Highly Parallel Optimisation of Nickel-Catalysed Suzuki Reactions through Automation and Machine Intelligence

Joshua W. Sin[1,2*], **Siu Lun Chau**[3], **Ryan P. Burwood**[4], **Kurt Püntener**[1], **Raphael Bigler**[1], **Philippe Schwaller**[2,5*]

[1]Process Chemistry & Catalysis, Synthetic Molecules Technical Development, F. Hoffmann-La Roche AG, Basel, Switzerland
[2]Laboratory of Artificial Chemical Intelligence (LIAC), EPFL, Lausanne, Switzerland
[3]Rational Intelligence Lab, CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
[4]Solid State Sciences, Synthetic Molecules Technical Development, F. Hoffmann-La Roche AG, Basel, Switzerland
[5]National Centre of Competence in Research (NCCR) Catalysis, EPFL, Lausanne, Switzerland

*Corresponding authors. Email: wing_pong.sin@roche.com; philippe.schwaller@epfl.ch

Abstract

We report the development and application of a scalable machine learning optimisation framework for batched multi-objective reaction optimisation. Through experimental data-derived benchmarks, we demonstrate our approach's capacity to efficiently handle large parallel batches and high-dimensional search spaces characteristic of high-throughput experimentation (HTE). We also establish the framework's robustness to reaction noise and handling of batch constraints encountered in real-world chemical laboratories. We applied our approach experimentally through an automated 96-well HTE reaction optimisation campaign for a nickel-catalysed Suzuki reaction, aiming to tackle challenges in non-precious metal catalysis. Our optimisation framework effectively navigates the complex reaction landscape with unexpected chemical reactivity, revealing advantages over traditional, purely experimentalist-driven HTE plate design. By integrating machine intelligence with highly parallel reaction execution via HTE robots, this work aims to accelerate reaction optimisation in academia and the pharmaceutical industry. This workflow can also be extended beyond HTE settings to any chemical reaction of interest.

1 Introduction

Chemical reaction optimisation is a challenging and resource-intensive yet essential process in chemistry. Chemists explore combinations of various reaction parameters (e.g., reagents, solvents, catalysts) to simultaneously optimise multiple objectives such as yield and selectivity. In process chemistry, reaction optimisation faces more rigorous demands on reaction objectives than in academic settings, encompassing additional economic, environmental, health, and safety considerations [1, 2]. These factors often necessitate the use of lower-cost, earth-abundant, and greener alternatives, such as replacing traditional palladium catalysts with nickel [3, 4], and selecting solvents that adhere to pharmaceutical guidelines [5]. Optimal reaction conditions satisfying these stringent criteria are often substrate-specific and challenging to identify generally for a given set of reactants.

Advancements in automation [6–11] have catalysed a shift in chemical reaction optimisation [12], particularly through the emergence of chemical high-throughput experimentation (HTE) adapted

Preprint. Under review.

from techniques in biology [13]. HTE platforms, utilising miniaturised reaction scales and automated robotic tools, enable parallel execution of numerous reactions. This allows for the exploration of many combinations of reaction conditions, making HTE more cost- and time-efficient than traditional techniques relying solely on chemical intuition and one-factor-at-a-time (OFAT) approaches [12, 14]. However, as additional reaction parameters multiplicatively expand the space of possible experimental configurations, exhaustive screening approaches remain intractable for larger design spaces, even with HTE [14]. Consequently, HTE chemists rely on chemical intuition to navigate vast reaction spaces effectively. A common approach is designing fractional factorial screening plates with grid-like structures [8, 15] (Fig. 1a), as used in our process chemistry HTE lab [16]. While these structures effectively distill chemical intuition into plate design, they explore only a limited subset of fixed combinations. This limitation, especially in broad reaction condition spaces, may lead experimentalists to overlook important regions of the chemical landscape.

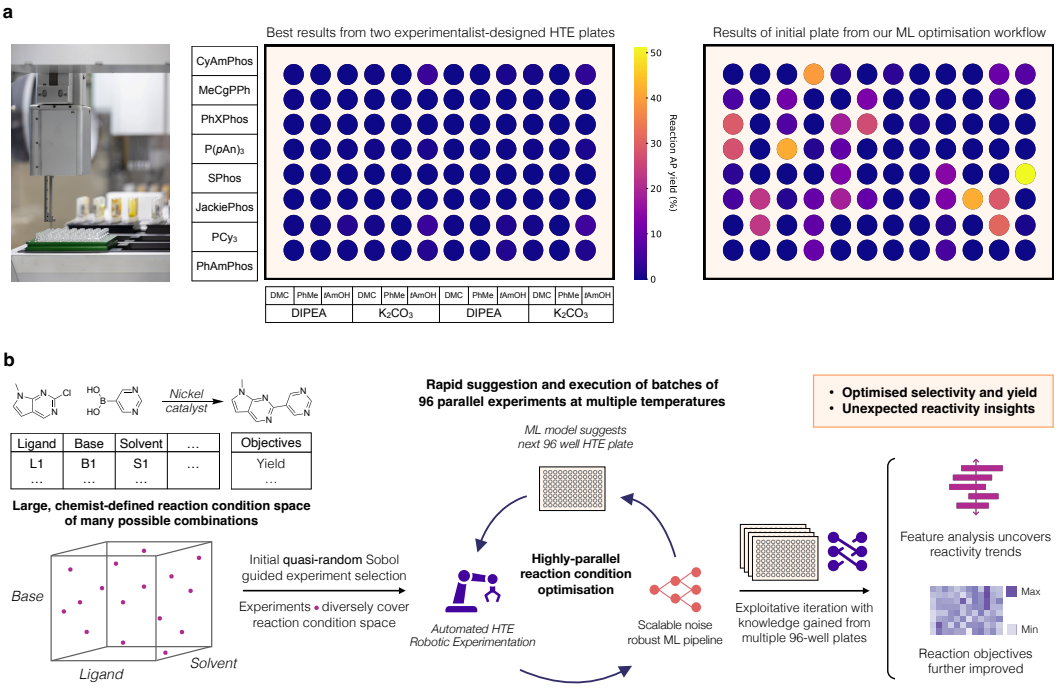

Figure 1: Strategies for HTE reaction optimisation and overview of this study. a, Comparing methods for HTE plate design. An example of a HTE plate designed with traditional fractional factorial grid-like structures and an initial HTE plate for our ML optimisation workflow. **b**, Experimental application of ML optimisation workflow to a Ni-catalysed Suzuki reaction. The experimentalist first defines promising reaction parameters (ligands, bases, solvents) comprising the reaction condition search space. The initial experiments are selected with quasi-random Sobol sampling [17], diversely sampling from the reaction condition space. Then, iterative Bayesian optimisation suggests subsequent HTE screening plates, optimising towards defined objectives. This process is typically repeated until convergence, stagnation in improvement, or exhaustion of experimental budget. Then, a fully exploitative ML approach leverages accumulated data from all HTE plates in the campaign to maximise final reaction objectives. Finally, feature analysis elucidates reactivity trends in the optimisation campaign.

From advances in computer science and statistics [18–21], machine learning (ML) techniques, particularly Bayesian optimisation, have gained popularity in chemistry for their ability to successfully guide experimental design [22–24]. Bayesian optimisation uses uncertainty-guided ML to balance exploration and exploitation of reaction spaces, identifying optimal reaction conditions in only a small subset of experiments. Bayesian optimisation has shown promising results in reaction optimisation, validated experimentally by multiple case studies [14, 25–27] and outperforming human experts in simulations [23]. However, existing applications in reaction optimisation have been largely limited to small numbers of experiments in parallel batches of up to sixteen [14, 23, 25–27]. Moreover, these approaches are often non-automated or are restricted to single reaction objectives. The constraints of small batch sizes necessitate more optimisation iterations to identify high-performing reaction configurations, particularly when exploring large reaction spaces. These limitations also impede

integration with large-scale automation. Consequently, the full potential of highly parallel, automated reaction optimisation remains underexplored. In the pharmaceutical industry, where many reactions prove unsuccessful and timelines for drug discovery and development are continuously accelerating, there is a need to expedite optimisation strategies in chemical synthesis [28, 29]. The natural synergy between ML optimisation and HTE platforms, leveraging efficient data-driven search strategies with highly parallel screening of numerous reactions, offers promising prospects for automated and accelerated chemical process optimisation.

