

Understanding Stability Mechanisms in Single-Atom Alloys with Theory-infused Deep Learning

Yang Huang,^{†,‡} Shih-Han Wang,^{†,‡} Shuyi Cao,[†] Luke E. K. Achenie,[†] and
Hongliang Xin^{*,†}

[†]*Department of Chemical Engineering, Virginia Polytechnic Institute and State University,
Blacksburg, VA 24061, USA*

[‡]*Contributed equally to this work*

E-mail: hxin@vt.edu

Abstract

We present an interpretable deep learning model that enhances the prediction of cohesive energy in transition metal alloys (TMAs) by incorporating cohesion theory into a graph neural network (GNN) framework. The model not only predicts the total cohesive energy—an indicator of crystal stability—but also disentangles its various contributing factors and underlying physical parameters. The physics insights extracted from the model clarify the stability trends of transition metal surfaces across the periodic table. Furthermore, by applying the model to single-atom alloys (SAAs), a class of catalytically significant next-generation TMAs, we analyze and explain the relative stability of monomer/dimer (in-plane symmetry breaking) and top-/sub-layer (out-of-plane symmetry breaking) configurations. These two types of symmetry breaking lead to distinct thermodynamic preferences in SAAs, governed by localized effects (e.g. d -orbital coupling) and delocalized effects (e.g. wavefunction renormalization). The model is thus positioned as a powerful tool for understanding and strategically designing TMAs, enabling the tailored development of materials with improved stability for advanced applications in catalysis and materials science.

1 Introduction

TMAs are pivotal in heterogeneous catalysis due to their remarkable activity, selectivity, and tunable functionality across a broad spectrum of reactions^{1–3}. However, a persistent challenge in TMA catalyst design lies in reconciling high catalytic performance with insufficient stability^{4–6}. For example, CuPt alloys, though effective in oxygen reduction, suffer from leaching and degradation under operational conditions⁷. Recently, SAAs—where active metal atoms are atomically dispersed on a host metal surface—have gained attention as a new generation of TMA catalysts^{8–13}. Their unique site geometry decouples the transition state from intermediate binding, promoting efficient bond activation while minimizing surface poisoning and enabling deviation from conventional scaling relations^{14–19}. Despite these

advantages, SAAs still face critical stability issues: guest metal atoms can aggregate^{20,21} or migrate into the bulk, resulting in the loss of isolated active sites. These challenges underscore the need for robust approaches to predict and understand the thermodynamic and kinetic stability of TMAs across their vast chemical space.

Cohesive energy and formation energy are critical indicators of stability for TMAs. Recent advancements in machine learning (ML) have led to models that can predict these energies with high precision.^{22–24} However, the opaque nature of many advanced ML models limits their interpretability, preventing the extraction of chemical insights for materials design and impeding further enhancements in model generalization for discovering new materials.

In this study, we introduce an interpretable deep learning approach—Theory-Infused Neural Network (TinNet)—for predicting the cohesive energy of TMAs by integrating cohesion theory into a graph neural network (GNN) framework. Beyond accurate predictions of total cohesive energy, which reflects the overall crystal stability, our model provides insight into the individual contributing factors and intuitive physical parameters that reveal the underlying mechanisms of cohesion. This theory-guided approach helps clarify the stability trends of transition metal surfaces across the periodic table. Furthermore, by applying TinNet to SAAs, we analyze the relative stability of monomer/dimer and top-/sub-layer configurations—two representative cases of in-plane and out-of-plane symmetry breaking, respectively—each governed by distinct physical effects. We propose that TinNet offers a valuable and generalizable tool for the strategic design of stable materials, particularly in catalysis and related fields.

2 Computational Methods

DFT calculations. Spin-polarized DFT calculations of metal systems were performed through Quantum ESPRESSO with ultrasoft pseudopotentials. The exchange-correlation functional was approximated within the generalized gradient approximation (GGA) with

Perdew–Burke–Ernzerhof (PBE). The standard Perdew–Burke–Ernzerhof (PBE) functional, which incorporates a generalized gradient approximation (GGA) to account for exchange–correlation effects²⁵, has been demonstrated to deliver reliable electronic structure insights for such alloys, as supported by benchmarks and computational catalysis literature²⁶. It is important to note that for strongly correlated systems, e.g., transition metal oxides, advanced approaches like DFT+U²⁷, hybrid functionals²⁸, or even the random phase approximation²⁹ are necessary to accurately capture these effects. Metal surfaces with fcc {111}-, {100}-, and {211}-terminations were simulated using $2 \times 2 \times 4$, $2 \times 2 \times 4$, and $3 \times 2 \times 4$ supercells, respectively. Each slab has a vacuum of 15 \AA between two adjacent images along the z -direction. The bottom two layers were fixed while the top two layers were allowed to relax until a force criterion of $0.1 \text{ eV}/\text{\AA}$. A plane-wave energy cutoff of 500 eV was used. A Monkhorst–Pack mesh of $6 \times 6 \times 1$, $6 \times 6 \times 1$, and $4 \times 6 \times 1$ k -points was used to sample the Brillouin zone for fcc-phase {111}, {100}, and {211} terminations, respectively. The Methfessel–Paxton smearing scheme was used with a smearing parameter of 0.1 eV . Electronic energies were extrapolated to $k_B T = 0 \text{ eV}$. Projected density of states were obtained by projecting the eigenvectors of the full system onto the site atom of interest at a denser k -point sampling ($12 \times 12 \times 1$) with an energy spacing of 0.01 eV . There are 24 elements involved in metal alloys, including Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Y, Zr, Nb, Mo, Ru, Rh, Pd, Ag, Ta, W, Re, Os, Ir, Pt, and Au. We focus on site metals with more than half-filled d -states ($f_d > 0.5$) because early transition metals are susceptible to transforming into metal compounds, e.g., metal oxides, or being poisoned under reaction conditions.

TinNet models. The TinNet models used in this study are developed based on the graph convolutional neural network (GCNN)²² with theory integration³⁰. Supplementary Fig. X illustrates the flowchart of GCNN and TinNet models. The GCNN model is originally designed to predict bulk properties, e.g., formation energy, band gap, Fermi level, and shear modulus. Each atom has an atomic feature vector to represent its atomic properties. The

distance between different atoms with the Gaussian basis function is used as bond features. Through convolution operations, information is passed between atoms, while each atom learns information about its surroundings over iterations. The pooling layer averages all the learned atomic feature vectors in the same image, and uses the averaged feature vector to represent the properties of the entire image. A set of fully-connected layers are used to learn the relationship between the image feature vector and target properties. However, in the cohesive energy model, since the properties of each atom, such as $N_s, N_d, \alpha, \beta, W_d$, etc., need to be predicted, the pooling layer is removed and instead the learned atomic features of each atom are used as input features to predict physical parameters of the atom in a fully-connected neural network (FCNN). The SoftPlus activation function is used for the output of neural state variables ($N_s, N_d, \alpha, \beta, W_d$) to have a positive sign constraint by theory. For the total cohesive energy E_{coh} , the linear activation functions are used to alleviate the sign constraint.

