Supporting Information for

Efficient mapping of CO adsorption on $Cu_{1-x}M_x$ bimetallic alloys via Machine Learning

Mattia Salomone,*[a] Michele Re Fiorentin,[a] Francesca Risplendi,[a] Federico Raffone,[a] Timo Sommer,[b] Max García-Melchor,*[b] and Giancarlo Cicero*[a]
[a] *Dipartimento di Scienza Applicata e Tecnologia, Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy*
[b] *CRANN and AMBER Research Centres, School of Chemistry, Trinity College Dublin, College Green, Dublin 2, Ireland.*

1 Feature sensitivity analysis

In Figure S1, we provide examples illustrating the relationships between CO adsorption energies and the chemical properties of guest species. We note that in the region surrounding the guest species (dark blue and blue dots) there is a correlation between the features and ΔE_{CO}, while far from the impurity no visible trend appears. This behavior is consistent with the one observed in Figure 6 of the manuscript.
We also evaluate the impact of feature selection on the model behaviors and in Table S1 we report a comparison of the performance of the selected algorithms (GBC for classification and GBR for regression) when considering different numbers of features.

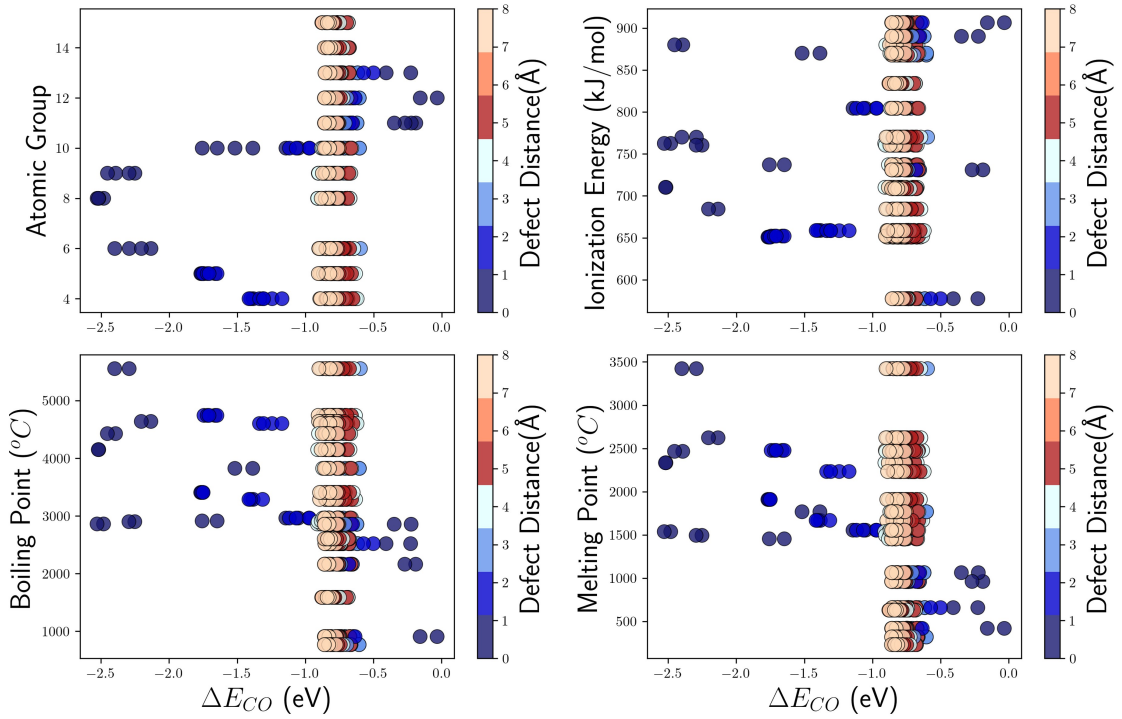

Figure S1: CO adsorption energies as a function of four chemical features of the guest species, namely AtGroup, IonEn, BoilPoint and MeltPoint. The color bar refers to the distance of the adsorption site from the M atom

Model	**All Features**	**7 Features**
GBC (min)	0.915	0.929
(maj)	0.985	0.987
GBR (R^2)	0.970	0.978
(MSE)	0.003	0.002

Table S1: Summary of F1 scores obtained for GBC, as well as MSE (in eV^2) and R^2 scores for GBR. The scores are presented for both complete (on the left) and reduced (on the right) feature vectors.

2 Cluster Expansion

The Cluster Expansion (CE) is performed by means of the Alloy Theoretic Automated Toolkit (ATAT) software package [1] coupled with Large-scale Atomic Molecular Massively Parallel Simulator [2–4] (LAMMPS). The use of LAMMPS ensures a faster evaluation of the alloy phase diagram. Additional information regarding the theoretical foundation and implementation of the CE can be found in our previous works. [5–7] Copper-based bimetallic alloys are described by using an EAM potential, parameterized as reported in reference, [8] and the minimization process is carried out using the Conjugate Gradient algorithm. Since we exclusively consider superficial substitutional impurities during the training process, the CE involves only the topmost layer of the alloy. The accuracy of the EAM potential is verified by comparing the formation energies of Au and Ag substitutional impurities on the surface, calculated using the EAM potential, with those obtained from Density Functional Theory (DFT). We observe a good agreement between the two sets of data, with an average Root Mean Square Error (RMSE) of 0.064 eV. For the comparison of the