In this work, we report the development and application of a ML-driven workflow for scalable batch reaction optimisation of multiple reaction objectives, applicable to any reaction of interest (Fig. 1b). Through in silico studies, we showed the scalability of our approach in handling large batch sizes of 96 and high-dimensional reaction search spaces of 530 dimensions. We also assessed the robustness of our workflow to chemical noise and accommodated for batch constraints typical in chemical laboratories, demonstrating its suitability for real-world applications. We applied our approach experimentally through an automated 96-well HTE optimisation campaign for a nickel-catalysed Suzuki reaction, exploring a search space of $88,000$ possible reaction conditions. Our optimisation workflow identified reactions with an area percent (AP) yield of 76% and selectivity of 92% for this challenging transformation, whereas two expert-designed HTE plates failed to find successful reaction conditions. This study demonstrated advantages of our approach over traditional, purely experimentalist-driven HTE plate design, highlighting its potential to accelerate automated reaction optimisation. The 672 HTE reactions conducted in this study are made available in the Simple User-Friendly Reaction Format (SURF) [30] with the custom code used in an open-source code repository, Minerva.

2 Results

2.1 Overview of optimisation pipeline

In our optimisation process, often involving reactions with sparse historical data, we prioritised thoroughly exploring a large set of categorical variables. From chemical experience, categorical variables such as ligands, solvents, and bases can substantially influence reaction outcomes, potentially creating distinct and isolated optima in the reaction yield landscape. Algorithmic exploration of these categorical variables enables identification of promising combinations, which can guide fine-tuning of continuous parameters such as catalyst loading and reaction time in later stages of the optimisation process. However, incorporating numerous categorical parameters increases the dimensionality and complexity of the search space, as molecular entities must be converted into numerical descriptors, unlike directly representable continuous variables. We represented the reaction condition space as a discrete combinatorial set of potential conditions composed of reaction parameters such as reagents, solvents, and temperatures deemed plausible by a chemist for a given reaction (Fig. 1b). This allowed for automatic filtering of impractical conditions such as those with reaction temperatures exceeding solvent boiling points or unsafe combinations like NaH and DMSO.

We then initiate our optimisation process with quasi-random Sobol sampling to select initial experiments, aiming to sample experimental configurations diversely spread across the reaction condition space [17] (Fig. 1b) (see Methods). By maximising reaction space coverage of the initial experiments, Sobol sampling increases the likelihood of discovering informative regions containing optima. Using this initial experimental data, we train a Gaussian Process (GP) regressor [31] to predict reaction outcomes (e.g., yield, selectivity) and their uncertainties for all reaction conditions. An acquisition function, balancing between exploration of unknown (uncertain) regions of the search space and exploitation of previous experiments, then evaluates all reaction conditions and selects the most promising next batch of experiments (see Methods for more details). After obtaining new experimental data, the chemist can choose to repeat this process for as many iterations as desired (Fig. 1b), usually terminating upon convergence, stagnation in improvement, or exhaustion of experimental budget.

2.2 Scalable multi-objective acquisition functions

In real-world scenarios, chemists often face the challenge of optimising multiple reaction objectives simultaneously, such as maximising yield while minimising cost. HTE campaigns, characterised by larger batch sizes and range of reaction parameters, further amplify optimisation complexity. Com-

putationally, scaling the parallel optimisation of multiple competing objectives towards high batch sizes is challenging and incurs considerable computational load [32]. For example, the q-Expected Hypervolume Improvement (q-EHVI) acquisition function [33], a popular multi-objective acquisition function applied previously in reaction optimisation [14], has time and memory space complexity scaling exponentially with batch size [34]. Given the limited scalability of such approaches (see Supplementary Information section 2), we sought to develop a more scalable optimisation framework for highly parallel HTE applications, incorporating several scalable multi-objective acquisition functions in our work: q-NParEgo [34], Thompson sampling with hypervolume improvement (TS-HVI) [32], and q-Noisy Expected Hypervolume Improvement (q-NEHVI) [34] (see Methods for further details).

2.3 Benchmarking and evaluating

To assess optimisation algorithm performance, practitioners often conduct retrospective in silico optimisation campaigns over existing experimental datasets [14, 23, 35, 36]. This enables comparison of optimisation performance against previously established experimental optima within a set evaluation budget. However, publicly available experimental datasets usually contain only $\sim$ 1000 reaction conditions per substrate pair, especially those with multiple reaction objectives. This limited scope is insufficient to benchmark HTE optimisation campaigns involving multiple 24/48/96-well plates. To address this limitation, we conducted benchmarks against emulated virtual datasets, following similar established practices [35, 36] (Fig. 2a).

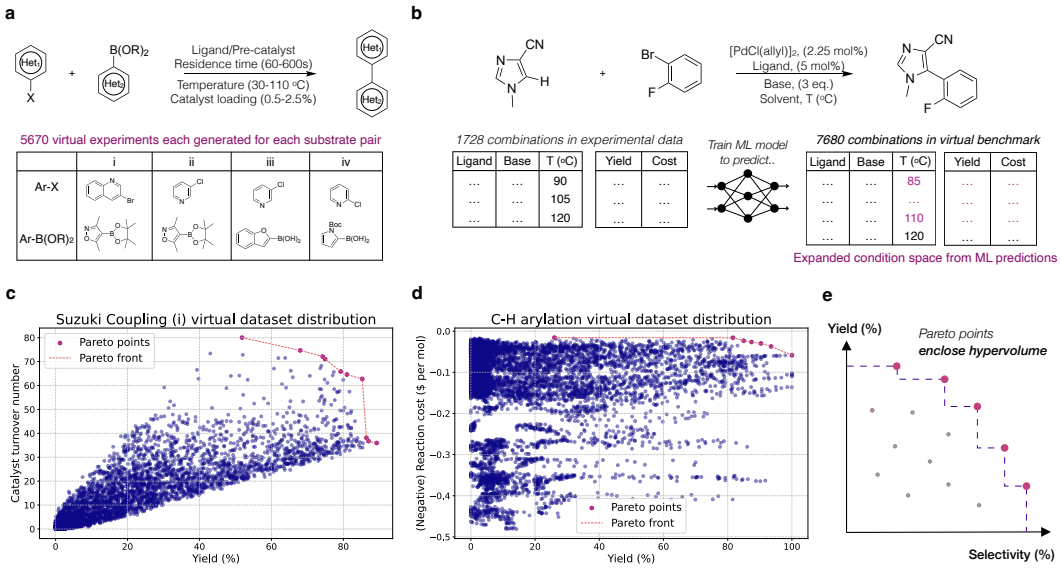

Figure 2: Benchmarking techniques and metrics to assess optimisation performance. **a**, Four Suzuki coupling virtual datasets from Olympus [35], derived from experimental data, used for benchmarking in this study. **b**, The C-H arylation virtual dataset is generated by training an ML model on 1728 experimentally collected reactions from Torres et al. [14] (EDBO+), then predicting reaction outcomes for a larger range of reaction conditions and variables not present in the original training data. This creates a large-scale virtual dataset suitable for benchmarking HTE optimisation campaigns (see Supplementary Information section 1). **c**, Distribution of reaction objectives, yield (%) and catalyst turnover number, for the first Suzuki Coupling (i) virtual dataset from Olympus [35]. Pareto points represent the optimal multi-objective combinations. **d**, Distribution of reaction objectives, yield (%) and reaction cost, for the C-H arylation virtual dataset generated in this study. **e**, The hypervolume is used to assess the performance of optimisation algorithms. The hypervolume quantifies the volume enclosed by the optimal objective (yield-selectivity) combinations (Pareto points) identified by each algorithm. The hypervolume is used to compare algorithm performance against the best reaction conditions in the benchmark dataset, evaluating how effectively the best existing solutions are identified.

We trained ML regressors on reaction datasets from Torres et al. [14] (EDBO+), using ML predictions to emulate reaction outcomes for a broader range of conditions (e.g., temperature, concentration) than present in the original experimental training data (Fig. 2b). This emulation expands smaller experimental datasets to larger-scale virtual datasets more suitable for benchmarking HTE optimisation campaigns (Fig. 2c and d). Additionally, we incorporated similar virtual datasets in our benchmarking

from Olympus [35] (Fig. 2a) (see Supplementary Information section 1 for details on all benchmark datasets). To evaluate optimisation performance, we used the hypervolume metric [37] to quantify the quality of reaction conditions identified by the algorithms. The hypervolume calculates the volume of objective space (e.g., yield, selectivity) enclosed by the set of reaction conditions selected by our algorithm (Fig. 2e). The hypervolume considers both the convergence towards optimal reaction objectives and diversity, providing a comprehensive measure of optimisation performance. We compared the hypervolume (%) of the reaction conditions obtained by each algorithm to that of the best conditions in the original benchmark dataset, using the latter as a reference for the true optimal solutions.