3 Results and Discussion

3.1 Cohesion Theory of Transition Metals

In a cohesion process (Fig. 1), according to the renormalized-atom theory developed by Watson and co-workers in the 1970s,^{31,32} the first step is atomic promotion, in which each free atom is excited from its ground valence electron configuration $d^{n-2}s^2$ to the excited state $d^{n-1}s$. This energy value of the atom for each element can be obtained from NIST atomic spectra database³³.

$$\Delta E_{prom} = (d^{n-1}s)_{avg} - (d^{n-2}s^2)_{min} \quad (1)$$

where avg and min are operated over all the multiplets corresponding to the same electron configuration. For noble metals Cu, Ag, Au with $d^{10}s$ as ground state, the promotion energy is zero.

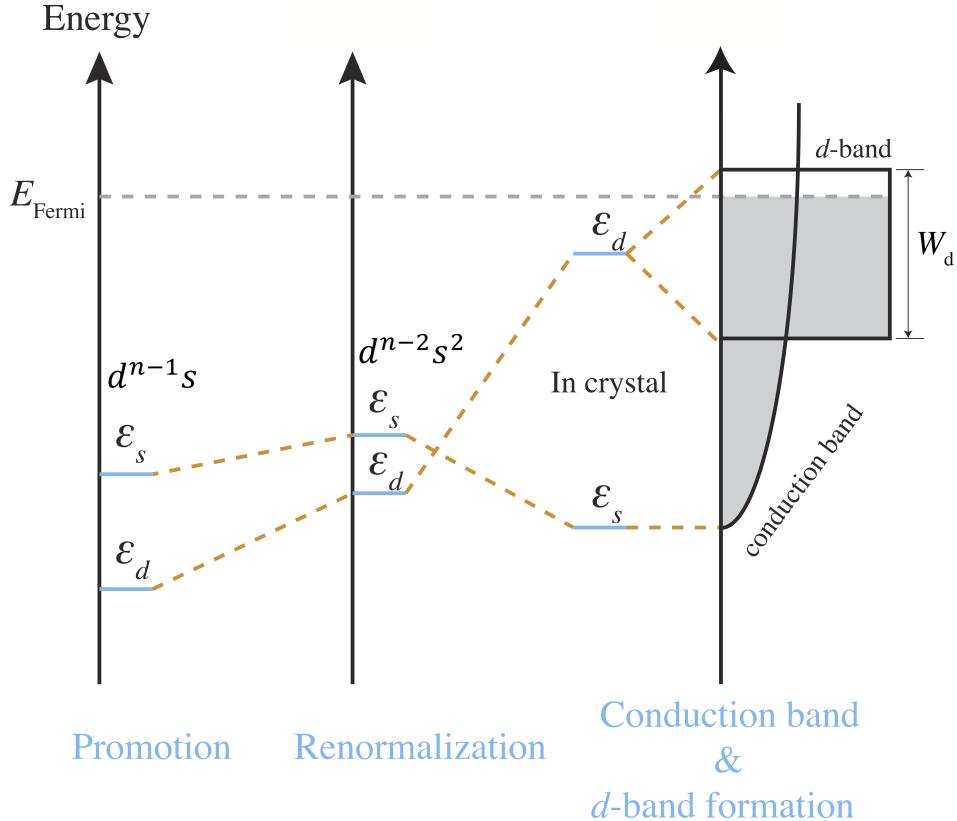


Figure 1: The cohesion processes according to the renormalized-atom cohesion theory of transition metals: promotion and renormalization of atomic s and d states, followed by conduction band and d -band formation.

The second step is renormalization. In this stage, the excited free atoms are squeezed into the crystal unit cell, feel the crystal potential and form effective atomic levels in solid state, called renormalized states. In 1970s, full band structure calculation was costly, so people tried non-self-consistent methods to get important band quantities like d -center and conduction band bottom. Renormalized-atom method is a successful one that can efficiently estimate these two important band quantities. Specifically, the name "renormalized" comes from renormalizing the atomic orbital wave function (the radial part $R_{nl}(r)$) within the Wigner-Seits cell,

$$R_{nl}(r) = \begin{cases} N_{nl} R_{nl}^{\text{atomic}}(r) & \text{if } r \leq R_{\text{WS}}, \\ 0 & \text{if } r > R_{\text{WS}}. \end{cases}$$

where N_{nl} is the renormalization constant

$$N_{nl}^{-2} = \int_0^{R_{\text{WS}}} [R_{nl}^{\text{atomic}}(r)]^2 r^2 dr \quad (2)$$

and the Wigner-Seitz cell is the space averagely occupied by each atom,

$$\frac{4\pi}{3} R_{\text{WS}}^3 = \frac{V}{N} \quad (3)$$

The renormalized atomic wave function is used to calculate the renormalized-atom potential for each orbital of each electron,

$$V_i^\alpha(r) = \sum_{j\{\alpha\}} \int_{\text{WS}} |\phi_j^\alpha(r')|^2 \frac{2}{|\vec{r} - \vec{r}'|} d^3 r' - \frac{2Z}{r} - \sum_{j\{\alpha\}} \delta(m_{si}, m_{sj}) \frac{\phi_j^\alpha(r)}{\phi_i^\alpha(r)} \int_{\text{WS}} \phi_j^{\alpha*}(r') \phi_i^\alpha(r') \frac{2}{|\vec{r} - \vec{r}'|} d^3 r' \quad (4)$$

where ϕ_i^α is the renormalized atomic wave function of orbital α of electron i . The first two terms are the Hartree potential that includes the electron-electron and nuclei-electron Coulomb potentials. The third term is the exchange potential written in the Hartree localized exchange approximation. The δ -function restricts the third term to electrons with the same spin. For renormalized d -orbital in the corresponding renormalized potential V^d , the solved energy level is called renormalized d -level, $\epsilon_{d,ren}$ and is close to the d -band center, ϵ_d . For orthogonalized plane wave with $k = 0$ in the corresponding renormalized potential V^s , the solved energy level is called renormalized s -level, $\epsilon_{s,ren}$ and is close to the conduction band bottom, ϵ_Γ . Therefore, we have the renormalization energy:

$$\Delta E_{ren} = N_s \epsilon_{s,ren} + N_d \epsilon_{d,ren} - (d^{m-1} s)_{avg} \approx N_s \epsilon_\Gamma + N_d \epsilon_d - (d^{n-1} s)_{avg} \quad (5)$$

