# Optimizing Active Learning Strategies for Neural Network Potentials in Catalyst Characterization Workflows

**Pol Sanz**[1], Zan Lian[1] , Nuria López[1]
[1] ICIQ – Institute of Chemical Research of Catalonia, Avinguda dels Països Catalans 16, Tarragona, Spain.

psanz@iciq.es

Supported clusters, with the cluster and the support being either metals or oxides, are an active area of research in catalysis[1,2], such as in the case of ZnO and Cu surfaces and nanoparticles, our target systems. Computer simulations can provide valuable information, but they face challenges such as the system size, which limits the use of accurate but demanding electronic structure methods. On the other hand, machine-learning potentials (MLPs) could bridge the complexity gap between electronic structure methods such as density functional theory (DFT) and traditional empirical potentials, if developed with enough accuracy and reliability by building upon datasets obtained from numerous electronic-structure calculations. During this project, we will leverage this modern methodology to generate large sets of structures in order to study more realistic structures of catalysts and replicate phase diagrams in order to provide insights into the catalytic activity of the material. A Python package has been developed in order to generate the initial training data. The data is generated following the phase diagram of the alloys, gathering template structures for each phase of the material from the Materials Project dataset[3]. Parting from these templates, the code then generates new structures by applying atomic replacements, thus changing the composition of the material. This is followed by random perturbations of the coordinates, which allow to obtain structures from different regions of the potential energy surface (PES) of the material and by applying small displacements to the lattice, which result in structures close to PES minima. Around 10000 initial structures are generated. The potential energy and forces of the structures were determined through DFT calculations using the VASP 5.4.4 code employing PBE pseudopotentials with DFT-D3 as the dispersion correction, and settings determined through a benchmarking process that resulted in a difference of energy per atom smaller than 1 meV/atom between different settings. Two libraries for the construction of NNPs, n2p2[4] and MACE[5] are tested. The potentials are trained using the generated initial database, and then a committee training procedure follows, where 4 different NNPs are trained using an active learning (AL) loop to generate new structures where there is a large discrepancy in prediction between the potentials using several extrapolation checking techniques. The resulting NNP is used to drive neural-network based molecular dynamics simulations (NN-MD) at different temperatures and timescales. The aiida[6] workflow library is used to automate the structure generation, extrapolation check, calculation submission and AL loop. In this study, large-scale NN-MD simulations were performed to understand the behavior of the the CuZn alloy as a catalyst for $CO_2$ hydrogenation. This potential was devised gradually, starting from pure Cu and CuZn and later adding ZnO. The implementation of this potential demonstrated excellent performance when used to simulate the behavior of the CuZn alloy. The phase diagram of the alloy was predicted based on the results of MD simulations simulations at different conditions, obtaining results that match the expected shape of the diagram. The most stable structure of the CuZn alloy was obtained from MD simulations ran at large spatial and temporal scales. The most common ensembles were identified by coordination environment and graph theory analysis. The study's findings reveal the structure of CuZn, providing a substantial basis for further investigations into the creation mechanism of long-chain hydrocarbons.

## References

[1] Amann, P. et al. Science 376, 603–608 (2022).
[2] Halim, H. H. & Morikawa, Y. ACS Phys. Chem Au 2, 430–447 (2022).
[3] Jain, A. et al. APL Materials 1, 011002 (2013).
[4] Singraber, A., Behler, J. & Dellago, C. J. Chem. Theory Comput. 15, 1827–1840 (2019).
[5] Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C. & Csányi, G. arXiv (2023).
[6] Huber, S. P. et al. Sci Data 7, 300 (2020).

## Figures

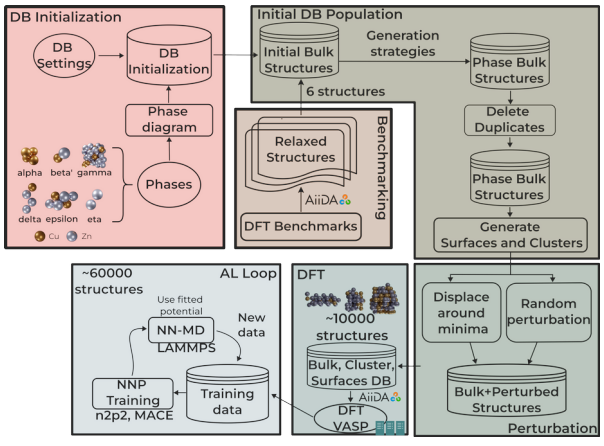


**Figure 1.** Diagram depicting the active learning procedure.