Aligning with the standard number of reaction vials in solid-dispensing HTE workflows, we benchmarked our optimisation approaches using batch sizes of 24, 48, and 96 for 5 iterations, using Sobol sampling for initial batch selection in the first iteration. We compared the three acquisition functions (q-NEHVI, q-NParEgo, and TS-HVI) against a Sobol baseline. All acquisition functions showed comparable performance, outperforming the Sobol baseline on all datasets and demonstrating scalability to large batch sizes of 96. As convergence to optimal solutions was generally achieved on most benchmark datasets within the lowest batch setting of batch size 24 (120 total experiments), we focused our investigation on these results (Fig. 3a). Although the best acquisition function varied across each dataset, q-NEHVI consistently performed well on all benchmarks (see Supplementary Information section 2 for statistical tests), and was thus selected for further analysis. We also compared our approach to other reaction optimisation software [14] (see Supplementary Information section 2), demonstrating improved scalability to large batch sizes and search spaces. Our benchmark datasets included various reaction condition featurisation methods including Density Functional Theory (DFT), with high-dimensional reaction representations up to 530 dimensions. While we have not explicitly tested upper computational limits of our approaches, we demonstrated robust scalability across dataset sizes, feature dimensions, and batch sizes examined in this study. All benchmarks results can be found in the Supplementary Information.

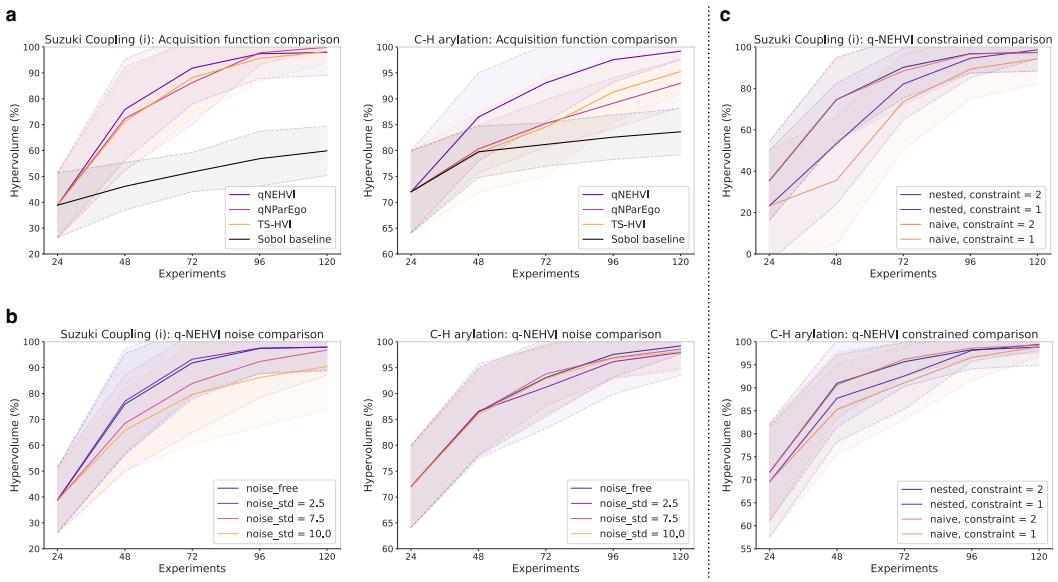

Figure 3: ML optimisation performance on the C-H arylation and Suzuki Coupling (i) virtual benchmark datasets. All displayed benchmark results used batch size of 24 for 5 iterations using the hypervolume (%) metric to assess reaction conditions identified by the optimisation algorithms compared to the best conditions in the ground truth dataset (see Supplementary Information section 2 for benchmarks on all datasets and batch sizes 48/96). All optimisations were repeated across 20 different random seeds, plotting the mean hypervolume (%) and $\pm$ 1 standard deviation. **a**, Comparison of different scalable multi-objective acquisition functions against a quasi-random Sobol baseline. **b**, Optimisation performance of the q-NEHVI acquisition function with varying magnitudes of Gaussian noise added to model training data. **c**, Optimisation performance of the q-NEHVI acquisition function with the nested and naive constrained strategies to accommodate constraints of 1 or 2 unique temperatures per batch.

2.4 Investigating noise robustness

Chemical reactions are stochastic in practice, resulting in yield variations even when reactions are repeated under identical conditions. Although HTE automation reduces this variability, some experimental noise is still expected. To assess the robustness of our optimisation workflow to such real-world variability, we simulated different noise levels by perturbing our model's input data with Gaussian noise of varying standard deviations (see Methods), comparing performance of the q-NEHVI acquisition function under these conditions. Our results showed that while there is some reduction in performance due to input noise, our optimisation approach remains robust even at high noise levels, such as a noise standard deviation of 10 (Fig. 3b). In practice, we expect HTE automation equipment to exhibit lower levels of noise, and our approach demonstrates reliable performance under these conditions. The ability of our workflow to handle appreciable levels of noise highlights its potential for real-world applications, where experimental variability is unavoidable. All noisy benchmark results can be found in Supplementary Information section 3.

2.5 Equipment constraints on experimental batches

Our HTE platform utilises four heating wells for temperature control, restricting a single batch of experiments to a maximum of 4 unique reaction temperatures. To maintain practicality and minimise plate splitting across different heating wells, we restricted optimisation campaigns to 2 unique temperatures per batch. Such temperature constraints are also commonly observed in non-automated laboratory settings, for example, the finite number of heating plates in a fume hood. To implement this constraint into batch selection with our acquisition functions, we extended the naive and nested approaches from Vellanki et al. [38] to accommodate multiple unique allowed temperatures per batch and multi-objective acquisition functions. The naive approach uses standard Bayesian optimisation to select batch experiments until the number of unique temperatures meets the constraint (e.g., 2). The remaining experiments are then restricted to being selected from only those 2 unique temperatures. The nested approach first algorithmically selects the most promising temperatures (see Methods), then identifies optimal reaction conditions within constraints of the selected temperatures. Both approaches can be extended to an arbitrary number of constrained temperatures per batch.

We evaluated these approaches in extreme cases, limiting batches to one or two unique temperatures (see Methods). Despite highly restrictive batch constraints, both approaches were able to identify high-performing experiments, yielding above 90 hypervolume (%) across all benchmarks (Fig. 3c). We observed similar performances between the two approaches for batch constraints of 2 unique temperatures. However, when constrained to 1 unique temperature, the naive approach consistently underperformed compared to the nested approach (see Supplementary Information section 4 for statistical tests). Hence, for the experimental deployment of our workflow, we proceeded with the nested approach for handling batch constraints, combined with the q-NEHVI acquisition function. We note that while these two strategies were observed to perform the best overall across the benchmark datasets evaluated in this study, the optimal algorithm may depend on the specific dataset considered.

2.6 Application to Nickel-catalysed Suzuki reactions

Inspired by successes of non-platinum group metal (PGM) catalysis in the pharmaceutical industry [3] and our lab's application of nickel catalysis to an active pharmaceutical ingredient (API), we challenged our optimisation workflow with a Ni-catalysed Suzuki-Miyaura coupling. Suzuki couplings are fundamental to pharmaceutical development, ranking among the top five reactions performed by medicinal chemists and the second largest reaction class at Roche and AbbVie [29, 39]. Although Pd-catalysed Suzuki reactions dominate the industrial landscape, palladium's increasing scarcity, high cost, and substantial environmental impact (3,880 kg CO_2 equivalents per kg of palladium) present challenges for its continued use [4]. In contrast, nickel offers a more sustainable and cost-effective alternative, with only 6.5 kg CO_2 equivalents per kg and a 3,000-fold lower cost than palladium on a molar basis. These benefits become even more pronounced in process chemistry, where reactions are conducted on multi-(hundred) kilogram scales [40]. However, Ni-catalysed Suzuki reactions are more challenging, prone to substrate inhibition, and generate more byproducts compared to their palladium counterparts. Consequently, the development of Ni-catalysed Suzuki reactions is important for the pharmaceutical industry to address the challenges associated with palladium catalysis.