Finally, the renormalized states will be expanded to form energy bands due to Pauli exclusion principle (for de-localized *s*-electrons) and orbital couplings (for localized *d*-electrons). Specifically, based on free electron approximation, we have the conduction band formation energy,

$$\Delta E_{s-band} = \int_{-\infty}^{E_F} \epsilon \rho_s(\epsilon) d\epsilon - N_s \epsilon_{s,ren} \approx N_s \epsilon_\Gamma + \frac{3}{5} k_F^2 - N_s \epsilon_{s,ren} \approx \frac{3}{5} k_F^2 = \frac{3\hbar}{10m} \left(\frac{3\pi^2 N_s}{V_{WS}} \right)^{2/3} \quad (6)$$

where ϵ_Γ is the conduction band bottom energy at the Γ point of the reciprocal space. Based on tight-binding and rectangular band approximation, we have the the *d*-band formation energy,

$$\Delta E_{d-band} = \int_{-\infty}^{E_F} \epsilon \rho_d(\epsilon) d\epsilon - N_d \epsilon_{d,ren} \approx \int_{-\infty}^{E_F} (\epsilon - \epsilon_d) \rho_d(\epsilon) d\epsilon \approx -\frac{W}{20} N_d (10 - N_d) \quad (7)$$

where ϵ_d is the *d*-band center. The sum of all these terms leads to the overall cohesive energy for each atom site. Among them, the promotion energy is associated with the atomic electronic configuration prior to crystal formation and is independent of the surrounding chemical environment. The *d*-band formation energy, originating from the hybridization and coupling of localized *d*-orbitals, can be classified as a localized effect. In contrast, the renormalization energy and the *s*-band formation energy are governed by the spatial redistribution of electrons in the crystal and hence are considered delocalized effects.

More specifically, renormalization energy reflects the change in on-site electronic energies due to the compression of valence electrons—particularly *s*-electrons—into the finite Wigner-Seitz volume of the solid. This effect is sensitive to the atomic size, crystal volumes, valence electron density, and the number of *d*-electrons that repel the *s*-electrons spatially. On the other hand, the *s*-band formation energy arises from the broadening of *s*-states into an extended conduction band, driven by Pauli exclusion and the need to accommodate the valence *s*-electrons across available momentum states. This term mainly depends on electron density. Together, these delocalized contributions encapsulate the kinetic and electrostatic

consequences of embedding atoms into a shared electronic environment.

3.2 Interpretable Deep Learning Framework

The formula based on renormalized-atom cohesion theory provides interpretation of cohesive energy prediction, however, the prediction is inaccurate due to the approximations made. In order to enhance prediction accuracy, we parameterized the formula and let the neural network determine all the physical parameters. Physical parameters that will be predicted by the neural network are: the renormalization energy ΔE_{ren} , the number of d -band electrons N_d , d -band width W_d in the rectangular band approximation, and the correction parameters α and β ,

$$E_{coh} = \frac{1}{N} \sum_i^N [(d^{n-1}s)_{avg} - (d^{n-2}s^2)_{min} + \Delta E_{ren} + \alpha \frac{3\hbar}{10m} (3\pi^2 N_s/V_{WS})^{2/3} + \beta \frac{W}{20} N_d (10 - N_d)] \quad (8)$$

The ground truth of cohesive energy E_{coh} is from DFT calculations. The promotion energy $(d^{n-1}s)_{avg} - (d^{n-2}s^2)_{min}$ is pre-tabulated values from NIST atomic spectra database³³. The ground truth of N_d is calculated from the integration up to Fermi level of the d -band of the site atom. The value of N_s is calculated from $n - N_d$ where n is the total number of valence electrons of the corresponding element. The ground truth of conduction band formation energy $\alpha \frac{3\hbar}{10m} (3\pi^2 N_s/V_{WS})^{2/3}$ is calculated using the projected s -band of each atom,

$$\alpha \frac{3\hbar}{10m} (3\pi^2 N_s/V_{WS})^{2/3} = \int_{-\infty}^{E_F} \epsilon \rho_s(\epsilon) d\epsilon - N_s \epsilon_\Gamma = \int_{-\infty}^{E_F} \epsilon \rho_s(\epsilon) d\epsilon - \epsilon_\Gamma \int_{-\infty}^{E_F} \rho_s(\epsilon) d\epsilon \quad (9)$$

therefore we can get the "ground truth" of α and put it into the loss function. The ground truth of the d -band formation energy $\beta \frac{W}{20} N_d (10 - N_d)$ is calculated using the projected d -band of each atom,

$$\beta \frac{W}{20} N_d (10 - N_d) = \int_{-\infty}^{E_F} \epsilon \rho_d(\epsilon) d\epsilon - N_d \epsilon_d = \int_{-\infty}^{E_F} (\epsilon - \epsilon_d) \rho_d(\epsilon) d\epsilon \quad (10)$$

therefore we can get the "ground truth" of β and put it into the loss function.

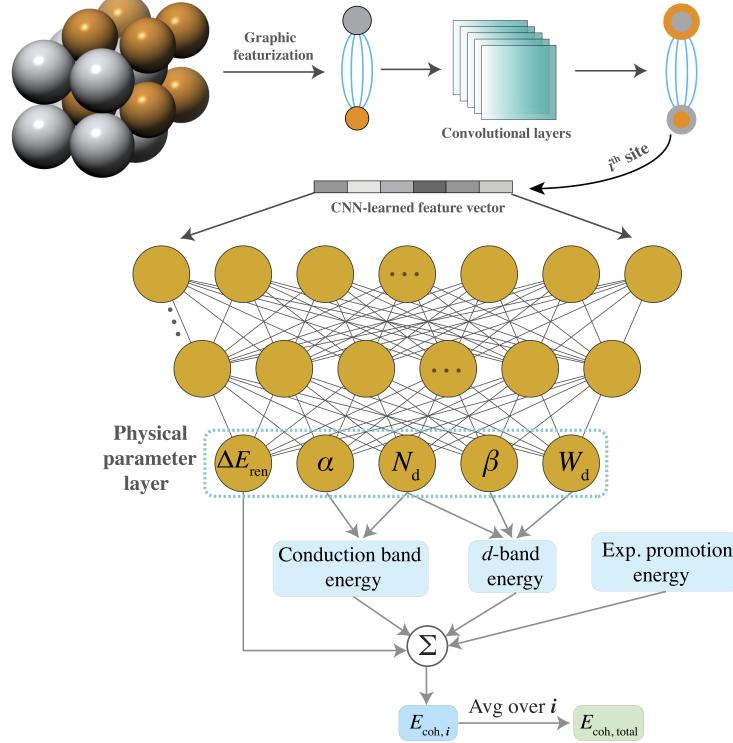


Figure 2: The TinNet architecture for accurate and interpretable prediction of cohesive energy.

