Multi-fidelity Bayesian optimisation for DFT-in-the-loop discovery of manufacturable homogeneous catalysts, auxiliaries, and ligands.

Yong Lee[1] Ehecatl Antonio del Rio Chanona[2] King Kuok (Mimi) Hii[1, 3]
[1] *Department of Chemistry, Molecular Science Research Hub, Imperial College London, United Kingdom*
[2] *Department of Chemical Engineering, South Kensington, Imperial College London, United Kingdom*
[3] *A*STAR Institute of Sustainable Chemicals Energy & Environment, Singapore*

Correspondence to: ***King Kuok (Mimi) Hii*** *mimi_hii@a-star.edu.sg*

1. Introduction

Density functional theory (DFT) has long been the workhorse for mechanistic analysis in homogeneous catalysis, yet it remains underutilized as a computational, predictive design tool for catalysts, auxiliaries, and ligands (LACs). The bottleneck is not only the absolute cost of computation, but also the combinatorial growth of synthetically plausible LAC space that is coupled to multi-criterion performance requirements (for example, reactivity, selectivity etc.). Recent AI-guided chemistry workflows have demonstrated the effectiveness of Bayesian optimisation (BO) and closed-loop decision-making, but most implementations either focus on experimental condition optimisation or rely on single-fidelity surrogate screening that struggles to scale to mechanistically grounded barrier predictions with DFT workflows.

2. Substantial section

We formulate LAC discovery as a mechanism-constrained, multi-objective optimisation problem: identify manufacturable LAC structures that minimise predicted kinetic barriers ($\Delta G^{\ddagger}$) and optimise thermodynamic driving forces (ΔG) within an a priori mechanistic manifold, while optionally penalising known failure modes (for example, off-cycle binding, catalyst degradation etc.).

To make DFT screening practical for LAC discovery, we introduce an asynchronous, multi-objective, multi-fidelity, multi-batch Bayesian optimisation framework that is tightly integrated with automated quantum chemistry calculations where input structures and job files are automatically generated. The central idea is to treat quantum mechanical evaluations of varying levels of theory as fidelities. For example, faster but more approximate methods (DFT with low level of theory with smaller basis sets or semiempirical tight-binding) as low-fidelity outputs, and higher-accuracy DFT as the high-fidelity outcome/ground truth. The optimiser is explicitly cost-aware, allocating compute to the cheapest evaluations that are most informative until uncertainty warrants promotion to higher-fidelity calculations. This yields a pragmatic "DFT-in-the-loop" engine that (i) reduces expensive high-fidelity DFT calculations, (ii) shortens overall calculation runtime through asynchronous scheduling, and (iii) returns a Pareto front of LAC candidates that reflect trade-offs rooted in explainable chemical theory rather than a single optimum.

2.1 Related work

BO has become a central tool in chemical discovery, especially in closed-loop experimental optimisation and self-driving lab settings, but these typically optimise reaction conditions and are not designed to natively couple with mechanistic DFT barrier evaluation with multi-fidelity cost control and multi-objective trade-offs.

On the computational chemistry side, DFT-guided active-learning workflows have been increasingly reported for catalyst/material screening. However, they often employ single-fidelity models, target one objective, or proceed synchronously, which limits throughput when barrier calculations fail, queue unpredictably on a high-performance computing cluster, or vary widely in runtime. In contrast, our framework combines (1) multi-fidelity decision-making, (2) parallel multi-objective BO, and (3) asynchronous execution in an end-to-end pipeline purpose-built for homogeneous LAC discovery with explicit manufacturability constraints.

3. Impact

By converting DFT from a post hoc explanatory tool into a budget-aware, autonomous design loop, this approach enables scalable identification of computational hits for (i) novel LAC design, (ii) substrate-scope expansion, and (iii) substrate-specific LAC tailoring. This work directly aligns established methods of mechanistic elucidation and interpretation with modern AI optimisation.

Acknowledgments

The authors gratefully acknowledge Imperial College London for financial support through

the President's Scholarship awarded to the primary author.

References

[1] J. Edmund, A. Mohammed, M. M. Austin, c. Antonio Del rio and E. J. Kim, presented in part at the AI for Accelerated Materials Design - NeurIPS, 2024.

[2] Y. Wu, A. Walsh and A. M. Ganose, ***Digital Discovery***, 2024, **3**, 1086-1100.

[3] L. Morán-González, A. L. Burnage, A. Nova and D. Balcells, ***ACS Catal***, 2025, **15**, 9089-9105.