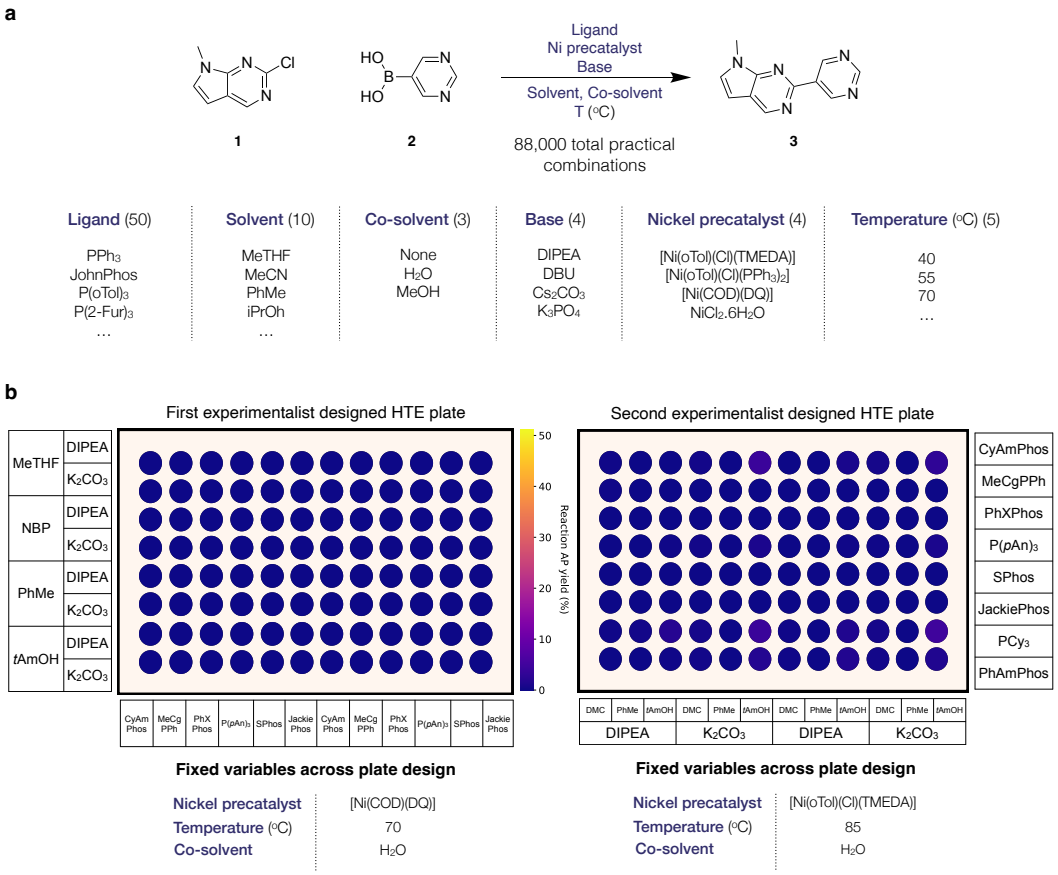

Figure 4: Applying the ML optimisation workflow to a nickel-catalysed Suzuki reaction. **a**, The Ni-catalysed Suzuki reaction chosen for experimental application and the reaction parameters comprising the reaction condition search space defined by experimentalist knowledge. The multiplicative combination of these parameters initially yielded $120,000$ possible reaction configurations. After removing configurations where reaction temperatures exceed the solvent boiling point, the final search space contained $88,000$ possible reactions. The full search space is specified in Supplementary Information section 5. **b**, Initial optimisation results for the Ni-catalysed Suzuki reaction using traditional expert-driven HTE plate design. Two 96-well HTE plates were executed, each containing 48 unique reactions run in duplicate to assess reproducibility. The schematic shows the grid-like structure plate design and fixed variables, specifying the conditions evaluated in the two HTE plates.

We selected a challenging nickel-catalysed heterocyclic Suzuki reaction between substrates **1** and **2** as an experimental case study (Fig. 4a). Based on expert knowledge, we defined a search space comprising 50 monophosphine ligands, 4 nickel precatalysts, 4 bases, 10 solvents, 3 co-solvents, and 5 temperatures. After removing conditions where the reaction temperature exceeded the solvent boiling point, we obtained a final reaction condition search space of $88,000$ combinations. Given the large number of categorical variables, one-hot-binary encoding (OHE) featurisation was impractical. Instead, we employed DFT descriptors to represent ligands and solvents (see Methods for full reaction condition parameterisation). All experimental details are fully specified in Supplementary Information section 5.

As a baseline comparison, we first attempted to optimise this Suzuki reaction using purely expert-designed HTE screening plates with the grid template structure shown in Figure 1a. Two expert-designed 96-well HTE plates were conducted (Fig. 4b), each containing 48 unique reactions run in duplicate. These plates yielded only trace amounts of product, with the best reaction achieving ~ 5 (area percent) AP yield based on Liquid Chromatography-Mass Spectroscopy (LC-MS) analysis (see Methods). Often, the absence of initial reaction hits could lead to premature conclusions regarding the reaction's viability. Subsequently, without incorporating the previous data, we re-attempted this optimisation from scratch, applying our ML optimisation workflow to this Suzuki reaction

within the defined search space of $88,000$ reaction conditions. We initiated the first iteration of this workflow using Sobol sampling to select the initial 96-well HTE experiments to maximise reaction space coverage, followed by model-guided Bayesian optimisation in subsequent iterations. The first Sobol sampling iteration surpassed the results of the two expert-designed plates, identifying multiple reaction hits with up to 51 AP yield and 86 AP selectivity (Fig. 5a). The second and third iterations of Bayesian optimisation further improved on both AP yield and AP selectivity. By the fourth iteration, we observed no improvement in maximum AP yield and selectivity over the prior three iterations, prompting us to conduct one final iteration before terminating the campaign. We noticed that the experiments selected by the Bayesian optimiser for the fifth iteration were still largely explorative, with many experiments suggesting the use of PhMe solvent, which had been minimally explored until this point. Therefore, we decided to implement a fully exploitative optimisation strategy in the final iteration of the campaign, aiming to maximally exploit all information accumulated in the campaign thus far, at the expense of any further exploration. We neglected model uncertainty and selected promising experiments with the highest predicted AP yield and AP selectivity values using Utopia point scalarisation (see Methods). This exploitative approach in the final iteration identified multiple reactions with improved AP yield (up to 76%) with high AP selectivity (up to 92%) compared to previous iterations. These high-performing reactions, conducted at relatively lower temperatures beneficial for process scale-up, represent promising hits that would typically progress to fine-tuning of quantitative factors like catalyst loading and reagent equivalents in industrial process development. For comparison, we also performed the originally suggested HTE plate for the final iteration, confirming the superior performance of the pure exploitative approach, highlighting the merits of leveraging accumulated information in the final stage of an optimisation campaign (see Supplementary Information section 5).

