and Formation Energies for Fe-based Systems

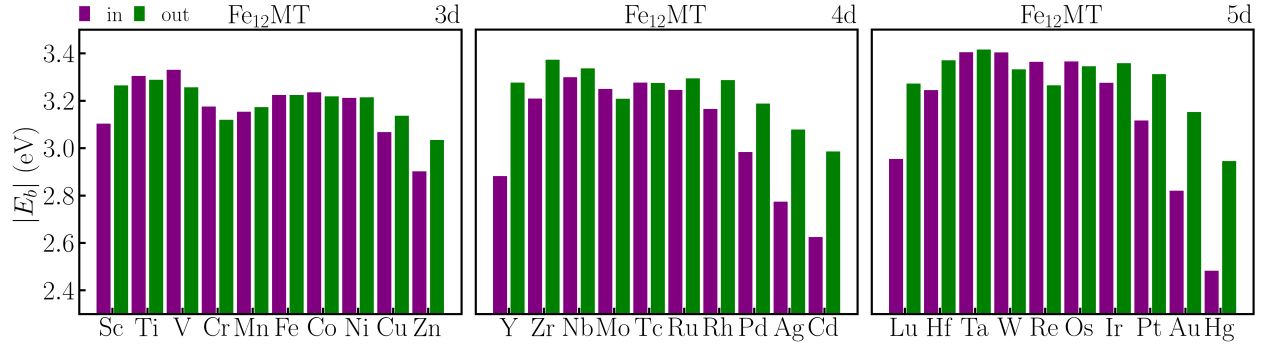


Figure S12: Binding energy for nanoclusters of Fe as a function of the atomic number. In color purple are the systems *in* and green the *out* systems.

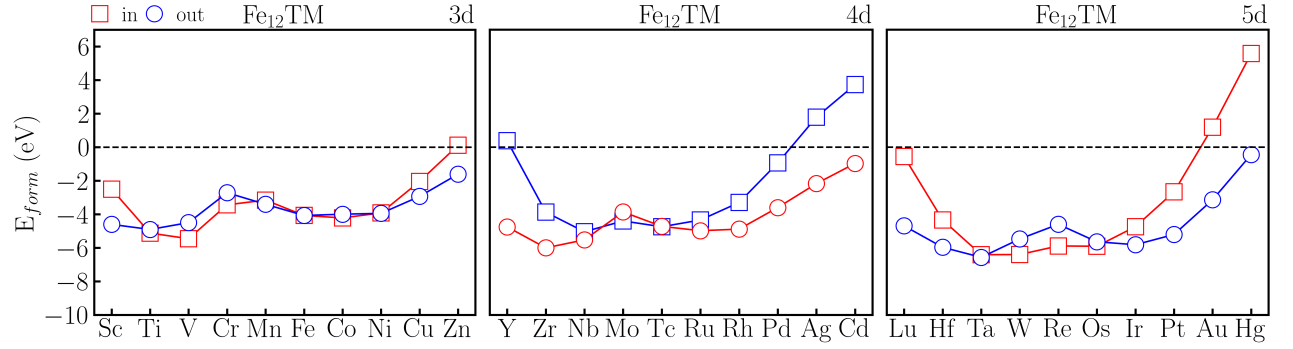


Figure S13: Formation Energy for nanoclusters of Fe as a function of the atomic number. In color, reds are the *in* systems, and blue is the *out* system.