The architecture of the neural network is shown in Fig. 2. The unit cell of the crystal is represented by an undirected graph and fed into a convolutional neural network. A fixed-length vector for each atomic site is learned and fed into a fully-connected neural network in order to predict all the 5 parameters in the formula. Then, the predicted parameters are put back into the formulas to calculate all the contribution terms and also the total cohesive energy. All the atomic cohesive energies are averaged over sites and the cohesive energy for each crystal is eventually obtained.

The training dataset consists of 8526 FCC binary surface alloys in 111, 100 and 211 facets. The test mean absolute error (MAE) of the cohesive energy prediction, 0.02 eV/atom, is as small as the state-of-art purely data-driven deep learning algorithms such as CGCNN²². Besides, all the physical and correction parameters are also accurately predicted. The MAE for W_d , N_d , α and β is 0.07 eV, 0.01, 0.008 and 0.007, respectively.

3.3 Origins of the Stability Trends of Transition Metal Surfaces

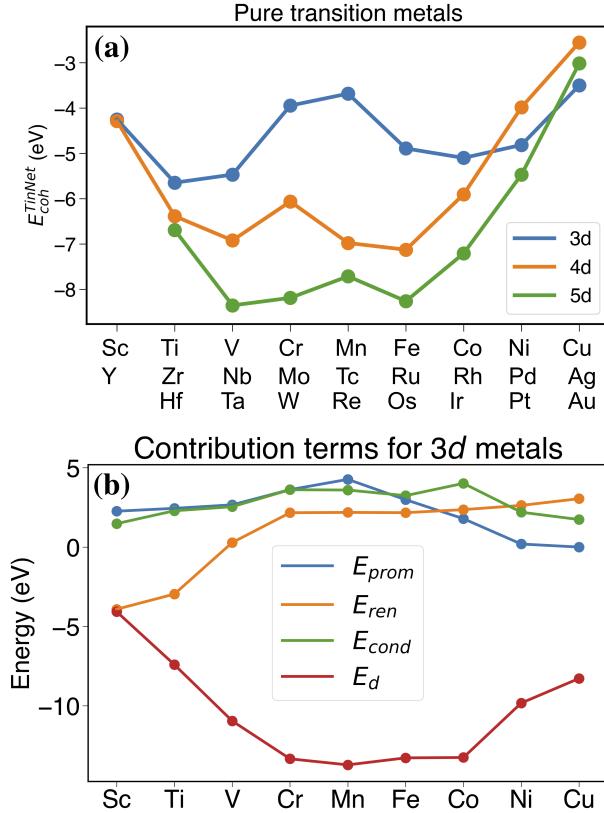


Figure 3: The TinNet-predicted (a) total cohesive energy of 3d, 4d and 5d pure metal surfaces and (b) contribution components in 3d-metal surfaces.

Our model not only provides accurate predictions on the total cohesive energy, but can also elucidate where such value is originated, by accurately predicting contribution terms at the same time. For example, as shown in Fig. 3a, for the pure 3d 4d and 5d metal surfaces (fcc-111), the cohesive energy is roughly parabolic across each row, except the high values near the middle. These trends can be explained in terms of the different contribution terms. We take 3d metals for illustration (Fig. 3b). The main contribution is from d -band formation energy which is parabolic across the row. Other contribution terms are roughly all parabolic but in the opposite direction, so they should all contribute to the irregularities in the middle. The dominant one is the promotion energy due to the stable ground state electronic configuration of mid-period elements, which is consistent with bulk results in

literatures^{32,34}.

The full physical picture underlying the stability trends of transition metals is as follows: The atomic promotion energy is uniformly positive—except for the noble metals, where it effectively vanishes—and is largest for elements with nearly half-filled *d*-shells due to Hund’s rule. The renormalization energy peaks near the center of the transition metal row, where the metallic Wigner–Seitz radius is smallest relative to the free-atom radius, resulting in maximal renormalization of atomic charge. The notable drop in renormalization energy from Cr to Mn arises from the abrupt increase in Wigner–Seitz radius between the two elements. Overall, its magnitude reflects the ratio between the Wigner–Seitz radius and the atomic radius. For mid-row elements, the *d*-bands strongly pull atoms together, compressing the conduction electrons. This compression elevates the conduction band relative to the renormalized atomic *s*-level, thereby reducing conduction-band bonding strength. The *d*-band formation energy—which also includes hybridization between the conduction and *d* bands—constitutes the largest contribution to the cohesive energy of transition metals. Its nearly parabolic trend across the series is a generic feature of the *d*-band contribution.

3.4 Stability Mechanisms in Single-Atom Alloys

We applied our TinNet model to systematically investigate the stability of SAAs in two representative scenarios: dimer/monomer conversion (Fig. 4a), which involves in-plane symmetry breaking, and top-/sub-layer conversion (Fig. 5a), which involves out-of-plane symmetry breaking. The guest metals analyzed include Pt, Pd, Co, Ni, Cu, Os, Ru, Rh, Re, and Ir, while Cu, Ag, and Au served as host metals. To analyze the underlying stability mechanism in the dimer/monomer configurations of SAAs, we constructed $4 \times 4 \times 4$ fcc-(111) slabs doped with two guest metal atoms to represent both configurations: the dimer, where the guest atoms are adjacent, and the monomer, where they are well-separated. This consistent setup allows for a direct comparison of total energies for the same guest–host combination. The relative stability is defined as the total energy difference between the dimer and monomer