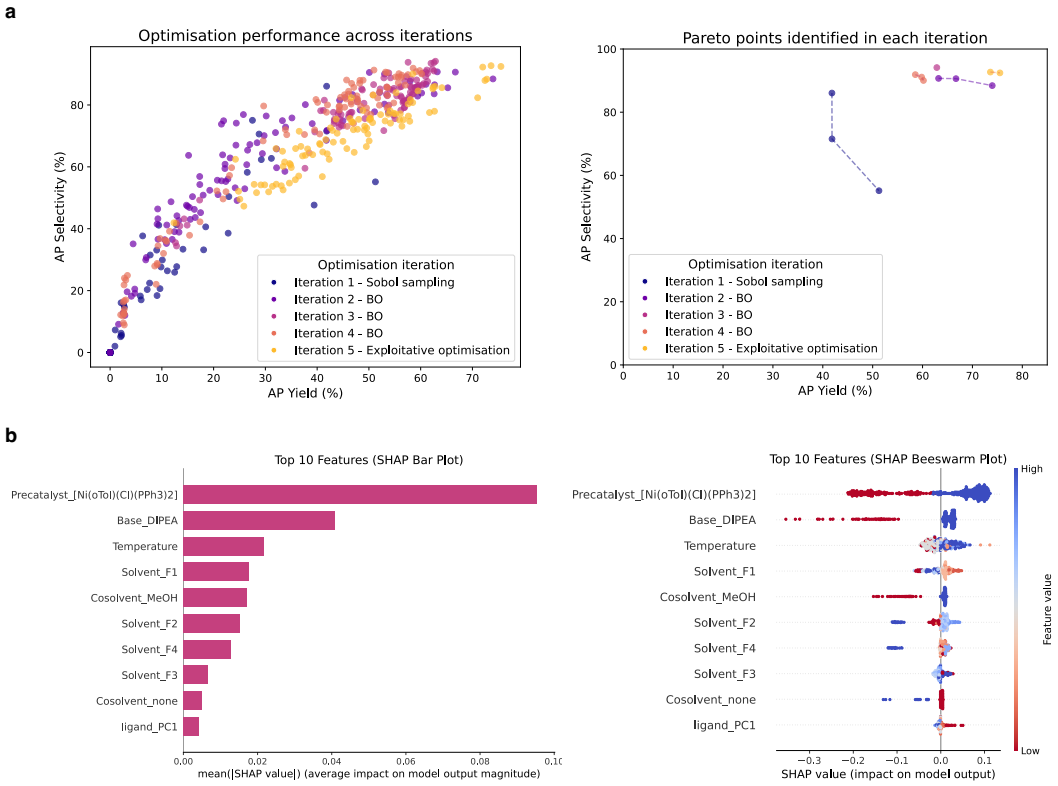

Figure 5: ML optimisation of the nickel-catalysed Suzuki reaction. a, The first scatter plot shows the AP yield and AP selectivity of experiments selected by the ML optimisation workflow at each iteration. The second plot shows the Pareto points, representing the best trade-offs between AP yield and selectivity, identified at each iteration. The Pareto points illustrate the optimal yield-selectivity combinations achieved. **b**, Using all collected experimental data from the campaign, we trained an ML model to perform SHAP [41] feature importance analysis. The SHAP values quantify the magnitude of each feature's contribution to the model's predicted AP yield, considering both positive and negative impacts to identify the most influential factors (see Methods for more details).

From experience, we expected initially that the choice of ligand would be the dominant factor influencing product formation. However, we observed throughout the campaign that this Suzuki reaction was highly sensitive to the choice of nickel precatalyst, with the [Ni(oTol)(Cl)(PPh_3)$_2$] precatalyst present in most of the high-performing reaction conditions. In contrast, different ligands seemed to have comparatively minimal impact on reaction outcomes. A Shapley value analysis of the collected data using SHAP (Shapley additive explanations) [41] corroborated these observations (see Methods), identifying the presence or absence of the [Ni(oTol)(Cl)(PPh_3)$_2$] precatalyst as the most important feature for AP yield, represented by the binary value of the one-hot encoded feature (Fig. 5b). Consistent with our observations, features representing bases, solvents, co-solvents, and temperature were also attributed higher importance than those of the ligands. In fact, when we assessed the most promising conditions on 500 mg scale, we observed essentially identical results with and without added ligand, indicating that [Ni(oTol)(Cl)(PPh_3)$_2$] catalyses the reaction. This departure from expected chemical reactivity explains the minimal product formed in the two expert-designed HTE plates, which used the less active [NiCl(oTol)(TMEDA)] and [Ni(COD)(DQ)] nickel precatalysts. This underscores the limitations of traditional grid-like HTE plate designs, which are restricted in the diversity of reaction parameters explored in each plate and risk overlooking promising regions. In large multidimensional reaction spaces with multiple categorical parameters, as in this case study, this limitation is further pronounced. In such complex landscapes, the presence of reactivity cliffs renders the reaction space challenging for experimentalists to navigate and ensure comprehensive exploration [42]. Consequently, this could lead to erroneous premature conclusions about reaction viability. Our optimisation workflow provides chemists with a valuable tool to efficiently navigate these challenging and intricate spaces, potentially uncovering unexpected reactivity patterns missed by traditional approaches.

3 Discussion

Our study demonstrates the efficacy of our optimisation framework in an automated, high-throughput setting. Initial HTE plate selection with quasi-random sampling ensures comprehensive coverage of large reaction spaces to identify promising local optima and reactivity crevices. Subsequent model-driven Bayesian optimisation iteratively refines reaction objectives, while purely exploitative optimisation in the final round leverages all accumulated knowledge to maximise final reaction outcomes. We successfully validated our ML-driven optimisation workflow on a challenging Ni-catalysed Suzuki reaction in a 96-well HTE campaign, demonstrating advantages over traditional, purely experimentalist-guided methods. By integrating machine learning strategies with highly parallel automation, our workflow demonstrates real-world applicability for accelerating reaction optimisation and is currently being implemented in several process chemistry projects at Roche. New values of parameters such as ligands and solvents could also be added to the search space as optimisation proceeds [14], and the workflow can be applied to other chemical reactions of interest, including those outside of HTE settings.

In this work, we only conducted the HTE optimisation campaign using a static search space across all iterations comprising $88,000$ possible reaction combinations, defined by the experimentalist prior to initialisation. Future applications could use dynamic, experimentalist-guided modifications to the search space during iterations, combining observed data with expert knowledge. Additionally, integrating more systematic experimentalist input could guide the desired balance between exploration and exploitation at various stages of the campaign, tailoring the process to specific campaign goals. The suggested improvements could be particularly effective when combined with explainable AI and visualisation methods, enhancing interpretability and decision-making. We envision that such collaborative approaches between ML algorithms and human experts could substantially enhance the adoption and trust in optimisation algorithms within the chemical community.

4 Methods

4.1 Emulated benchmark datasets

We used ML regressors trained on experimental data to generate expanded reaction condition spaces by predicting reaction outcomes for a larger set of experiments that, while not present in the original training data, fall primarily within the domain of the training set. We refer to these ML regressors as emulators. We used emulators from the Olympus [35] benchmarking framework ($suzuki\_i$ to

$suzuki\_iv$) to generate the Suzuki Coupling (i to iv) virtual benchmark datasets. Similarly, we also trained a Multi-layer Perceptron (MLP) using TensorFlow [43] on experimental data from Torres et al. [14] (EDBO+) to generate the C-H arylation virtual benchmark dataset. A detailed description of all (emulated) virtual benchmark datasets are included in Supplementary Information section 1 and the data accompanying this paper.

4.2 Computational implementation

Bayesian optimisation is an iterative approach for identifying the local optima of 'black-box' functions, where the functional form is unknown or cannot be expressed analytically. It relies on constructing a probabilistic surrogate model, usually a Gaussian Process (GP) [31], to approximate the unknown function. An acquisition function, using the GP model's predicted mean and uncertainty, selects the next experiments to evaluate based on previously observed data. The acquisition function balances exploration and exploitation, simultaneously enabling efficient exploration of uncertain regions in the parameter space whilst exploiting promising regions, allowing the algorithm to converge towards local optima. Frazier [18] provides additional background on a more comprehensive mathematical treatment of Bayesian optimisation. In this work, we used the multi-objective acquisition functions: q-NParEgo [34], Thompson sampling with hypervolume improvement (TS-HVI) [32], and q-Noisy Expected Hypervolume Improvement (q-NEHVI) [34]. q-NParEgo simplifies the multi-objective problem by combining the multiple objectives into a single optimisation objective through various scalarisations. TS-HVI provides a scalable approximation of EHVI [32]. q-NEHVI is a more efficient and noise-robust method of evaluating EHVI, with time and memory space complexity scaling polynomially instead of exponentially [34]. We used PyTorch [44], GPyTorch [45], and BoTorch [46] to build, initialise, test, benchmark, and apply all of our GP surrogate models and acquisition functions.

Initial training data for the GP in all cases was generated using Sobol sampling with PyTorch [44], employing low-discrepancy Sobol sequences to obtain quasi-random points uniformly covering the multi-dimensional reaction space, providing superior distribution and space filling properties than standard random sampling [17]. Similar methods like centroidal Voronoi tessellation (CVT) and Latin hypercube sampling (LHS), have shown superior optimisation performance compared to random initialisation [14]. Training inputs and targets for all GPs are normalised and standardised, respectively, according to the specifications in BoTorch [46]. As we aimed to focus the benchmarks on comparing differences between acquisition functions, we used a GP with general purpose kernel hyperparameters adapted from EDBO+ [14] for all benchmarks. All acquisition functions were implemented using BoTorch [46]. We implemented KeOps [47] and fast variance estimates from GPyTorch [45] to enable memory-efficient computations. We used PyTorch Lightning [48] to set random seeds, running 20 repeats from seed 1 to 20 for each benchmark result. All computations were run on a workstation with an AMD Ryzen 9 5900X 12-Core CPU and a RTX 3090 (24GB) GPU.