4.6 Geometric Properties for Fe-based Systems

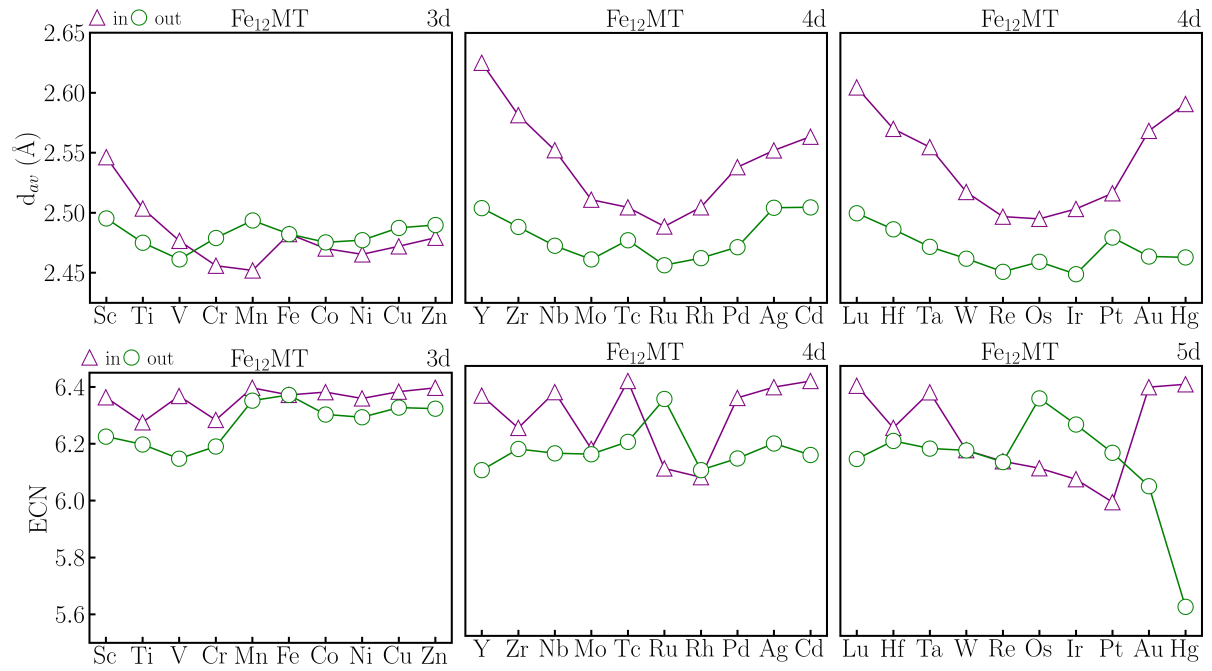


Figure S14: d_{av} and ECN for nanoclusters of Fe. In color purple are the in systems, and in green are the out systems.

