

Supporting Information

A Transferable Machine-learning Scheme from Pure Metals to Alloys in Predicting Adsorption Energies

Xin Li^a, Bo Li^a, Ze Yang^a, Zhiwen Chen^b, Wang Gao^{*a}, and Qing Jiang^a

^aKey Laboratory of Automobile Materials, Ministry of Education, Department of Materials Science and Engineering, Jilin University 130022, Changchun, China.

^bDepartment of Materials Science and Engineering, University of Toronto, Toronto, ON, M5S, 3E4, Canada.

Note S1: A database of adsorption energies for machine learning models.

Among the considered datasets, pure transition metals (TMs) contain Ag, Au, Cu, Pt, Pd, Ir, Ni, Rh and Ru, single-atom alloys (SAAs) contain Ag@Cu, Au@Ni, Pt@Rh and Pd@Ir, and AB intermetallics (ABs) contain AgAu, AgPt, PtRh and IrRu, which are from reference 1. We consider the (100) facet with 2 adsorption sites, the (110) facet with 4 sites, the (111) facet with 3 sites and (211) facets with 12 sites. Each metallic adsorption systems contain six adsorbates of C, CH, CO, H, O and OH. For the high-entropy alloys (HEAs) with the adsorbate of OH on IrPdPtRhRu, we consider the (100) facet with 1 bridge site, (110) facet with 2 bridge sites, (111) facet with 1 bridge site, (211) facet with 3 bridge sites and (532) facet with 6 bridge sites, which are all from reference 2. The considered adsorption sites are shown in Figure S1. For the case of the interchanges of nearest-neighboring atoms on HEAs, we consider the HEA structures with the adsorption sites of RhRh, PtPt and PdIr. For the TM nanoparticles (NPs) from reference 3, the data contain three morphologies of Cube, Cuboctahedron and Icosahedron as well as three sizes of 55-, 147- and 172-atoms, with three adsorbates of CH₃, CO and OH. The structures are shown in Figure S2. The datasets on TMs with adatoms come from reference 4, which include (100), (111), (211) facets and three adsorbates of CH₃, CO and OH.

Note S2: A Python script for determining the adsorption sites and active sites of targeted structures and calculating the corresponding descriptors.

We design a Python script to make it easy for users to obtain a large number of descriptors of the targeted structures. Based on the package of Atoms Simulation Environment (ASE)⁵, this script is used for identifying the adsorption sites and active sites of the targeted structures built for Material Studio, VASP and other accessible files or softwares supported by ASE. The script determines the nearest neighboring atoms by collision sphere detection for ASE, namely given a number of spheres of the different radius located at the different points, it calculates the pairs of spheres that overlap. We use this detection to identify the neighboring atoms and coordination numbers. Taking the given package as an example, we input the Trajectory files that include the structure information. The script first identifies the adsorbates and substrates by searching every atoms and then determine the atoms at adsorption site and nearest neighbor according to the position of the adsorbates. Based on the collected information of active centers, the ψ , \overline{CN} and their coupling terms are calculated.