Comparisons against EDBO+ [14] were implemented according to instructions in the code accompanying its publication, using the same scripts but replacing the Pd-catalysed C-H arylation data set in the EDBO+ directory with the expanded C-H arylation virtual benchmark dataset generated in this work. To ensure fair comparison, EDBO+ was tested with both CPU and GPU (see Supplementary Information section 2). For noisy benchmarks in section 2.4, we perturbed objective values only for yield and catalyst turnover number for all datasets, excluding noiseless reaction costs in the C-H arylation virtual dataset. The noisy values were clamped at 0% and 100% for yield, and for turnover in the Suzuki Coupling virtual datasets, at 0 and the max observed values. The nested acquisition function described in section 2.5 first selected promising temperatures by ranking the temperatures with the highest average q-NEHVI acquisition function value, followed by standard acquisition function evaluation on the search space restricted to the obtained top performing temperatures. For the benchmarks constrained to 1 and 2 unique temperatures per batch, model initialisation data was constructed by randomly selecting 1, or 2 unique temperatures according to constraints. Then, Sobol samples were drawn from experimental conditions restricted to those temperatures to initialise optimisation process. All Wilcoxon statistical significance tests included in the Supplementary Information were implemented using SciPy [49].

Exploitative greedy reaction optimisation strategies in a single-objective case select experiments based solely on only the highest predicted mean yield % [23], neglecting any GP model uncertainty

and hence exploration. For our implementation of multi-objective exploitative optimisation, we used Utopia point scalarisation to combine AP yield and AP selectivity into a single value. Utopia point scalarisation measures the euclidean distance of all AP yield and AP selectivity data values from a hypothetical ideal Utopia point, set in this case to 110 AP yield and AP selectivity, and has shown effectiveness in prior Bayesian optimisation applications [26]. This scalarisation allowed us to rank and select the most promising reactions considering both objectives simultaneously. Analogous to the single-objective case, we selected experiments with the predicted mean AP yield and AP selectivity values closest to the Utopia point. Consistent with an exploitative greedy approach, we selected the two temperatures for this experimental batch using the experiments with the closest predicted Utopia point distances.

4.3 Experimental application

The reaction condition space comprised 50 monophosphine ligands, 4 nickel precatalysts, 4 bases, 10 solvents, 3 co-solvents, and 5 temperatures to give a total of $120,000$ combinatorial reaction configurations (see Supplementary Information section 5 for full list). After removing conditions where the reaction temperature exceeded the solvent boiling point, we obtained a reaction condition search space of $88,000$ configurations. Given the large number of categorical variables, which would require 71 one-hot-encoded (OHE) features to describe, we parameterised monophosphine ligands and solvents using quantum mechanical descriptors, which have shown good performance in prior optimisation studies [14, 27, 50–52], to provide a more informative representation. We used monophosphine ligand descriptors from Kraken [53], applying principal component analysis (PCA) to narrow down the 190 ligand descriptors from Kraken to 37 principal components based on a 99% explained variance threshold, labeled $ligand\_PC1$ to $ligand\_PC37$. Solvent descriptors were obtained from Moity et al. [54] with parameterisation using COSMOtherm and represented with 4 DFT-based descriptors labeled $Solvent\_F1$ to $Solvent\_F4$ The rest of the categorical variables (co-solvents, bases, and precatalysts) were featurised using OHE, with temperature remaining numerical. The resulting encoded reaction feature space is provided in the data section.

Our HTE platform from Unchained Labs uses four distinct heating wells for reaction execution. We used two heating plates for each batch of HTE experiments in our ML optimisation workflow, constrained to two unique temperatures per batch. The experiments used to initialise the ML experimental workflow were selected using quasi-random Sobol sampling, restricting initial reaction temperatures to 70 and 100 degrees Celsius (see Supplementary Information section 5 for experimental procedures). All HTE reactions were evaluated using area percent (AP) yield. These yields, derived from Liquid Chromatography (LC), are uncorrected from differences in LC response factors between the Suzuki coupling product and the limited starting material. While approximate, AP yields provide a useful measure for elucidating reactivity trends and comparing performance of different reaction conditions [40].

4.4 Experimental data analysis and visualisation

To support empirical observations from the collected HTE experimental data and further elucidate underlying chemical relationships governing the nickel-catalysed Suzuki reaction, we employed several analysis and visualisation methods. First, we generated box plots comparing the average experimental AP yield value of all ligands, solvents, bases, precatalysts, co-solvents, and temperature to the overall average AP yield of all reactions. These plots are included in Supplementary Information section 5. To gain deeper insights into feature importance, we trained a Random Forest surrogate model on the collected HTE data to approximate the AP yield function. We then applied SHAP [41] analysis to obtain feature importances for features of each experimental parameter. SHAP uses cooperative game theory concepts to quantify the magnitude of each feature's contribution to the model's prediction, assigning each feature a SHAP value that represents its impact on model output. Beeswarm and bar plots of the SHAP results were generated using the default settings of the SHAP [41] package. For a more comprehensive explanation of the SHAP methodology and its implementation, we refer the reader to the SHAP package documentation, which provides more detailed information on the calculation and interpretation of SHAP values.

Data availability

All virtual and experimental data and reaction condition spaces generated in this study are included in the manuscript, Supplementary Information, and on the accompanying public GitHub repository.

Code availability

The Python code used in this study is made available in a public GitHub repository under the MIT open source license: github.com/schwallergroup/minerva

Acknowledgements

We thank T. Vögtlin, M. Müller, T. Chi, and J. Krizic from the High-Output Reaction Screening System (HORSS) team at F. Hoffmann-La Roche Ltd. for experimental and analytical support. J.W.S., R.P.B., K.P., and R.B. thank Roche and its Technology Innovation and Science initiative for generous financial support. S.L.C. acknowledges support from the Helmholtz Association of German Research Centres. P.S. acknowledges support from NCCR Catalysis (grant no. 180544), a National Centre of Competence in Research funded by the Swiss National Science Foundation. We are grateful to Z. Jončev, J.J. Dotson, R.C. Walroth, and K.A. Mack for useful discussions.

Author contributions

J.W.S. contributed to the conceptualisation, methodology, code development, analysis, visualisation, and writing of the manuscript. S.L.C. contributed to methodology, code development, and writing of the manuscript. R.P.B. and K.P. contributed to conceptualisation, supervision, and writing of the manuscript. R.B. contributed to conceptualisation, supervision, experimental results, and writing of the manuscript. P.S. contributed to conceptualisation, methodology, supervision, and writing of the manuscript.

Competing interests

J.W.S., R.P.B., K.P., and R.B. declare potential financial and non-financial conflict of interest as full employees of F. Hoffmann-La Roche Ltd. The other authors declare no competing interests.

References

[1] Michael F. Lipton and Anthony G. M. Barrett. Introduction: Process chemistry. *Chemical Reviews*, 106:2581–2582, 7 2006. ISSN 0009-2665. doi: 10.1021/cr068400d.

[2] Tony Y. Zhang. Process chemistry: The science, business, logic, and logistics. *Chemical Reviews*, 106(7):2583–2595, March 2006. ISSN 1520-6890. doi: 10.1021/cr040677v. URL http://dx.doi.org/10.1021/cr040677v.

[3] Hui Zhao, Anne K. Ravn, Michael C. Haibach, Keary M. Engle, and Carin C. C. Johansson Seechurn. Diversification of pharmaceutical manufacturing processes: Taking the plunge into the non-pgm catalyst pool. *ACS Catalysis*, page 9708–9733, June 2024. ISSN 2155-5435. doi: 10.1021/acscatal.4c01809. URL http://dx.doi.org/10.1021/acscatal.4c01809.

[4] Xuelei Guo, Hester Dang, Steven R. Wisniewski, and Eric M. Simmons. Nickel-catalyzed suzuki–miyaura cross-coupling facilitated by a weak amine base with water as a cosolvent. *Organometallics*, 41(11):1269–1274, May 2022. ISSN 1520-6041. doi: 10.1021/acs.organomet.2c00197. URL http://dx.doi.org/10.1021/acs.organomet.2c00197.

[5] Coby J. Clarke, Wei-Chien Tu, Oliver Levers, Andreas Bröhl, and Jason P. Hallett. Green and sustainable solvents in chemical processes. *Chemical Reviews*, 118(2):747–800, January 2018. ISSN 1520-6890. doi: 10.1021/acs.chemrev.7b00571. URL http://dx.doi.org/10.1021/acs.chemrev.7b00571.