4.7 Center of Gravity of the Occupied d -states

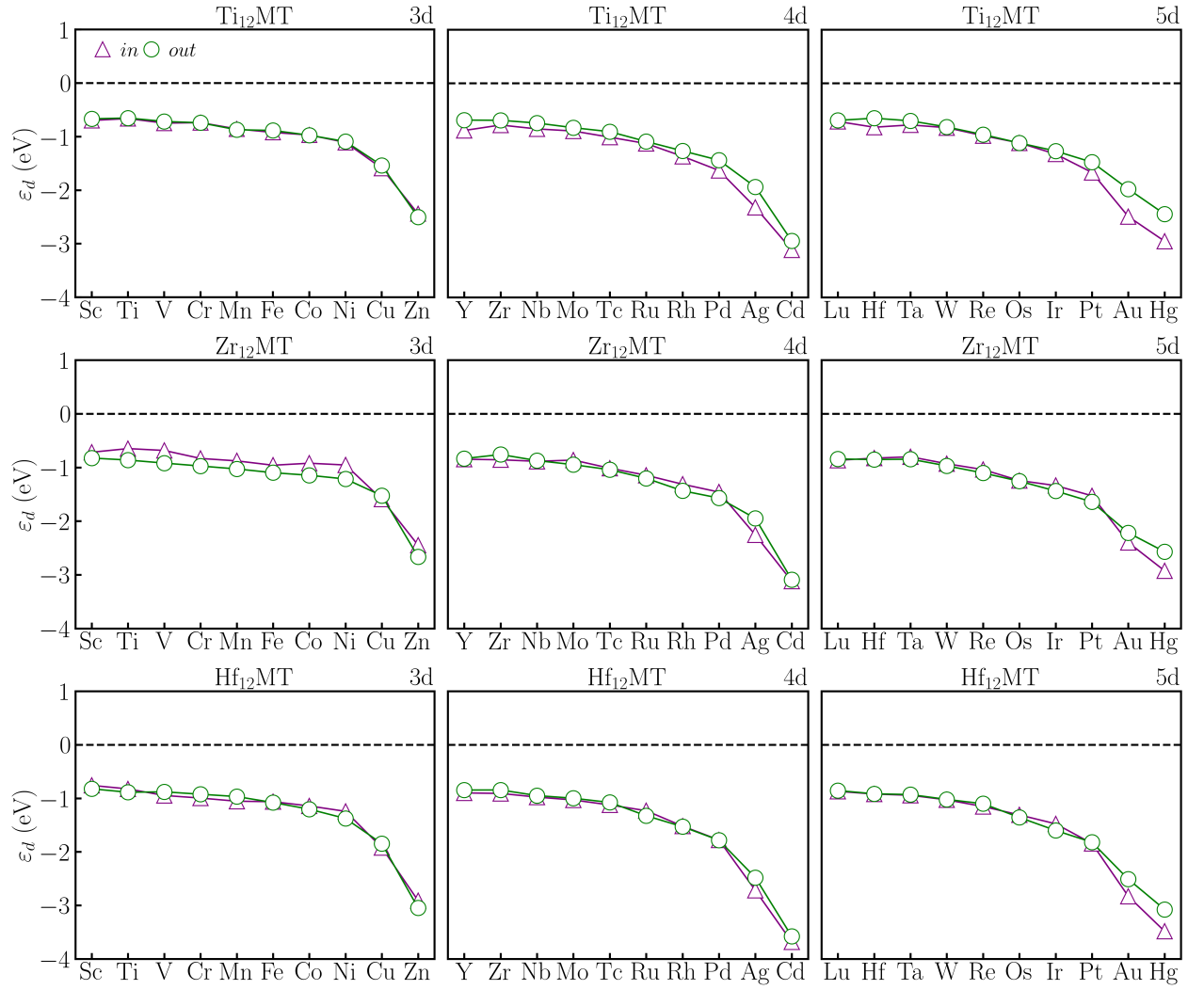


Figure S15: Center of gravity of the occupied d states for nanoclusters of Ti, Zr, and Hf as a function of the atomic number. In purple are the *in* systems, and in green the *out* systems.