configurations. DFT calculations reveal that in most cases, the dimer configuration tends to be more stable than the monomer. As shown in Fig. 4b, the overall trend is primarily governed by the localized *d*-band formation energy, which dominates due to the absence of out-of-plane symmetry breaking during in-plane migration. The structural distinction between the two configurations lies in the replacement of guest–host bonds (in monomer) by guest–guest bonds (in dimer). Therefore, the difference in orbital coupling strength is considered the main contributor to the observed stability trends. According to tight-binding theory, the *d*-band width is proportional to the coupling strength^{35,36}, which is positively correlated with the *d*-orbital radius r_d . For a guest metal site on both configurations, based on Equation 7, their *d*-band formation energy difference can be approximated as the descriptor: $-N_d^g(10 - N_d^g)r_d^g(r_d^g - r_d^h)$, where *g* denotes guest and *h* denotes host metals. As shown in Fig. 4c, this descriptor exhibits strong correlation with the computed energy differences, effectively capturing the trend in dimer stability. A local interpretability analysis is carried out on Re₂Cu system which highlights a notably large energy difference between dimer and monomer configurations, as demonstrated in Fig. 4d where the contribution terms of each atom site are plotted. The enhanced stability of the dimer is attributed to stronger Re–Re than Re–Cu *d*-orbital coupling as the result of larger *d*-orbital radius of Re (0.88) compared with Cu (0.49), which significantly lowers the *d*-band formation energy at Re sites. This specific example reflects a broader trend observed across many SAAs: the dimer configuration tends to be more stable than the monomer. This is because Cu, Ag, and Au, serving as host metals, are located at the far right of the transition metal series in the periodic table and thus have relatively small *d*-orbital radius, which results in stronger guest–guest bonding when guest atoms are brought into proximity.

In the case of SAAs with guest atoms located in the top versus sub-layer positions, DFT calculations reveal that the sub-layer configuration tends to be more stable in most systems. Since inter-layer migration involves out-of-plane symmetry breaking, the delocalized effects, namely the combined contribution of renormalization energy and *s*-band formation energy,

dominate the overall energy difference (Fig. 5b). Compared to intra-layer migration (e.g., the dimer/monomer case), the inter-layer migration of guest atoms in SAAs—from sub-layer to top-layer—introduces two fundamental differences. First, the degree of electron confinement varies: when located in the top layer, the guest atom’s electrons experience less confinement, whereas in the sub-layer, they are fully embedded within the bulk environment. Second, the out-of-plane symmetry breaking can lead to larger changes in available free space around the migrating atom than in-plane symmetry breaking. To capture the energetic consequences of these two effects, we construct a simple descriptor defined as $\Delta N_d \cdot \Delta R$. Here, ΔN_d is the ratio of d -electron count between guest and host atoms (i.e., $N_d^{\text{guest}}/N_d^{\text{host}}$), serving as a proxy for the degree of electronic confinement: a smaller ΔN_d indicates fewer guest d -electrons can “spread out” on the surface, leading to stronger confinement in the bulk and, consequently, stronger $d-s$ repulsion and higher renormalization energy. The term ΔR is the ratio between relative local space radius R of the two configurations. The relative local space radius R of each configuration is defined as the radius ratio between the polyhedral centered at the sub-layer guest atom position (constructed from its 12 nearest-neighbor atom positions) and the free isolated center atom³⁷. This geometric measure reflects both the available free space for electron delocalization—which affects the kinetic energy of conduction band electrons—and the extent of spatial confinement, which in turn influences the renormalization energy through wavefunction compression. As shown in Fig.5c, this composite descriptor qualitatively captures the trends in total energy differences between top- and sub-layer configurations across different systems. A local interpretation on Co₁Au, illustrated in Fig.5d, supports this physical picture: the sub-layer configuration is more stable due to less confined d -electrons and a larger relative local space size, both of which contribute to a reduced total cohesive energy. Moreover, in most systems, the sub-layer configuration is thermodynamically favored. This trend can be explained by the fact that guest metals typically have fewer d -electrons than the host d^{10} metals Cu, Ag, and Au. As a result, when the guest atom occupies the sub-layer and the host atom remains in the top layer, more

electrons reside near the surface and less electrons get confined in the bulk region, thereby lowering the overall renormalization energy.

It is important to note that all the above analyses are based solely on thermodynamic preferences. However, actual surface configurations result from both thermodynamics and kinetics. If the system is kinetically limited—i.e., significant energy barriers exist along the atomic diffusion pathways—then even thermodynamically favorable transformations such as monomer-to-dimer aggregation or top-to-sub-layer migration may not occur in practice. Moreover, our current study considers only bare surfaces, neglecting adsorbate-induced effects, which are often significant under realistic reaction conditions and may alter the preferred configurations of surface atoms.

3.5 Outlook

While this work focuses on the application of TinNet to the stability analysis of SAAs, the framework’s potential is not limited to this domain. Due to its integration of general physics theory with graph-based learning, the model is well-suited for extension to broader and more complex compositional spaces, such as high-entropy alloys (HEAs)^{38,39} and doped dimer or trimer SAA clusters⁴⁰, where stability predictions remain a major challenge. These systems present a combinatorially vast design space with intricate bonding environments, making them ideal candidates for TinNet-driven analysis. In future work, we plan to explore these directions and further refine TinNet to support the rational discovery and design of next-generation alloy catalysts with both high stability and targeted functionalities.

4 Conclusions

In conclusion, we introduce TinNet, an interpretable graph neural network model that successfully integrates cohesion theory into deep learning for accurate and physically meaningful prediction of cohesive energy in TMAs. TinNet not only captures total energy trends but

also reveals the underlying contributions from localized and delocalized interactions. By applying the model to SAAs, we uncover the stability mechanisms behind monomer/dimer and top/sub-layer configurations, demonstrating how electronic structure and spatial confinement interplay to govern thermodynamic preferences. While the present study offers an initial demonstration of TinNet’s capabilities in stability analysis, we believe this framework holds broader potential for analyzing more complex alloy systems with larger design spaces. TinNet thus provides a promising foundation for interpretable and scalable materials discovery.

Acknowledgements

Y.H. and H.X. acknowledge the financial support from the US Department of Energy, Office of Basic Energy Sciences under contract no. DE-SC0023323. S.H.W. and L.E.K.A. especially thank to the NSF Non-Academic Research Internships for Graduate Students (INTERN) program (CBET-1845531) for supporting part of the work in NIST under the guidance of Dr. Kamal Choudhary. The computational resource used in this work is provided by the advanced research computing at Virginia Polytechnic Institute and State University.

Data availability

Currently, all source data along with Jupyter Notebooks for data preprocessing, model development, and *post hoc* analysis are available from the GitHub repository: .