[6] Alexander Buitrago Santanilla, Erik L. Regalado, Tony Pereira, Michael Shevlin, Kevin Bateman, Louis-Charles Campeau, Jonathan Schneeweis, Simon Berritt, Zhi-Cai Shi, Philippe Nantermet, Yong Liu, Roy Helmy, Christopher J. Welch, Petr Vachal, Ian W. Davies, Tim Cernak, and Spencer D. Dreher. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science*, 347(6217):49–53, January 2015. ISSN 1095-9203. doi: 10.1126/science.1259203. URL http://dx.doi.org/10.1126/science.1259203.

[7] Nathan Gesmundo, Kevin Dykstra, James L. Douthwaite, Yu-Ting Kao, Ruheng Zhao, Babak Mahjour, Ron Ferguson, Spencer Dreher, Bérengère Sauvagnat, Josep Saurí, and Tim Cernak. Miniaturization of popular reactions from the medicinal chemists' toolbox for ultrahigh-throughput experimentation. *Nature Synthesis*, 2(11):1082–1091, June 2023. ISSN 2731-0582. doi: 10.1038/s44160-023-00351-1. URL http://dx.doi.org/10.1038/s44160-023-00351-1.

[8] Babak Mahjour, Rui Zhang, Yuning Shen, Andrew McGrath, Ruheng Zhao, Osama G. Mohamed, Yingfu Lin, Zirong Zhang, James L. Douthwaite, Ashootosh Tripathi, and Tim Cernak. Rapid planning and analysis of high-throughput experiment arrays for reaction discovery. *Nature Communications*, 14(1), July 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-39531-0. URL http://dx.doi.org/10.1038/s41467-023-39531-0.

[9] Loïc M. Roch, Florian Häse, Christoph Kreisbeck, Teresa Tamayo-Mendoza, Lars P. E. Yunker, Jason E. Hein, and Alán Aspuru-Guzik. Chemos: Orchestrating autonomous experimentation. *Science Robotics*, 3(19), June 2018. ISSN 2470-9476. doi: 10.1126/scirobotics.aat5559. URL http://dx.doi.org/10.1126/scirobotics.aat5559.

[10] Connor W. Coley, Dale A. Thomas, Justin A. M. Lummiss, Jonathan N. Jaworski, Christopher P. Breen, Victor Schultz, Travis Hart, Joshua S. Fishman, Luke Rogers, Hanyu Gao, Robert W. Hicklin, Pieter P. Plehiers, Joshua Byington, John S. Piotti, William H. Green, A. John Hart, Timothy F. Jamison, and Klavs F. Jensen. A robotic platform for flow synthesis of organic compounds informed by ai planning. *Science*, 365(6453), August 2019. ISSN 1095-9203. doi: 10.1126/science.aax1566. URL http://dx.doi.org/10.1126/science.aax1566.

[11] Sebastian Steiner, Jakob Wolf, Stefan Glatzel, Anna Andreou, Jarosław M. Granda, Graham Keenan, Trevor Hinkley, Gerardo Aragon-Camarasa, Philip J. Kitson, Davide Angelone, and Leroy Cronin. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science*, 363(6423), January 2019. ISSN 1095-9203. doi: 10.1126/science.aav2211. URL http://dx.doi.org/10.1126/science.aav2211.

[12] Connor J. Taylor, Alexander Pomberger, Kobi C. Felton, Rachel Grainger, Magda Barecka, Thomas W. Chamberlain, Richard A. Bourne, Christopher N. Johnson, and Alexei A. Lapkin. A brief introduction to chemical reaction optimization. *Chemical Reviews*, 123:3089–3126, 3 2023. ISSN 0009-2665. doi: 10.1021/acs.chemrev.2c00798.

[13] Steven M. Mennen, Carolina Alhambra, C. Liana Allen, Mario Barberis, Simon Berritt, Thomas A. Brandt, Andrew D. Campbell, Jesús Castañón, Alan H. Cherney, Melodie Christensen, David B. Damon, J. Eugenio de Diego, Susana García-Cerrada, Pablo García-Losada, Rubén Haro, Jacob Janey, David C. Leitch, Ling Li, Fangfang Liu, Paul C. Lobben, David W. C. MacMillan, Javier Magano, Emma McInturff, Sebastien Monfette, Ronald J. Post, Danielle Schultz, Barbara J. Sitter, Jason M. Stevens, Iulia I. Strambeanu, Jack Twilton, Ke Wang, and Matthew A. Zajac. The evolution of high-throughput experimentation in pharmaceutical development and perspectives on the future. *Organic Process Research & Development*, 23: 1213–1242, 6 2019. ISSN 1083-6160. doi: 10.1021/acs.oprd.9b00140.

[14] Jose Antonio Garrido Torres, Sii Hong Lau, Pranay Anchuri, Jason M. Stevens, Jose E. Tabora, Jun Li, Alina Borovika, Ryan P. Adams, and Abigail G. Doyle. A multi-objective active learning platform and web app for reaction optimization. *Journal of the American Chemical Society*, 144:19999–20007, 11 2022. ISSN 0002-7863. doi: 10.1021/jacs.2c08592.

[15] Michael Shevlin. Practical high-throughput experimentation for chemists. *ACS Medicinal Chemistry Letters*, 8(6):601–607, May 2017. ISSN 1948-5875. doi: 10.1021/acsmedchemlett.7b00165. URL `http://dx.doi.org/10.1021/acsmedchemlett.7b00165`.

[16] Georg Wuitschik, Vera Jost, Torsten Schindler, and Michal Jakubik. Hte os: A high-throughput experimentation workflow built from the ground up. *Organic Process Research & Development*, 28(7):2875–2884, 2024. doi: 10.1021/acs.oprd.4c00160. URL `https://doi.org/10.1021/acs.oprd.4c00160`.

[17] Sebastian Burhenne, Dirk Jacob, and Gregor Henze. Sampling based on sobol' sequences for monte carlo techniques applied to building simulations. *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association*, pages 1816–1823, 01 2011.

[18] Peter I. Frazier. A tutorial on bayesian optimization, 2018. URL `https://arxiv.org/abs/1807.02811`.

[19] Xilu Wang, Yaochu Jin, Sebastian Schmitt, and Markus Olhofer. Recent advances in bayesian optimization, 2022. URL `https://arxiv.org/abs/2206.03301`.

[20] Siu Lun Chau, Jean-Francois Ton, Javier Gonz'alez, Yee Teh, and Dino Sejdinovic. Bayesimp: Uncertainty quantification for causal data fusion. *Advances in Neural Information Processing Systems*, 34:3466–3477, 2021.

[21] Masaki Adachi, Brady Planden, David A Howey, Krikamol Maundet, Michael A Osborne, and Siu Lun Chau. Looping in the human: Collaborative and explainable bayesian optimization. In *Artificial intelligence and statistics*, 2024.

[22] Florian Häse, Loïc M. Roch, Christoph Kreisbeck, and Alán Aspuru-Guzik. Phoenics: A bayesian optimizer for chemistry. *ACS Central Science*, 4(9):1134–1145, August 2018. ISSN 2374-7951. doi: 10.1021/acscentsci.8b00307. URL `http://dx.doi.org/10.1021/acscentsci.8b00307`.

[23] Benjamin J. Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I. Martinez Alvarado, Jacob M. Janey, Ryan P. Adams, and Abigail G. Doyle. Bayesian reaction optimization as a tool for chemical synthesis. *Nature*, 590:89–96, 2 2021. ISSN 0028-0836. doi: 10.1038/s41586-021-03213-y.

[24] Jeff Guo, Bojana Ranković, and Philippe Schwaller. Bayesian optimization for chemical reactions. *CHIMIA*, 77(1/2):31, February 2023. ISSN 0009-4293. doi: 10.2533/chimia.2023.31. URL `http://dx.doi.org/10.2533/chimia.2023.31`.

[25] Elena Braconi and Edouard Godineau. 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 Chemistry & Engineering*, 11(28):10545–10554, July 2023. ISSN 2168-0485. doi: 10.1021/acssuschemeng.3c02455. URL `http://dx.doi.org/10.1021/acssuschemeng.3c02455`.

[26] Derek M. Dalton, Richard C. Walroth, Caroline Rouget-Virbel, Kyle A. Mack, and F. Dean Toste. Utopia point bayesian optimization finds condition-dependent selectivity for n-methyl pyrazole condensation. *Journal of the American Chemical Society*, 146(23):15779–15786, May 2024. ISSN 1520-5126. doi: 10.1021/jacs.4c01616. URL `http://dx.doi.org/10.1021/jacs.4c01616`.