4.8 HOMO-LUMO Gap

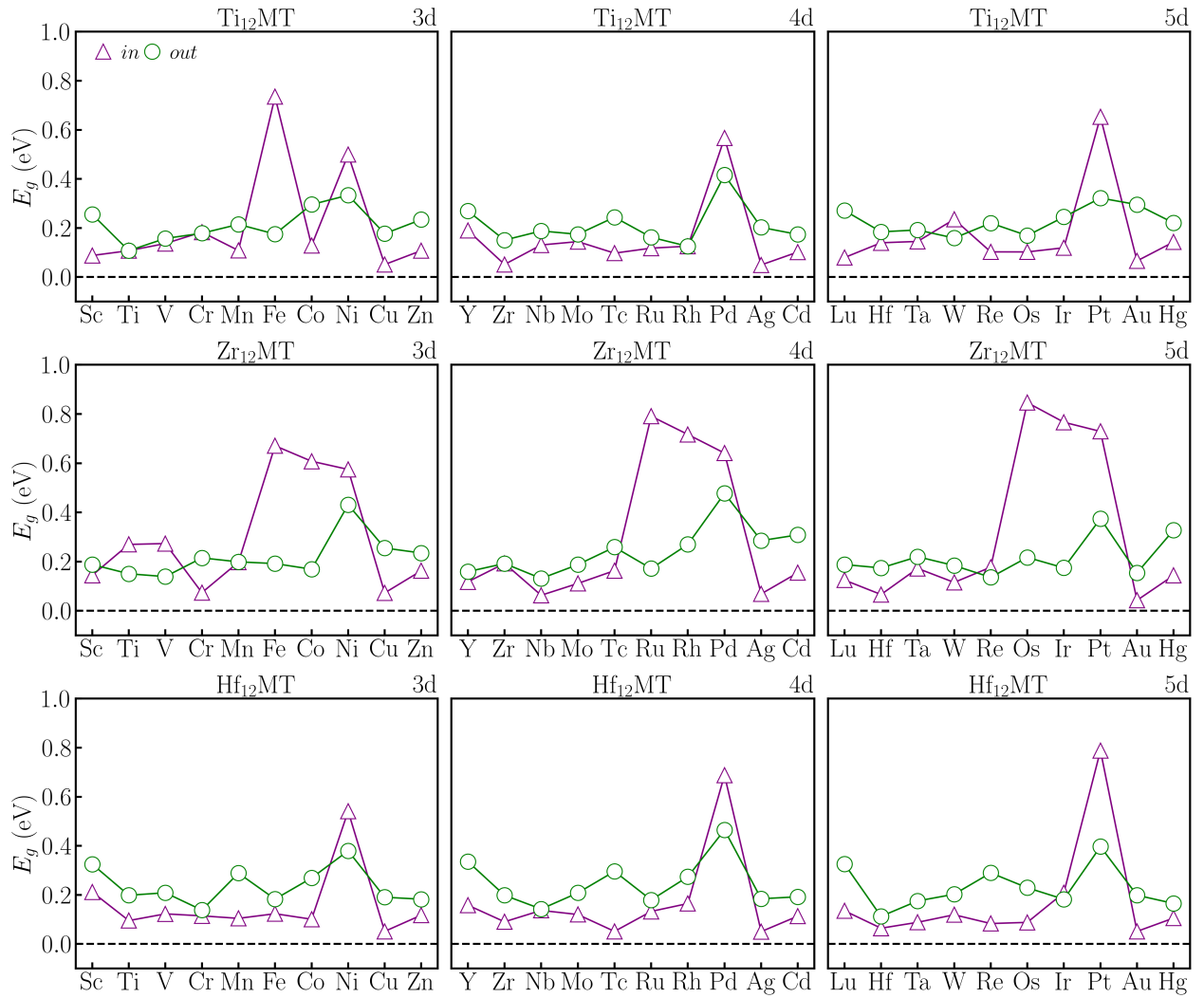


Figure S16: HOMO-LUMO gap for nanoclusters of Ti, Zr, and Hf as a function of the atomic number. In purple are the *in* systems, and in green the *out* systems.

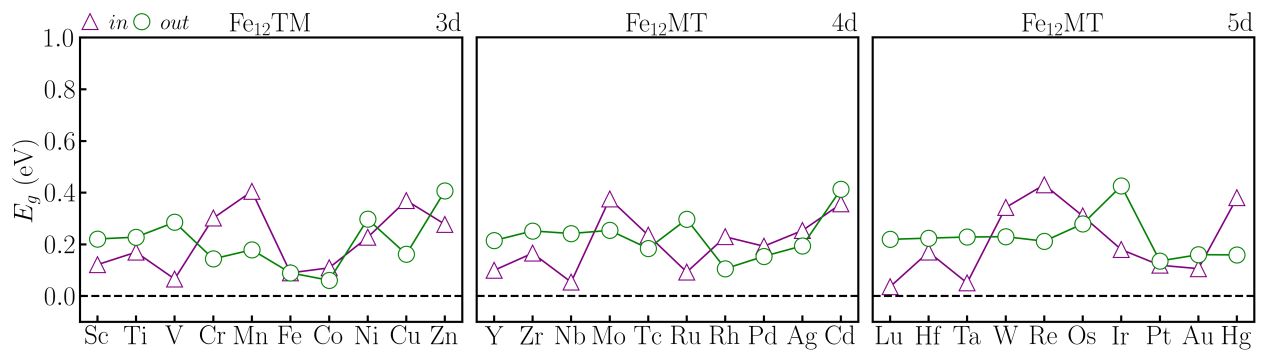


Figure S17: HOMO-LUMO gap for nanoclusters of Fe as a function of the atomic number. In purple are the *in* systems, and in green the *out* systems.