Competing interests

The authors declare no competing interests.

References

- (1) Schwank, J. W. Bimetallic catalysts: Discoveries, concepts, and applications. By John H. Sinfelt, John Wiley & Sons, 1983. XI+ 164 pp. **1985**,

- (2) Kitchin, J.; Nørskov, J. K.; Barteau, M.; Chen, J. Modification of the surface electronic and chemical properties of Pt (111) by subsurface 3d transition metals. *The Journal of chemical physics* **2004**, *120*, 10240–10246.
- (3) Zhang, J.; Vukmirovic, M. B.; Xu, Y.; Mavrikakis, M.; Adzic, R. R. Controlling the catalytic activity of platinum-monolayer electrocatalysts for oxygen reduction with different substrates. *Angewandte Chemie* **2005**, *117*, 2170–2173.
- (4) Antolini, E.; Salgado, J. R.; Gonzalez, E. R. The stability of Pt–M (M= first row transition metal) alloy catalysts and its effect on the activity in low temperature fuel cells: a literature review and tests on a Pt–Co catalyst. *Journal of Power Sources* **2006**, *160*, 957–968.
- (5) Bezerra, C. W.; Zhang, L.; Liu, H.; Lee, K.; Marques, A. L.; Marques, E. P.; Wang, H.; Zhang, J. A review of heat-treatment effects on activity and stability of PEM fuel cell catalysts for oxygen reduction reaction. *Journal of Power Sources* **2007**, *173*, 891–908.
- (6) Meier, J. C.; Galeano, C.; Katsounaros, I.; Witte, J.; Bongard, H. J.; Topalov, A. A.; Baldizzone, C.; Mezzavilla, S.; Schüth, F.; Mayrhofer, K. J. Design criteria for stable Pt/C fuel cell catalysts. *Beilstein journal of nanotechnology* **2014**, *5*, 44–67.
- (7) Zhang, J.; Yuan, Y.; Gao, L.; Zeng, G.; Li, M.; Huang, H. Stabilizing Pt-based electrocatalysts for oxygen reduction reaction: fundamental understanding and design strategies. *Advanced Materials* **2021**, *33*, 2006494.
- (8) Gao, Q.; Han, X.; Liu, Y.; Zhu, H. Electrifying Energy and Chemical Transformations with Single-Atom Alloy Nanoparticle Catalysts. *ACS catalysis* **2024**, *14*, 6045–6061.
- (9) Zhang, T.; Walsh, A. G.; Yu, J.; Zhang, P. Single-atom alloy catalysts: structural analysis, electronic properties and catalytic activities. *Chemical Society Reviews* **2021**, *50*, 569–588.

- (10) Giannakakis, G.; Flytzani-Stephanopoulos, M.; Sykes, E. C. H. Single-atom alloys as a reductionist approach to the rational design of heterogeneous catalysts. *Accounts of chemical research* **2018**, *52*, 237–247.
- (11) Liu, J.; Lucci, F. R.; Yang, M.; Lee, S.; Marcinkowski, M. D.; Therrien, A. J.; Williams, C. T.; Sykes, E. C. H.; Flytzani-Stephanopoulos, M. Tackling CO poisoning with single-atom alloy catalysts. *Journal of the American Chemical Society* **2016**, *138*, 6396–6399.
- (12) Zhang, X.; Cui, G.; Feng, H.; Chen, L.; Wang, H.; Wang, B.; Zhang, X.; Zheng, L.; Hong, S.; Wei, M. Platinum–copper single atom alloy catalysts with high performance towards glycerol hydrogenolysis. *Nature Communications* **2019**, *10*, 5812.
- (13) Huang, Y.; Wang, S.-H.; Achenie, L. E.; Choudhary, K.; Xin, H. Origin of unique electronic structures of single-atom alloys unraveled by interpretable deep learning. *The Journal of Chemical Physics* **2024**, *161*.
- (14) Lucci, F. R.; Darby, M. T.; Mattera, M. F.; Ivimey, C. J.; Therrien, A. J.; Michaelides, A.; Stamatakis, M.; Sykes, E. C. H. Controlling hydrogen activation, spillover, and desorption with Pd–Au single-atom alloys. *The journal of physical chemistry letters* **2016**, *7*, 480–485.
- (15) Réocreux, R.; Kress, P. L.; Hannagan, R. T.; Çınar, V.; Stamatakis, M.; Sykes, E. C. H. Controlling Hydrocarbon (De) Hydrogenation Pathways with Bifunctional PtCu Single-Atom Alloys. *The Journal of Physical Chemistry Letters* **2020**, *11*, 8751–8757.
- (16) Cheng, X.; Wang, Y.; Lu, Y.; Zheng, L.; Sun, S.; Li, H.; Chen, G.; Zhang, J. Single-atom alloy with Pt-Co dual sites as an efficient electrocatalyst for oxygen reduction reaction. *Applied Catalysis B: Environmental* **2022**, *306*, 121112.
- (17) Gao, Q.; Yao, B.; Pillai, H. S.; Zang, W.; Han, X.; Liu, Y.; Yu, S.-W.; Yan, Z.; Min, B.; Zhang, S. et al. Synthesis of core/shell nanocrystals with ordered intermetallic single-

- atom alloy layers for nitrate electroreduction to ammonia. *Nature Synthesis* **2023**, *2*, 624–634.
- (18) Darby, M. T.; Sykes, E. C. H.; Michaelides, A.; Stamatakis, M. Carbon monoxide poisoning resistance and structural stability of single atom alloys. *Topics in catalysis* **2018**, *61*, 428–438.
- (19) Hannagan, R. T.; Giannakakis, G.; Flytzani-Stephanopoulos, M.; Sykes, E. C. H. Single-atom alloy catalysis. *Chemical Reviews* **2020**, *120*, 12044–12088.
- (20) McCue, A. J.; Anderson, J. A. CO induced surface segregation as a means of improving surface composition and enhancing performance of CuPd bimetallic catalysts. *Journal of Catalysis* **2015**, *329*, 538–546.
- (21) Ouyang, M.; Papanikolaou, K. G.; Boubnov, A.; Hoffman, A. S.; Giannakakis, G.; Bare, S. R.; Stamatakis, M.; Flytzani-Stephanopoulos, M.; Sykes, E. C. H. Directing reaction pathways via in situ control of active site geometries in PdAu single-atom alloy catalysts. *Nature Communications* **2021**, *12*, 1549.
- (22) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters* **2018**, *120*, 145301.
- (23) Hart, G. L.; Mueller, T.; Toher, C.; Curtarolo, S. Machine learning for alloys. *Nature Reviews Materials* **2021**, *6*, 730–755.
- (24) Ubaru, S.; Międlar, A.; Saad, Y.; Chelikowsky, J. R. Formation enthalpies for transition metal alloys using machine learning. *Physical Review B* **2017**, *95*, 214102.
- (25) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *The Journal of chemical physics* **1999**, *110*, 5029–5036.

