# PlateOpt: Bayesian Optimization for Organic Catalysis in Combinatorial Well Plates

Florian Boser[*1] Felix Katzenburg[*1] Joshua W. Sin[*2] Bojana Ranković[2] Sarina Kopf[2] Philippe Schwaller[2] Frank Glorius[1]

[*]*Equal contribution* [1]*[Organisch-Chemisches-Institut, Universität Münster, Corrensstraße 36, 48149 Münster, Germany* [2]*[Laboratory of Artificial Chemical Intelligence (LIAC), National Centre of Competence in Research (NCCR) Catalysis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland]*

*Correspondence to: [Florian Boser]* florian.boser@uni-muenster.de

## 1. Introduction

Reaction optimization remains a cornerstone of organic synthesis, yet it has long been bottlenecked by One-Factor-at-a-Time (OFAT) methodologies. While conceptually simple, OFAT fails to capture synergistic interactions between reaction variables, such as the non-linear interdependencies between ligand, base, and solvent. To address this, the field has pivoted toward Design of Experiments (DoE) and Machine Learning (ML) techniques to navigate chemical space more efficiently. This transition is critical because a typical cross-coupling reaction space—spanning catalyst precursors, ligands, bases, solvents, and temperatures—can easily exceed $10^6$ discrete combinations. In this high-dimensional search space, Bayesian Optimization (BO) has emerged as the gold standard, utilizing probabilistic surrogate models (typically Gaussian Processes) to direct the search toward global optima with high sample and iteration efficiency.[1]

Alongside these algorithmic advances, High-Throughput Experimentation (HTE) has risen as the laboratory solution to the "curse of dimensionality".[2] By miniaturizing and parallelizing reactions, HTE allows for the dense exploration of chemical space in a resource-efficient manner.[3] Furthermore, the structured, metadata-enriched datasets generated by HTE have proven uniquely suitable for training predictive molecular ML models.[3–5]

Recently, Sin et al. bridged the gap between BO and HTE with the *Minerva* platform, utilizing a multi-objective acquisition function (qNEHVI) to optimize yield and selectivity simultaneously.[6] However, a critical limitation persists: treating each well as an independent experiment rather than as part of a coordinated HTE plate design results in a logistical complexity that grows rapidly. This "unconstrained" approach, while dense in information, necessitates either discrete solid-dosing for every well or an intractable number of unique stock solutions. Such experimental workflows remain prohibitively expensive for most academic facilities. While the authors successfully implemented batch constraints for temperature (necessitating separate plates), the granular logistical burdens of liquid handling and reagent preparation remain unsolved.

In practice, the majority of laboratories rely on manual multichannel pipettes or affordable liquid handlers (e.g., Opentrons) for HTE. In these setups, the bottleneck is not only temperature, but also stock solution preparation and management. To maintain throughput, well plates are traditionally designed in grids, conserving stock solutions row- or column-wise.[7] Integrating BO into this workflow poses a unique mathematical challenge: the optimal experiments in a batch are no longer independent. As illustrated in Fig. 1, selecting Ligand L1 for well A1 fixes that ligand for the entire row A. This rigid dependency structure breaks the independence assumptions underlying standard BO acquisition functions, nesting a complex combinatorial optimization problem within the BO loop.

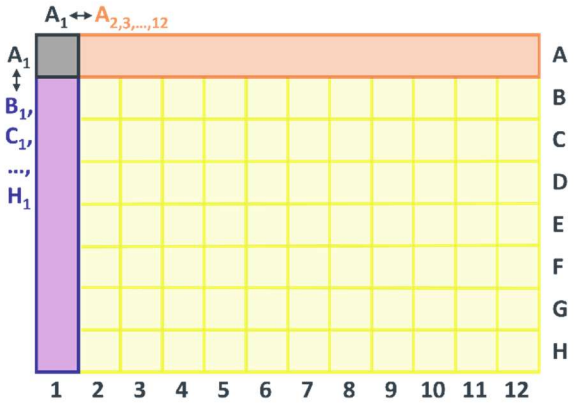


Fig. 1: Dependence of row and column wise experiments upon selection of well A1.

