

Supplementary Information

Exploring the physical origin of amorphous alloy catalyst
by using Machine Learning accelerated DFT study

Siyan Gao^{1#}, Huijie Zhen^{1#}, Bo Wen¹, Jiang Ma², Xi Zhang^{1,2*}

¹Institute of Nanosurface Science and Engineering, Guangdong Provincial Key Laboratory of Micro/Nano, Shenzhen University, Shenzhen 518060, China.

²College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China.

These authors contributed equally to this work.

* E-mail: zh0005xi@szu.edu.cn

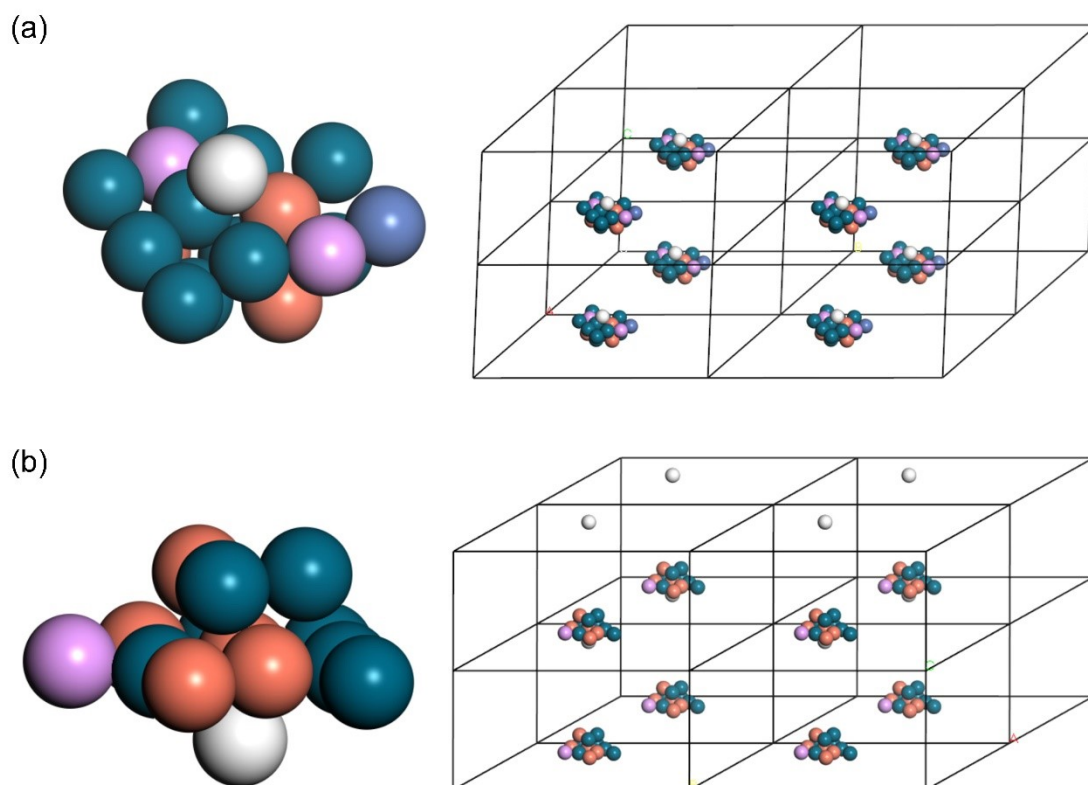


Fig. S1 (a) The schematic diagram of the adsorption of hydrogen atoms on the upper surface. (b) The schematic diagram of hydrogen atoms adsorbed on the lower surface. The white balls are the hydrogen atoms adsorbed on the surface of the amorphous alloy.

The amorphous alloy is spatially periodic. Hence, we need to expand the unit cell into supercell before sampling, otherwise there are errors occurring when sampling at the edges of the amorphous alloy. We extended the unit cell in the X and Y axes to make the substrate continuous when sampling at edges. Since the catalytic performance of the amorphous alloy is closely related to the local atomic environment, we uniformly sampled 270 adsorption sites on upper and lower surfaces of $\text{Pd}_{40}\text{Ni}_{10}\text{Cu}_{30}\text{P}_{20}$ amorphous alloy.

Here, we set α on the diagonal of the kernel function to be in the range of $10^{-3} \sim 4 \times 10^{-2}$, and the step size is 0.0026. The amplitude factor σ_f is set to $10^{-1} \sim 10^2$, and its initial value is 10. The characteristic length factor σ_l is set to $1 \sim 10^3$, and its initial

value is 80. When α is given, the values of σ_f and σ_l can be obtained through the L-BFGS algorithm optimization whose number of optimization iterations is set to 10. Fig. S2 shows the search results. As the value of α increases (adding noise to the model to prevent over-fitting), the training error of the model gradually increases, and the test error gradually decreases. When $\alpha = 0.0114$, it achieves balance where the corresponding $\sigma_f = 10.9$ and $\sigma_l = 326$. This means that the model achieves the optimal performance and has good generalization ability.

We employed our home-made code (Chinese software No. 2020SR0773207) to perform machine learning, including Atomic Simulation Environment (ASE) and Dscribe packages.

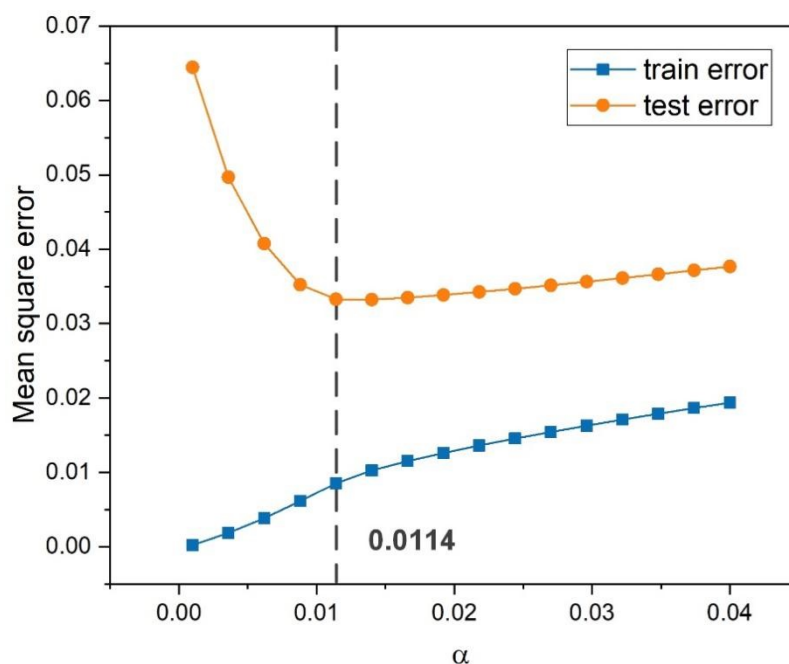


Fig. S2 Gaussian process regression model hyperparameter adjustment process.

Table S1 The representative feature data set was extracted using Smooth Overlap of Atomic Positions (SOAP) method

	0	1	2	3	4	5	6
0	0.028129	0.341014	0.780525	0.815767	4.134202	9.462517	9.889763
1	0.028129	0.341014	0.780525	0.815767	4.134202	9.462517	9.889763
2	0.028129	0.341014	0.780525	0.815767	4.134202	9.462517	9.889763
3	0.028129	0.341014	0.780525	0.815767	4.134202	9.462517	9.889763