- (26) Nørskov, J. K.; Abild-Pedersen, F.; Studt, F.; Bligaard, T. Density functional theory in surface chemistry and catalysis. *Proceedings of the National Academy of Sciences* **2011**, *108*, 937–943.
- (27) Capdevila-Cortada, M.; Łodziana, Z.; López, N. Performance of DFT+ U approaches in the study of catalytic materials. 2016.
- (28) Walsh, A.; Butler, K. T. Prediction of electron energies in metal oxides. *Accounts of chemical research* **2014**, *47*, 364–372.
- (29) Yan, J.; Hummelshøj, J. S.; Nørskov, J. K. Formation energies of group I and II metal oxides using random phase approximation. *Physical Review B* **2013**, *87*, 075207.
- (30) Wang, S.-H.; Pillai, H. S.; Wang, S.; Achenie, L. E.; Xin, H. Infusing theory into deep learning for interpretable reactivity prediction. *Nature communications* **2021**, *12*, 5288.
- (31) Hodges, L.; Watson, R.; Ehrenreich, H. Renormalized atoms and the band theory of transition metals. *Physical Review B* **1972**, *5*, 3953.
- (32) Gelatt Jr, C.; Ehrenreich, H.; Watson, R. Renormalized atoms: cohesion in transition metals. *Physical Review B* **1977**, *15*, 1613.
- (33) Ralchenko, Y. NIST atomic spectra database. *Memorie della Societa Astronomica Italiana Supplementi* **2005**, *8*, 96.
- (34) Brooks, M.; Johansson, B. Exchange integral matrices and cohesive energies of transition metal atoms. *Journal of Physics F: Metal Physics* **1983**, *13*, L197.
- (35) Huang, Y.; Wang, S.-H.; Kamanuru, M.; Achenie, L. E.; Kitchin, J. R.; Xin, H. Unifying theory of electronic descriptors of metal surfaces upon perturbation. *Physical Review B* **2024**, *110*, L121404.

- (36) Kitchin, J. R.; Nørskov, J. K.; Bartheau, M. A.; Chen, J. Role of strain and ligand effects in the modification of the electronic and chemical properties of bimetallic surfaces. *Physical review letters* **2004**, *93*, 156801.
- (37) Clementi, E.; Raimondi, D.-L. Atomic screening constants from SCF functions. *The Journal of Chemical Physics* **1963**, *38*, 2686–2689.
- (38) Bueno, S. L.; Leonardi, A.; Kar, N.; Chatterjee, K.; Zhan, X.; Chen, C.; Wang, Z.; Engel, M.; Fung, V.; Skrabalak, S. E. Quinary, Senary, and Septenary High Entropy Alloy Nanoparticle Catalysts from Core@ Shell Nanoparticles and the Significance of Intraparticle Heterogeneity. *ACS nano* **2022**, *16*, 18873–18885.
- (39) Huang, Y.; Wang, S.-H.; Wang, X.; Omidvar, N.; Achenie, L. E.; Skrabalak, S. E.; Xin, H. Unraveling Reactivity Origin of Oxygen Reduction at High-Entropy Alloy Electrocatalysts with a Computational and Data-Driven Approach. *The Journal of Physical Chemistry C* **2024**, *128*, 11183–11189.
- (40) Kress, P. L.; Zhang, S.; Wang, Y.; Çınar, V.; Friend, C. M.; Sykes, E. C. H.; Montemore, M. M. A priori design of dual-atom alloy sites and experimental demonstration of ethanol dehydrogenation and dehydration on PtCrAg. *Journal of the American Chemical Society* **2023**, *145*, 8401–8407.