4.9 Validation for Fe-based Systems

To critically validate the generality of our trends and the predictive capability of our ML model, we selected Fe_{12}TM nanoclusters as a test case and analyzed both *in* and *out* configurations. The choice of Fe as a host system is particularly justified given the robust literature indicating that Fe_{13} adopts an ICO-like geometry as its ground state, consistent with earlier studies.^{3,4}

Below, we present a direct comparison between DFT-calculated and ML-predicted E_{form} for Fe_{12}TM nanoclusters across the 3d, 4d, and 5d TM series. Both *in* and *out* configurations are depicted, with excellent agreement between ML and DFT predictions, even for dopants in the unseen (validation) set, highlighting again the model’s predictive capacity and transferability. The accurate reproduction of trends of stability, i.e., the increased stabilization of *in* configurations towards mid-series TMs (e.g., Mn, Fe, Co) and the trend to *out* configurations at the series boundaries, highlights the model’s capability to capture small electronic and structural effects.

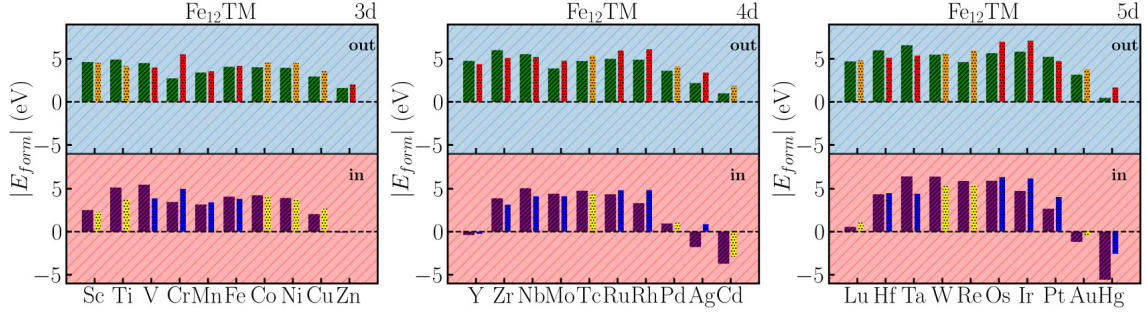


Figure S18: Comparison of DFT (solid bars) and (seen and unseen) ML-predicted (hatched bars) formation energies ($|E_{\text{form}}|$) for Fe_{12}TM nanoclusters with 3d, 4d, and 5d TM dopants. Top and bottom panels show *out* and *in* configurations, respectively. The high consistency across series and configurations, including unseen validation cases, corroborates the reliability and transferability of the ML model.

Consequently, our results for Fe_{12}TM nanoclusters also strongly confirm the generalizability of the group IV-based nanocluster structural and energetic design principles developed. In addition, the even better agreement between ML predictions and DFT calculations in Figure S18 further confirms the predictive capability and universal

applicability of our model across hosts with varying atomic sizes, electronic structures, and magnetic properties.