To resolve this, we introduce *PlateOpt*, a HTE-aware BO architecture that optimizes templates rather than individual experiments. PlateOpt accommodates the physical realities of the laboratory, specifically the use of multichannel pipettes and limited reagent capacities. By restricting designs to a maximum of 30 stock solutions per iteration, our approach accelerates

physical setup and minimizes dead-volume waste, reconciling the proven capabilities of BO with the logistical constraints of HTE.

## 2. Template-Based Bayesian Optimization

Standard Bayesian Optimization (BO) algorithms typically propose batches of $N$ experiments under the assumption of experimental independence to maximize total information gain. In contrast, PlateOpt evaluates predefined experimental templates—structured grids (e.g., 12 ligands with 8 bases or 6 bases with 2 solvents and 8 precatalysts) designed to minimize stock solution usage and setup time.

To score these templates, we employ an aggregated expected improvement strategy. Rather than evaluating a single reaction condition in isolation, we assess the potential of a specific chemical component (e.g., Ligand L1) by aggregating its acquisition function score (e.g., qNEI) across all combinatorial pairings (precatalysts, solvents, bases, temperatures) within the search space. The algorithm ranks valid plate layouts based on these combined scores and selects the template that maximizes collective batch improvement while strictly adhering to a defined stock solution budget (e.g. 30).

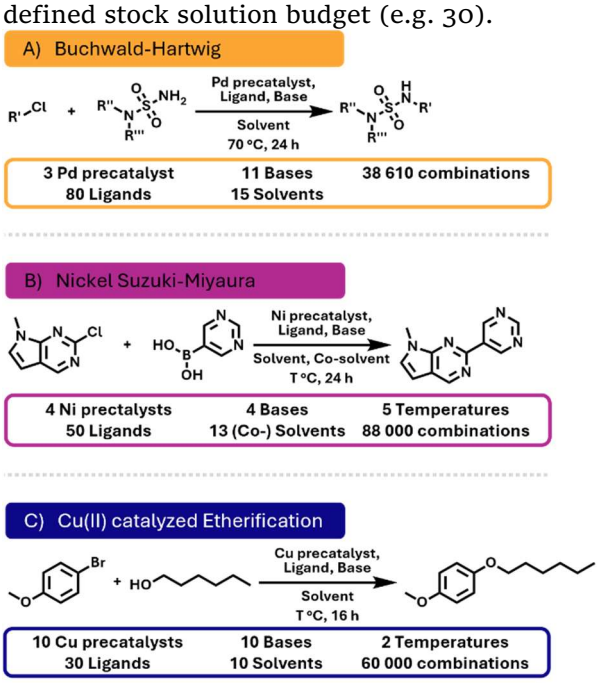


Fig. 2: Organic catalytic reactions optimized with PlateOpt.

### 2.1 Computational Validation

We validated PlateOpt using existing HTE datasets for a Pd-catalyzed Buchwald-Hartwig amination (38 610 combinations) and a Ni-catalyzed Suzuki-Miyaura coupling (88 000 combinations). The virtual campaigns were seeded with 12 random unconstrained experiments to gather initial diversity before switching to template based HTE optimization. The defining advantage of PlateOpt is visualized in Fig. 3B. While standard unconstrained BO requires an average of above 150 unique stock solutions per iteration, PlateOpt remains strictly bounded at 30 unique stock solutions. This represents a >80% reduction in stock solution preparation, significantly mitigating dead-volume waste and enabling rapid manual or automated setup. In terms of optimization performance (Fig. 3A), unconstrained BO achieves a marginally higher average normalized yield after 6 iterations due to its freedom to explore more diverse regions of the landscape. However, PlateOpt closely trails this performance, demonstrating that the restriction to combinatorial grids does not prevent the algorithm from locating high-performing regions, effectively trading a minimal delay in convergence for a massive gain in experimental feasibility.