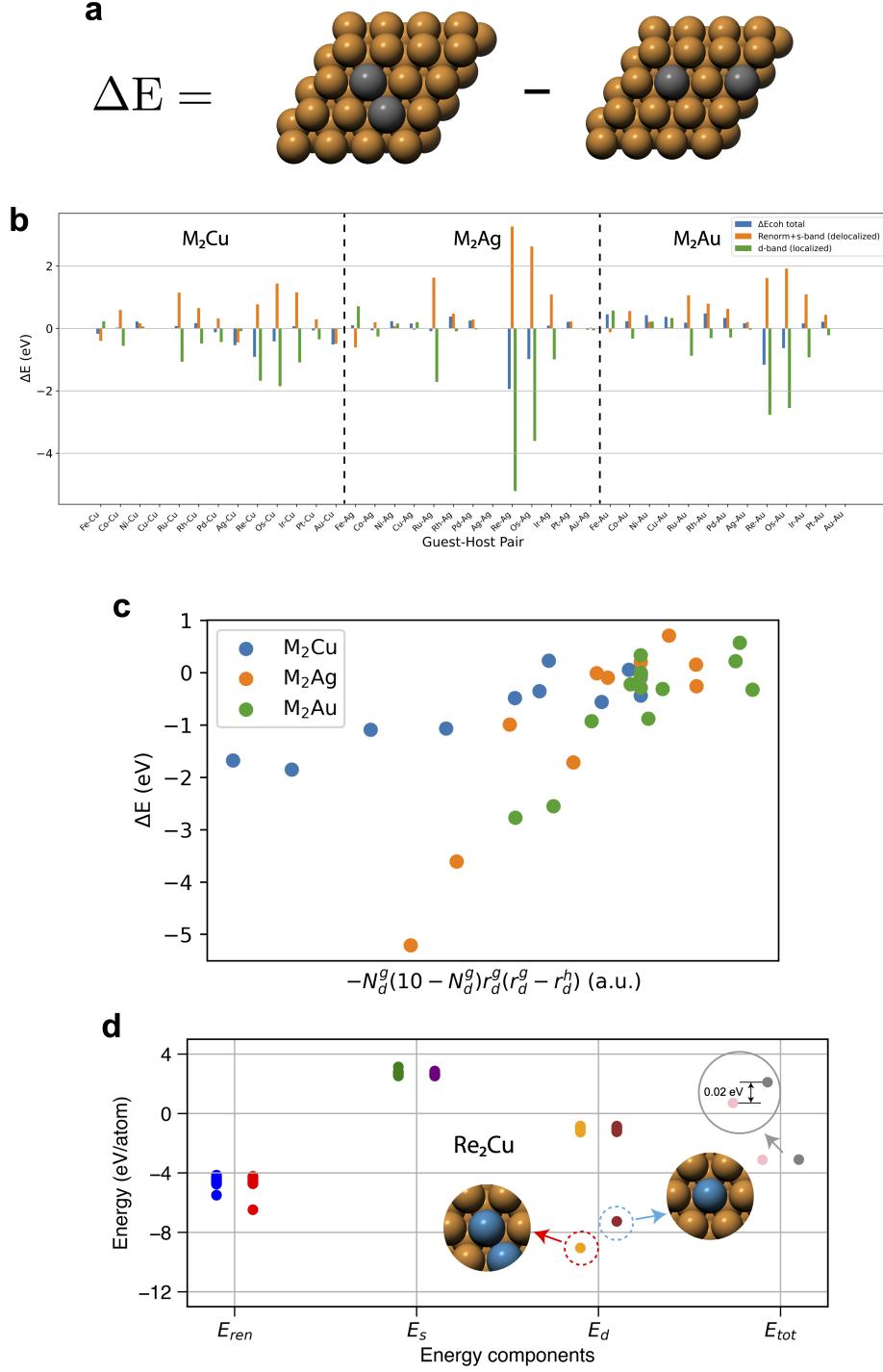


Figure 4: (a) Model structures to represent dimer and monomer SAAs. The ΔE is defined as the total energy difference between these two model structures. (b) Total and decomposed terms of cohesive energy difference between dimer and monomer SAA. (c) The correlation between the descriptor of localized effects and the energy difference between monomer and dimer configurations with Cu, Ag and Au substrates. The descriptor of delocalized effects characterizes chemical bond environment change, specifically the d -orbital coupling strength change around the guest metal atom. (d) The distribution of each cohesive energy component of atoms in the two Re_2Cu configurations.

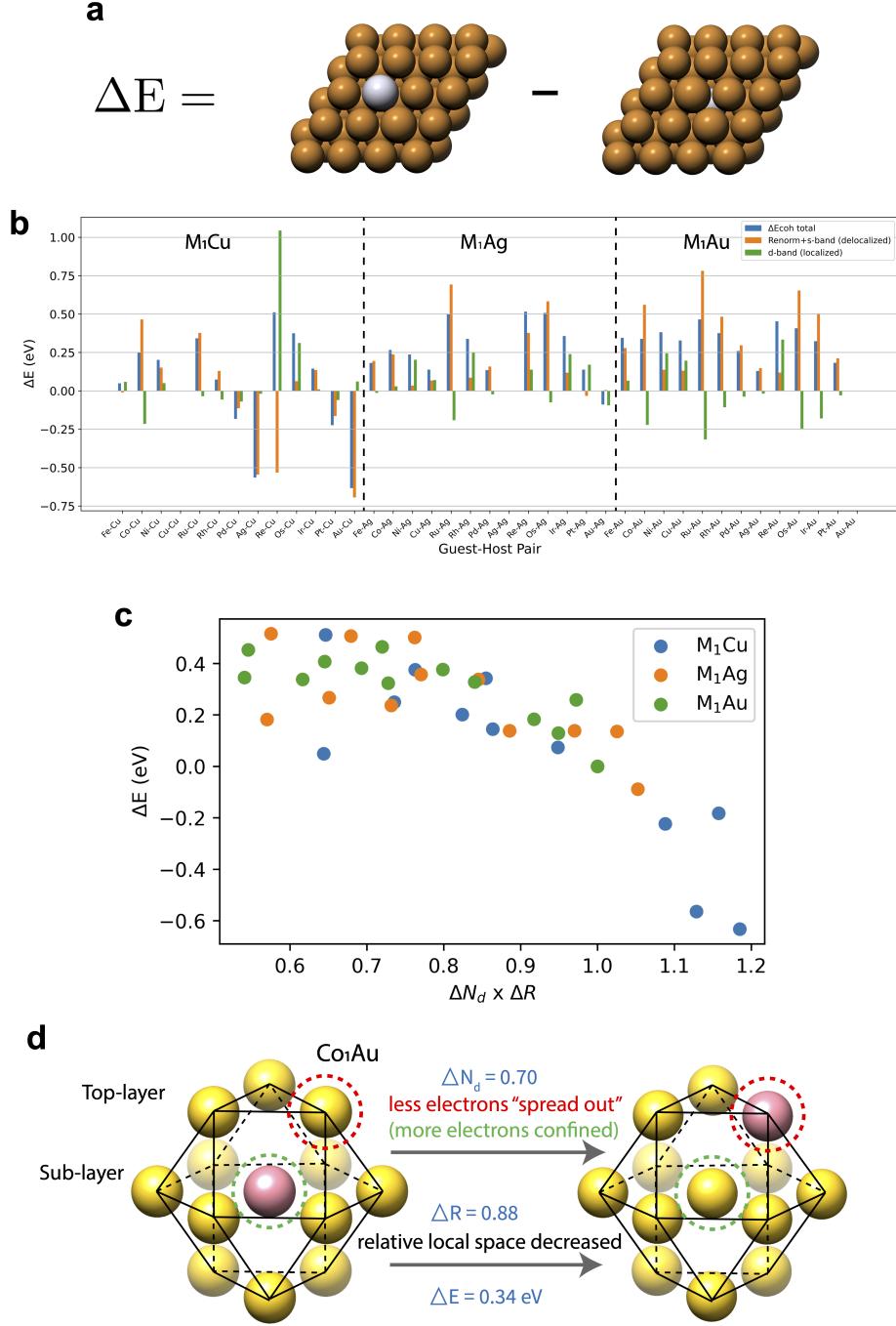


Figure 5: (a) Model structures to represent SAAs with guest metal at top- and sub-layers. The ΔE is defined as the total energy difference between these two model structures. (b) Total and decomposed terms of cohesive energy difference between SAAs in these two configurations. (c) The correlation between descriptor of delocalized effects and the total energy difference. The descriptor of delocalized effects contains two components, the number of d -electrons change at top-layer and relative local space size change around sub-layer guest atom site, characterizing the degree of wavefunction renormalization and electron confinement. (d) Mechanisms of relative stability of Co₁Au SAA between its top- and sub-layer configurations.