5 Predictive Artificial Intelligence

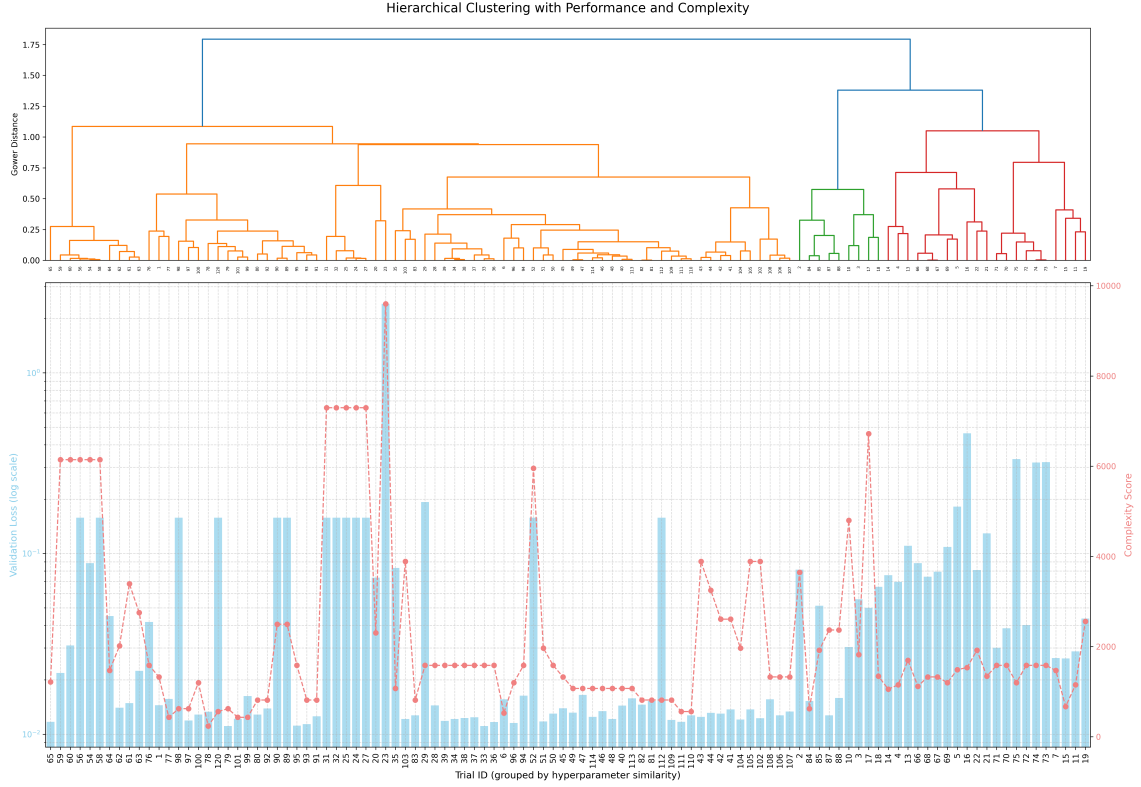


Figure S19: Hierarchical clustering of hyperparameter trials and their performance–complexity trade-off. The upper panel shows a dendrogram built from the Gower distance between all FTTransformer hyperparameter configurations explored during the search, grouping trials with similar architectural and training settings. The lower panel reports, for the exact trial ordering, the validation loss (blue bars, left y -axis, logarithmic scale) and the corresponding model–complexity score (red dashed line with markers, right y -axis), highlighting clusters of configurations that simultaneously minimize loss while avoiding unnecessarily complex models.

Figure S19 summarizes the structure of the hyperparameter search space and its relationship to model performance and complexity. Hierarchical clustering based on the Gower distance organizes all trials into families of similar architectures and optimization settings, which appear as contiguous blocks along the x -axis. Plotting the validation loss and complexity score for this clustered ordering reveals clear Pareto-like regions: some clusters contain overly simple models with high loss, others contain highly complex models with only marginal performance gains, and a narrow band of intermediate-complexity models achieves the best trade-off. The final FTTransformer configuration used in this

work was selected from this latter region, combining low validation loss with moderate complexity.

References

- 1 Hoppe, R. The Coordination Number – an "Inorganic Chameleon". *Angew. Chem. Int. Ed.* **1970**, *9*, 25–34, DOI: 10.1002/anie.197000251.
- 2 Hoppe, R. Effective Coordination Numbers (ECoN) and Mean Active Fictive Ionic Radii (MEFIR). *Z. Kristallogr.* **1979**, *150*, 23–52, DOI: 10.1524/zkri.1979.150.1-4.23.
- 3 Piotrowski, M. J.; Piquini, P.; Silva, J. L. F. D. Density Functional Theory Investigation of 3d, 4d, and 5d 13-atom Metal Clusters. *Phys. Rev. B* **2010**, *81*, 155446, DOI: 10.1103/PhysRevB.81.155446.
- 4 Chaves, A. S.; Piotrowski, M. J.; Silva, J. L. F. D. Evolution of the Structural, Energetic, and Electronic Properties of the 3d, 4d, and 5d Transition-Metal Clusters (30 TM_n Systems for n= 2–15): A Density Functional Theory Investigation. *Phys. Chem. Chem. Phys.* **2017**, *19*, 15484–15502, DOI: 10.1039/C7CP02240A.