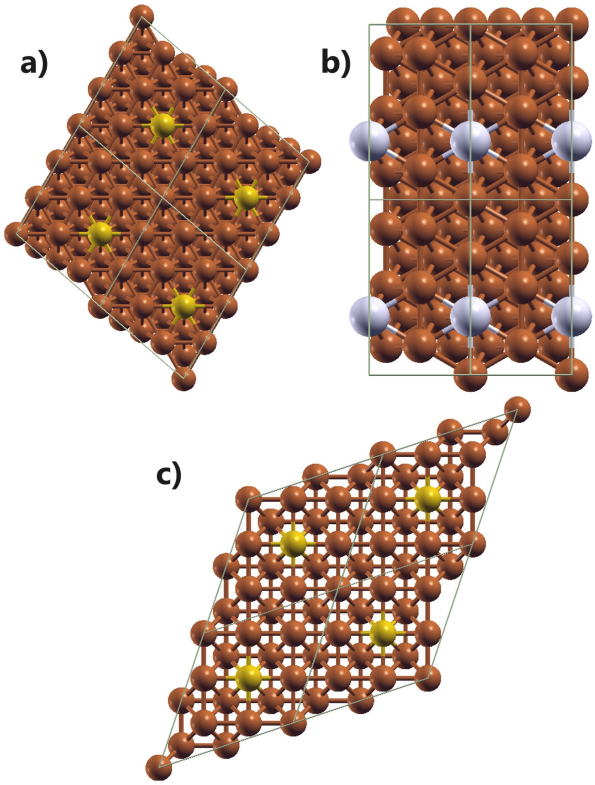

Figure S2: Structures obtained with the CE of the $Cu_{0.889}Au_{0.111}(111)$ (a), $Cu_{0.833}Ag_{0.167}(111)$ (b) and $Cu_{0.875}Au_{0.125}(100)$ (c) alloys. Color code: Cu (orange), Au (yellow) and Ag (light gray)

formation energies, we use the same structure employed in DFT calculations, i.e. a 4-layers thick slab with 6×6 surface supercell. In the CE, we add 6 layers to the original structure, resulting in a 10-layers thick slab with 6×6 surface supercell. Then we calculate the energies of the structures automatically generated by ATAT. To ensure coverage of impurity concentrations in the phase diagram that are not typically addressed by ATAT, we also provide ad hoc structures to the algorithm. The CE is performed for an impurity concentration interval corresponding to $0 \leq x \leq 0.3$ for both $Cu_{1-x}Ag_x$ and $Cu_{1-x}Au_x$ alloys. This range is chosen because it extends to concentration values higher than those used to train the ML algorithm, but at the same time it ensures that the surface does not undergo significant structural rearrangement typical of strained surface layers.
In Figure S2 we report the three structures obtained with the CE that we used as test set for our ML models, while in Figure S3 we show the CO adsorption energy maps on the $Cu_{0.889}Au_{0.111}(111)$ as a representative case. The stability of the different adsorption sites is predicted with an accuracy of 100% and the CO binding energy forecast is very precise (this is confirmed by the parity plot reported in Figure 7, red dots).

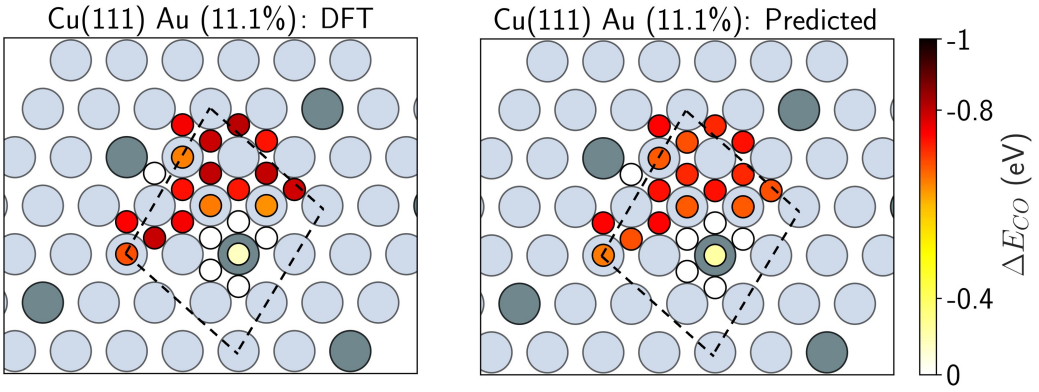

Figure S3: DFT-calculated (left) and ML-predicted (right) CO adsorption energy maps over the $Cu_{0.889}Au_{0.111}(111)$ surface.

References

[1] A. van de Walle, G. Ceder, *J Phase Equilib* **23**, 348 (2002).

[2] A. P. Thompson, *et al.*, *Comp. Phys. Comm.* **271**, 108171 (2022).

[3] F. Raffone, C. Ataca, J. C. Grossman, G. Cicero, *The Journal of Physical Chemistry Letters* **7**, 2304 (2016). PMID: 27225447.

[4] F. Raffone, F. Savazzi, G. Cicero, *Phys. Chem. Chem. Phys.* **23**, 11831 (2021).

[5] M. Salomone, F. Raffone, M. Re Fiorentin, F. Risplendi, G. Cicero, *Nanomaterials* **12**, 2079 (2022).

[6] F. Raffone, F. Savazzi, G. Cicero, *The Journal of Physical Chemistry Letters* **10**, 7492 (2019). PMID: 31735028.

[7] V. Ankit Kumar, R. Federico, C. Giancarlo, *Nanomaterials and Nanotechnology* **10**, 1847980420955093 (2020).

[8] X. W. Zhou, J. R. A., W. H. N. G., *Phys. Rev. B* **69**, 144113 (2004).