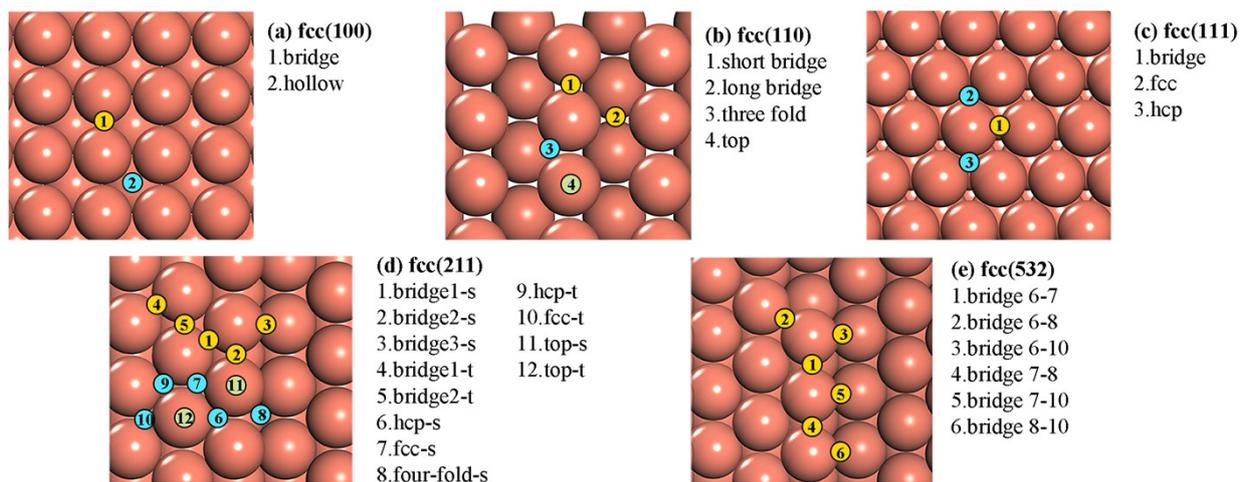


Figure S1. Overview of structures used for the considered adsorption sites on (a) the (100) facet, (b) the (110) facet, (c) the (111) facet, (d) the (211) facet and (e) the (532) facet.

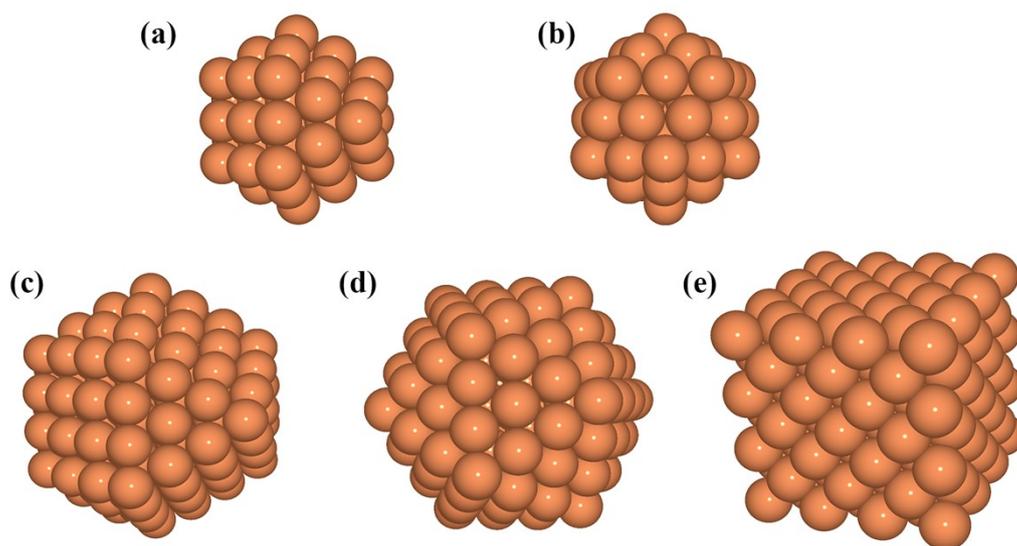


Figure S2. The structures of nanoparticles. (a) 55-atom Cuboctahedron. (b) 55-atom Icosahedron. (c) 147-atom Cuboctahedron. (d)147-atom Icosahedron. (e)172-atom Cube.

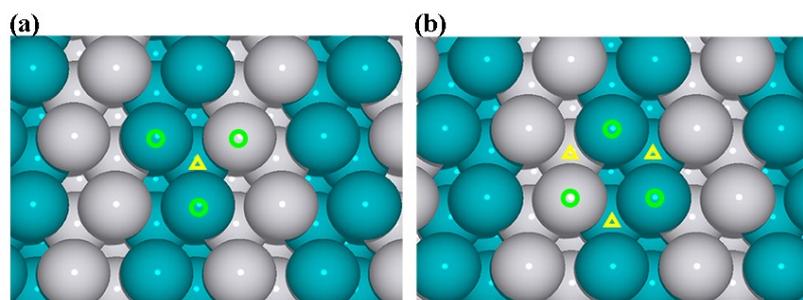


Figure S3. The central atoms (the adsorption sites, represented by the green empty circles) and the atoms with more bonding numbers with central atoms (the atoms in category C₂, represented by the yellow empty triangles) for fcc (a) and hcp (b) site.