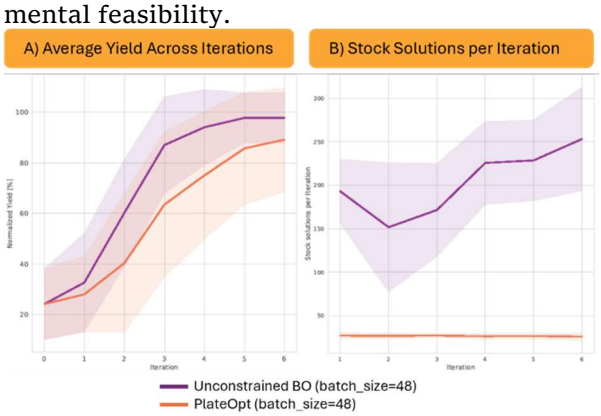


Fig. 3: Computational validation of PlateOpt on a Buchwald-Hartwig coupling.

### 2.2 Experimental Validation

Moving from *in-silico* validation to physical application, we deployed PlateOpt to optimize the C–O coupling of aryl bromides (Fig. 2C). We selected a Cu(II)-catalyzed etherification as a case study to target sustainable, earth-abundant catalysis. Unlike standard Pd-cross couplings, Cu(II) precursors are air-stable and inexpensive, aligning with Green Chemistry principles. However, the mechanistic landscape is complex, involving a Cu(II)/Cu(III) cycle that is sensitive to ligand-base interactions. Despite these challenges, PlateOpt successfully navigated the search space while adhering to the restricted stock budget. In the exploration phase of this ongoing campaign, the algorithm identified a local optimum yielding 19% after three wellplates, validating that template-constrained BO can effectively drive discovery in complex, real-world chemical environments.

**Acknowledgments**

Generous financial support by the European Research Council (Advanced Grant Agreement No. 101098156, HighEnT.) and the Deutsche Forschungsgemeinschaft (Priority Program "Molecular Machine Learning", SPP 2363) is gratefully acknowledged.

**References**

(1) Braconi, E.; Godineau, E. Bayesian Optimization as a Sustainable Strategy for Early-Stage Process Development? A Case Study of Cu-Catalyzed C–N Coupling of Sterically Hindered Pyrazines. *ACS Sustainable Chem. Eng.* **2023**, *11* (28), 10545–10554. https://doi.org/10.1021/acssuschemeng.3c02455.

(2) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio de Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23* (6), 1213–1242. https://doi.org/10.1021/acs.oprd.9b00140.

(3) Katzenburg, F.; Boser, F.; Schäfer, F. R.; Pflüger, P. M.; Glorius, F. Calibration-Free Quantification and Automated Data Analysis for High-Throughput Reaction Screening. *Digital Discovery* **2025**, *4* (2), 384–392. https://doi.org/10.1039/D4DD00347K.

(4) Saebi, M.; Nan, B.; Herr, J. E.; Wahlers, J.; Guo, Z.; Zurański, A. M.; Kogej, T.; Norrby, P.-O.; Doyle, A. G.; Chawla, N. V.; Wiest, O. On the Use of Real-World Datasets for Reaction Yield Prediction. *Chem. Sci.* **2023**, *14* (19), 4997–5005. https://doi.org/10.1039/D2SC06041H.

(5) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360* (6385), 186–190. https://doi.org/10.1126/science.aar5169.

(6) Sin, J. W.; Chau, S. L.; Burwood, R. P.; Püntener, K.; Bigler, R.; Schwaller, P. Highly Parallel Optimisation of Chemical Reactions through Automation and Machine Intelligence. *Nat. Commun.* **2025**, *16* (1), 6464. https://doi.org/10.1038/s41467-025-61803-0.

(7) Mahjour, B.; Zhang, R.; Shen, Y.; McGrath, A.; Zhao, R.; Mohamed, O. G.; Lin, Y.; Zhang, Z.; Douthwaite, J. L.; Tripathi, A.; Cernak, T. Rapid Planning and Analysis of High-Throughput Experiment Arrays for Reaction Discovery. *Nat Commun* **2023**, *14* (1), 3924. https://doi.org/10.1038/s41467-023-39531-0.

(8) Fromer, J.; Wang, R.; Manjrekar, M.; Tripp, A.; Hernández-Lobato, J. M.; Coley, C. W. Batched Bayesian Optimization by Maximizing the Probability of Including the Optimum. *J. Chem. Inf. Model.* **2025**, *65* (10), 4808–4817. https://doi.org/10.1021/acs.jcim.5c00214.