

# MULTI-FIDELITY MACHINE LEARNING TO ACCELERATE MATERIALS RESEARCH

M. Kuchelmeister<sup>1</sup>, J. Löfgren<sup>1</sup>, M. Todorović<sup>2</sup>, and P. Rinke<sup>1</sup>

<sup>1</sup>*Department of Applied Physics, Aalto University*

<sup>2</sup>*Department of Mechanical and Materials Engineering, University of Turku*

email: manuel.kuchelmeister@aalto.fi

The computational optimization and exploration of materials is a challenging task, due to the high dimensionality of the search space and the high cost of accurate quantum mechanical calculations. To reduce the number of costly calculations, the Bayesian Optimization Structure Search (BOSS) [1] has been developed. BOSS combines sample-efficient active learning with Gaussian process regression.

We here present a multi-fidelity approach (Fig. 1a) that can reduce the number of costly, accurate calculations even further by incorporating information from inexpensive but less accurate calculations. We implemented the intrinsic model of coregionalization [2] method into BOSS to sample data from multiple atomistic calculations based on quantum chemistry (Gaussian 16, using CCSD(T)), density-functional theory (FHI-aims, using a PBE exchange-correlation functional) and force fields (AMBER18). Multi-fidelity BOSS is initialized with lower-fidelity data and continues to sample only higher-fidelity calculations, maintaining CCSD(T) accuracy for the global minimum inference.

We tested our multi-fidelity model on 4D alanine conformer search (Fig. 1b). The efficiency of our approach is measured by computational cost in CPU hours, comparing runs with and without lower-fidelity (LF) data (Fig. 1c). We were able to reduce the computational cost for a CCSD(T) run, when using DFT as LF data, by about 70%. We found that the efficiency of the ICM model depends on both the correlation and computational cost difference between the fidelities, as well as the number of LF data. Our test serves as a first benchmark for the great potential that multi-fidelity learning can have to reduce expensive structure-search problems.

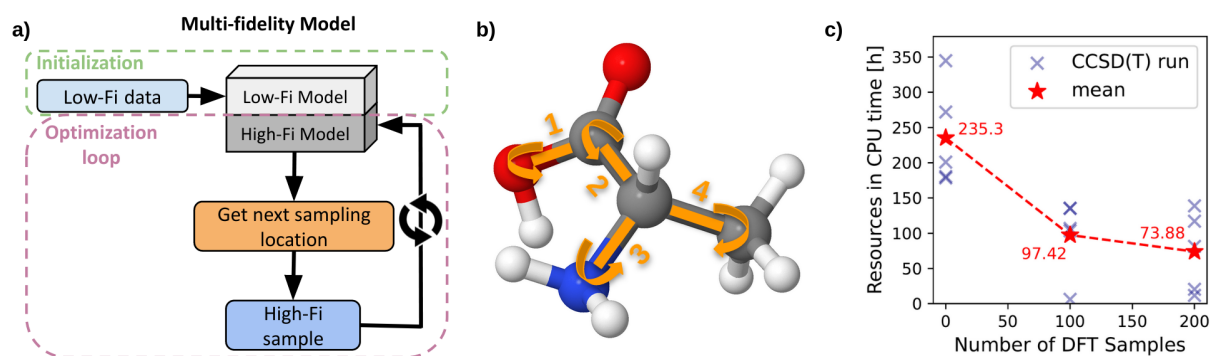


Figure 1: a) Multi-fidelity BOSS framework. b) Atomistic model for the 4D alanine conformer search. c) Reduction in computational cost for CCSD(T) structure search with DFT as LF data.

[1] M. Todorović, M. U. Gutmann, J. Corander, and P. Rinke, *npj Comput. Mater.* **5**, 35 (2019).

[2] E.V. Bonilla et al., *Advances in Neural Information Processing Systems* pp.153-160 (2007).