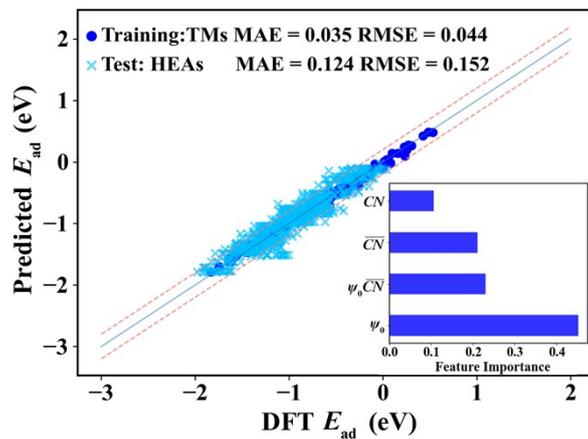


Figure S4. The transferability of our models in predicting the adsorption energies of HEAs by training the data of TMs only based on four descriptors ψ_0 , \overline{CN} , CN and $\psi_0 \overline{CN}$.

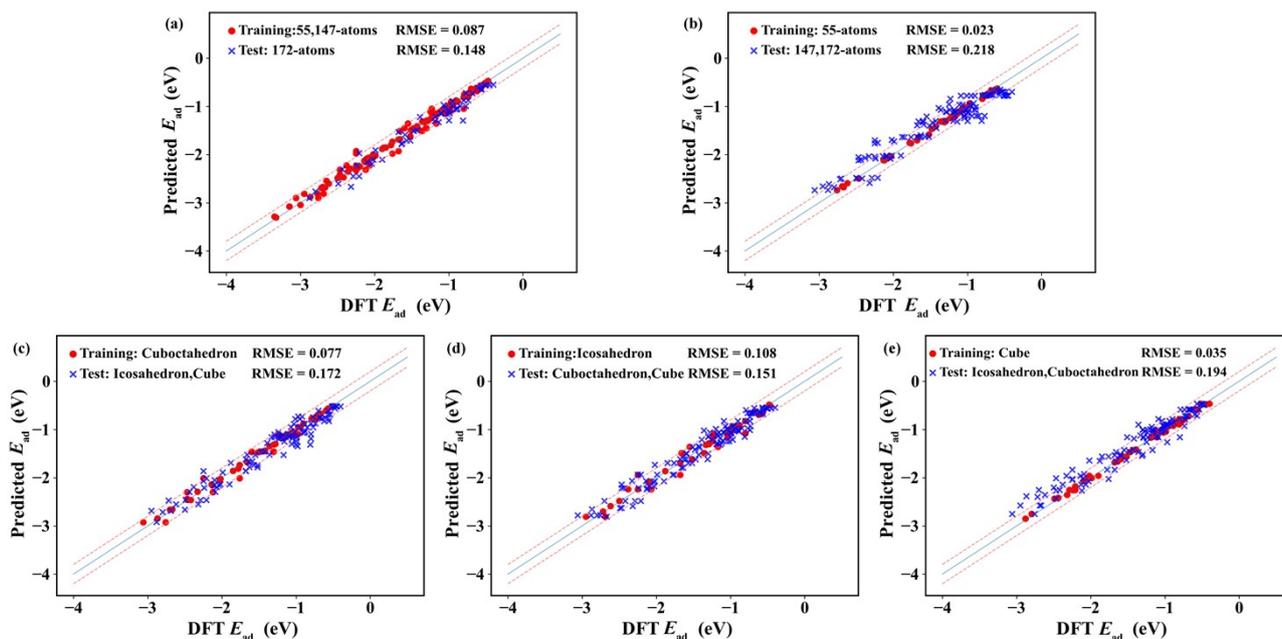


Figure S5. The mutual predictions for different sizes and morphologies of nanoparticles (NPs). (a) shows the performance to predict the adsorption energies of 172-atoms NPs by training the properties of 55- and 147-atoms clusters. (b) shows the performance to predict the adsorption energies of the 147- and 172-atoms NPs by training the properties of 55-atoms NPs. (c), (d) and (e) shows the performance to predict the adsorption energies of two types of NPs by training the properties of only one types of NPs.

Supplementary Table S1. The results of ML model on the datasets of TMs, SAAs and ABs with various training sizes. The training scores and test scores are obtained based on 5-fold cross-validation (CV). The 5-fold CV scores and 10-fold CV scores on the training sets are also listed.