[27] Melodie Christensen, Lars P. E. Yunker, Folarin Adedeji, Florian Häse, Loïc M. Roch, Tobias Gensch, Gabriel dos Passos Gomes, Tara Zepel, Matthew S. Sigman, Alán Aspuru-Guzik, and Jason E. Hein. Data-science driven autonomous process optimization. *Communications Chemistry*, 4:112, 8 2021. ISSN 2399-3669. doi: 10.1038/s42004-021-00550-x.

[28] Kenneth Atz, David F. Nippa, Alex T. Müller, Vera Jost, Andrea Anelli, Michael Reutlinger, Christian Kramer, Rainer E. Martin, Uwe Grether, Gisbert Schneider, and Georg Wuitschik. Geometric deep learning-guided suzuki reaction conditions assessment for applications in medicinal chemistry. *RSC Medicinal Chemistry*, 2024. ISSN 2632-8682. doi: 10.1039/d4md00196f. URL http://dx.doi.org/10.1039/D4MD00196F.

[29] Priyanka Raghavan, Alexander J. Rago, Pritha Verma, Majdi M. Hassan, Gashaw M. Goshu, Amanda W. Dombrowski, Abhishek Pandey, Connor W. Coley, and Ying Wang. Incorporating synthetic accessibility in drug design: Predicting reaction yields of suzuki cross-couplings by leveraging abbvie's 15-year parallel library data set. *Journal of the American Chemical Society*, May 2024. ISSN 1520-5126. doi: 10.1021/jacs.4c00098. URL http://dx.doi.org/10.1021/jacs.4c00098.

[30] David F. Nippa, Alex T. Müller, Kenneth Atz, David B. Konrad, Uwe Grether, Rainer E. Martin, and Gisbert Schneider. Simple user-friendly reaction format, May 2024. URL http://dx.doi.org/10.26434/chemrxiv-2023-nfq7h-v2.

[31] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006. ISBN 026218253X.

[32] Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Multi-objective bayesian optimization over high-dimensional search spaces, 9 2021.

[33] Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization, 6 2020.

[34] Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Parallel bayesian optimization of multiple noisy objectives with expected hypervolume improvement, 5 2021.

[35] Florian Häse, Matteo Aldeghi, Riley J Hickman, Loïc M Roch, Melodie Christensen, Elena Liles, Jason E Hein, and Alán Aspuru-Guzik. Olympus: a benchmarking framework for noisy optimization and experiment planning. *Machine Learning: Science and Technology*, 2:035021, 9 2021. ISSN 2632-2153. doi: 10.1088/2632-2153/abedc8.

[36] Kobi C. Felton, Jan G. Rittig, and Alexei A. Lapkin. Summit: Benchmarking machine learning methods for reaction optimisation. *Chemistry–Methods*, 1:116–122, 2 2021. ISSN 2628-9725. doi: 10.1002/cmtd.202000051.

[37] Andreia P. Guerreiro, Carlos M. Fonseca, and Luís Paquete. The hypervolume indicator: Problems and algorithms, 2020. URL https://arxiv.org/abs/2005.00515.

[38] Pratibha Vellanki, Santu Rana, Sunil Gupta, David Rubin, Alessandra Sutti, Thomas Dorin, Murray Height, Paul Sanders, and Svetha Venkatesh. Process-constrained batch bayesian optimisation. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/1f71e393b3809197ed66df836fe833e5-Paper.pdf.

[39] John D. Hayler, David K. Leahy, and Eric M. Simmons. A pharmaceutical industry perspective on sustainable metal catalysis. *Organometallics*, 38(1):36–46, October 2018. ISSN 1520-6041. doi: 10.1021/acs.organomet.8b00566. URL http://dx.doi.org/10.1021/acs.organomet.8b00566.

[40] Matthew J. Goldfogel, Xuelei Guo, Jeishla L. Meléndez Matos, John A. Gurak, Matthew V. Joannou, William B. Moffat, Eric M. Simmons, and Steven R. Wisniewski. Advancing base-metal catalysis: Development of a screening method for nickel-catalyzed suzuki–miyaura reactions of pharmaceutically relevant heterocycles. *Organic Process Research & Development*, 26(3):785–794, July 2021. ISSN 1520-586X. doi: 10.1021/acs.oprd.1c00210. URL http://dx.doi.org/10.1021/acs.oprd.1c00210.

[41] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages

4765–4774. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[42] Dennis W. Lendrem, B. Clare Lendrem, David Woods, Ruth Rowland-Jones, Matthew Burke, Marion Chatfield, John D. Isaacs, and Martin R. Owen. Lost in space: design of experiments and scientific exploration in a hogarth universe. *Drug Discovery Today*, 20(11):1365–1371, November 2015. ISSN 1359-6446. doi: 10.1016/j.drudis.2015.09.015. URL http://dx.doi.org/10.1016/j.drudis.2015.09.015.

[43] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[44] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019. URL https://arxiv.org/abs/1912.01703.

[45] Jacob R. Gardner, Geoff Pleiss, David Bindel, Kilian Q. Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration, 9 2018.

[46] Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. Botorch: A framework for efficient monte-carlo bayesian optimization, 10 2019.

[47] Benjamin Charlier, Jean Feydy, Joan Alexis Glaunès, François-David Collin, and Ghislain Durif. Kernel operations on the gpu, with autodiff, without memory overflows, 3 2020.

[48] William Falcon, Jirka Borovec, Adrian Wälchli, Nic Eggert, Justus Schock, Jeremy Jordan, Nicki Skafte, Ir1dXD, Vadim Bereznyuk, Ethan Harris, Tullie Murrell, Peter Yu, Sebastian Præsius, Travis Addair, Jacob Zhong, Dmitry Lipin, So Uchida, Shreyas Bapat, Hendrik Schröter, Boris Dayma, Alexey Karnachev, Akshay Kulkarni, Shunta Komatsu, Martin.B, Jean-Baptiste SCHIRATTI, Hadrien Mary, Donal Byrne, Cristobal Eyzaguirre, Cinjon, and Anton Bakhtin. Pytorchlightning/pytorch-lightning: 0.7.6 release, 2020. URL https://zenodo.org/record/3828935.

[49] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[50] Jordan J. Dotson, Lucy van Dijk, Jacob C. Timmerman, Samantha Grosslight, Richard C. Walroth, Francis Gosselin, Kurt Püntener, Kyle A. Mack, and Matthew S. Sigman. Data-driven multi-objective optimization tactics for catalytic asymmetric reactions using bisphosphine ligands. *Journal of the American Chemical Society*, 145(1):110–121, December 2022. ISSN 1520-5126. doi: 10.1021/jacs.2c08513. URL http://dx.doi.org/10.1021/jacs.2c08513.

[51] Derek T. Ahneman, Jesús G. Estrada, Shishi Lin, Spencer D. Dreher, and Abigail G. Doyle. Predicting reaction performance in c–n cross-coupling using machine learning. *Science*, 360:186–190, 4 2018. ISSN 0036-8075. doi: 10.1126/science.aar5169.

[52] Andrzej M. Żurański, Jason Y. Wang, Benjamin J. Shields, and Abigail G. Doyle. Autoqchem: an automated workflow for the generation and storage of dft calculations for organic molecules. *Reaction Chemistry & Engineering*, 7(6):1276–1284, 2022. ISSN 2058-9883. doi: 10.1039/d2re00030j. URL http://dx.doi.org/10.1039/D2RE00030J.

[53] Tobias Gensch, Gabriel dos Passos Gomes, Pascal Friederich, Ellyn Peters, Théophile Gaudin, Robert Pollice, Kjell Jorner, AkshatKumar Nigam, Michael Lindner-D'Addario, Matthew S. Sigman, and Alán Aspuru-Guzik. A comprehensive discovery platform for organophosphorus ligands for catalysis. *Journal of the American Chemical Society*, 144(3):1205–1217, January 2022. ISSN 1520-5126. doi: 10.1021/jacs.1c09718. URL `http://dx.doi.org/10.1021/jacs.1c09718`.

[54] Laurianne Moity, Morgan Durand, Adrien Benazzouz, Christel Pierlot, Valérie Molinier, and Jean-Marie Aubry. Panorama of sustainable solvents using the cosmo-rs approach. *Green Chemistry*, 14(4):1132, 2012. ISSN 1463-9270. doi: 10.1039/c2gc16515e. URL `http://dx.doi.org/10.1039/C2GC16515E`.