Training Sizes	Training Scores	Test Scores	5-fold CV Scores	10-fold CV Scores
5%	1.000	0.714	0.588	0.253
10%	1.000	0.819	0.752	0.737
15%	0.999	0.874	0.839	0.839
20%	0.999	0.913	0.887	0.894
25%	0.998	0.940	0.914	0.921
30%	0.997	0.952	0.925	0.932
35%	0.997	0.960	0.944	0.950
40%	0.997	0.965	0.952	0.956
45%	0.997	0.969	0.959	0.963
50%	0.997	0.973	0.965	0.967
55%	0.997	0.974	0.968	0.971
60%	0.997	0.976	0.971	0.973
65%	0.997	0.978	0.974	0.975
70%	0.997	0.979	0.975	0.977
75%	0.998	0.981	0.976	0.978
80%	0.997	0.981	0.978	0.979
85%	0.998	0.982	0.979	0.980
90%	0.998	0.983	0.980	0.981
95%	0.998	0.984	0.981	0.982

Supplementary Table S2. The results of ML model on the datasets of HEAs with various training sizes. The training scores and test scores are obtained based on 5-fold cross-validation (CV). The 5-fold CV scores and 10-fold CV scores on the training sets are also listed.

Training Sizes	Training Scores	Test Scores	5-fold CV	10-fold CV
5%	0.998	0.893	0.860	0.828
10%	0.996	0.927	0.912	0.907
15%	0.993	0.940	0.931	0.929
20%	0.990	0.948	0.939	0.939
25%	0.988	0.952	0.946	0.946
30%	0.986	0.954	0.950	0.950
35%	0.985	0.956	0.953	0.954
40%	0.983	0.958	0.954	0.954
45%	0.982	0.958	0.956	0.956
50%	0.981	0.960	0.957	0.957
55%	0.981	0.960	0.958	0.959
60%	0.980	0.961	0.959	0.959
65%	0.979	0.961	0.960	0.960
70%	0.979	0.963	0.960	0.960
75%	0.978	0.962	0.961	0.961
80%	0.978	0.963	0.961	0.961
85%	0.977	0.963	0.962	0.962
90%	0.977	0.963	0.962	0.962
95%	0.977	0.962	0.962	0.963

Supplementary Table S3. Comparison between the DFT-calculated ΔE_{ad} and the predicted ΔE_{ad} by our machine learning scheme for CO. Cu(100) is selected as the reference to calculate ΔE_{ad} .

Alloys	Predicted ΔE_{ad} (eV)	DFT-calculated ΔE_{ad} (eV)	Error (eV)
Cu@Ag(211)	0.002	-0.102	0.104
Cu@Au(211)	-0.043	-0.083	0.040
Cu(100)	0	0.014	0.014
Cu@Ni(211)	-0.089	-0.157	0.068
Cu@Ir(211)	-0.128	-0.120	0.008
Cu@Pd(211)	0.022	-0.210	0.232
Cu@Pt(211)	-0.037	-0.172	0.135
Cu@Rh(211)	0.022	-0.141	0.163
CuAg(211)	-0.089	-0.090	0.001
CuAu(211)	-0.017	-0.059	0.042
CuIr(211)	0.009	-0.062	0.071
CuPd(211)	-0.095	-0.198	0.103
CuPt(211)	0.015	-0.138	0.153
CuRh(211)	-0.095	-0.169	0.074
HEA(100)-1	-0.007	-0.025	0.018
HEA(100)-2	-0.007	-0.128	0.119
HEA(100)-3	-0.007	-0.027	0.020
HEA(100)-4	-0.007	-0.031	0.024
HEA(100)-5	-0.007	0.004	0.011

References

1. M. Andersen; S. V. Levchenko; M. Scheffler; K. Reuter, *Acs Catal.* 2019, **9**, 2752-2759.
2. Z. L. Lu; Z. W. Chen; C. V. Singh, *Matter* 2020, **3**, 1318-1333.
3. J. Dean; M. G. Taylor; G. Mpourmpakis, *Sci. Adv.* 2019, **5**, eaax5101.
4. L. T. Roling; F. Abild-Pedersen, *Chemcatchem* 2018, **10**, 1643-1650.
5. A. H. Larsen; J. J. Mortensen; J. Blomqvist; I. E. Castelli; R. Christensen; M. Dulak; J. Friis; M. N. Groves; B. Hammer; C. Hargus; E. D. Hermes; P. C. Jennings; P. B. Jensen; J. Kermode; J. R. Kitchin; E. L. Kolsbjerg; J. Kubal; K. Kaasbjerg; S. Lysgaard; J. B. Maronsson; T. Maxson; T. Olsen; L. Pastewka; A. Peterson; C. Rostgaard; J. Schiøtz; O. Schütt; M. Strange; K. S. Thygesen; T. Vegge; L. Vilhelmsen; M. Walter; Z. H. Zeng; K. W. Jacobsen, *J. Phys-Condens. Mat.* 2017, **29**, 273002-273032.