Discrete and mixed-variable experimental design with surrogate-based approach

Mengjia Zhu, Austin Mroz, Lingfeng Gui, Kim Jelfs, Alberto Bemporad, Ehecatl Antonio del Río Chanona*, and Ye Seol Lee*

Corresponding Authors: lauren.lee@ucl.ac.uk, a.del-rio-chanona@imperial.ac.uk

MOTIVATIONS

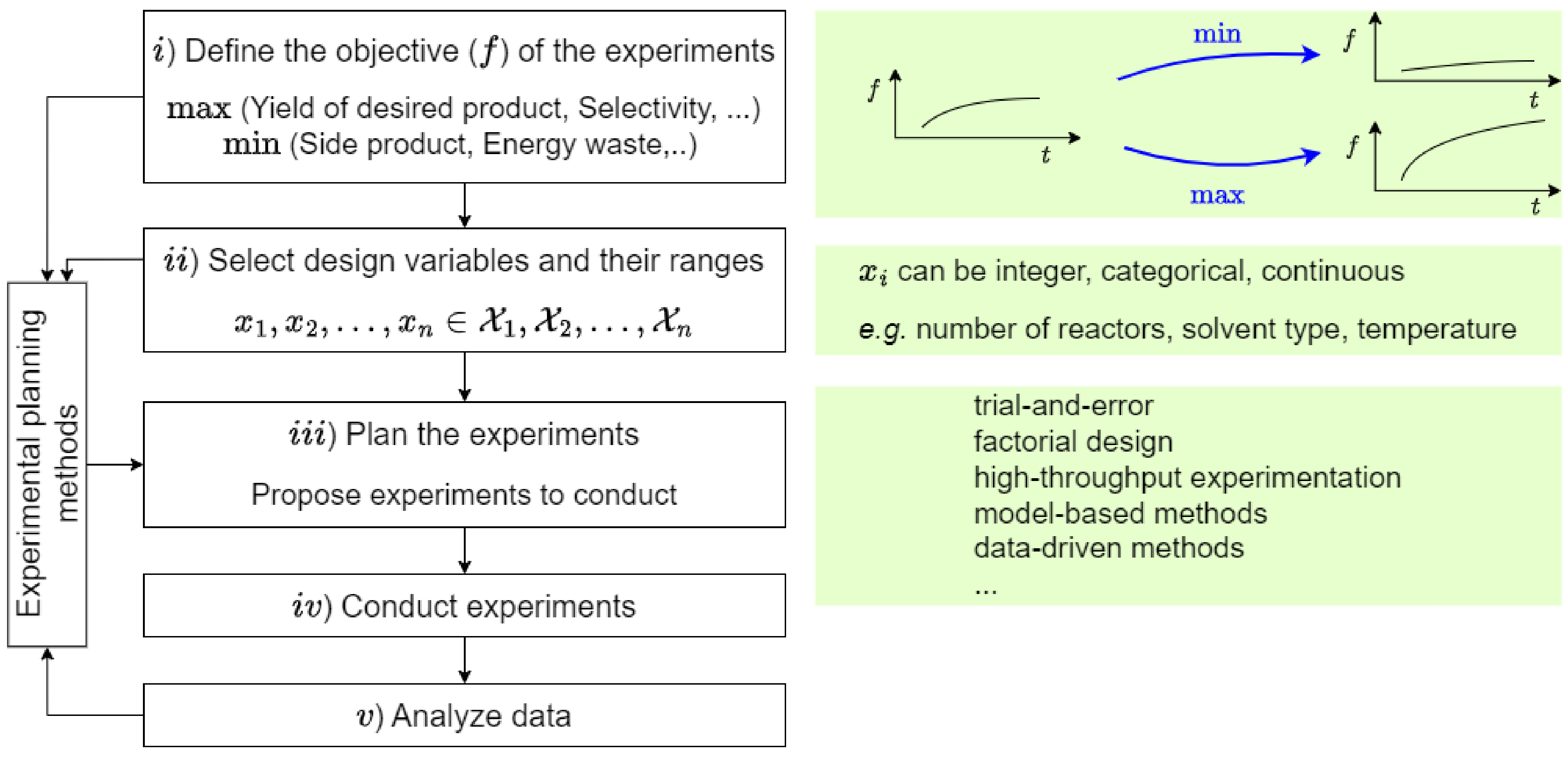
1. Experimental planning in chemistry often involves discrete and mixed variables, with known discrete/mixed-variable constraints
2. These problems can be challenging for conventional Bayesian Optimization (BO) approaches to find feasible samples while maintaining exploration capability

OBJECTIVES

1. Propose alternative surrogate and acquisition models for realistic design space representation while preserving exploration
2. Integrate mixed-integer optimization for feasible sampling
3. Benchmark against state-of-the-art algorithms to demonstrate effectiveness

GENERAL METHODOLOGY

GENERAL STEPS OF EXPERIMENTAL DESIGN



Experimental planning methods

i) Define the objective (f) of the experiments
$\max$ (Yield of desired product, Selectivity, ...)
$\min$ (Side product, Energy waste,..)

ii) Select design variables and their ranges
$x_1, x_2, \ldots, x_n \in \mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$

iii) Plan the experiments
Propose experiments to conduct

iv) Conduct experiments

v) Analyze data

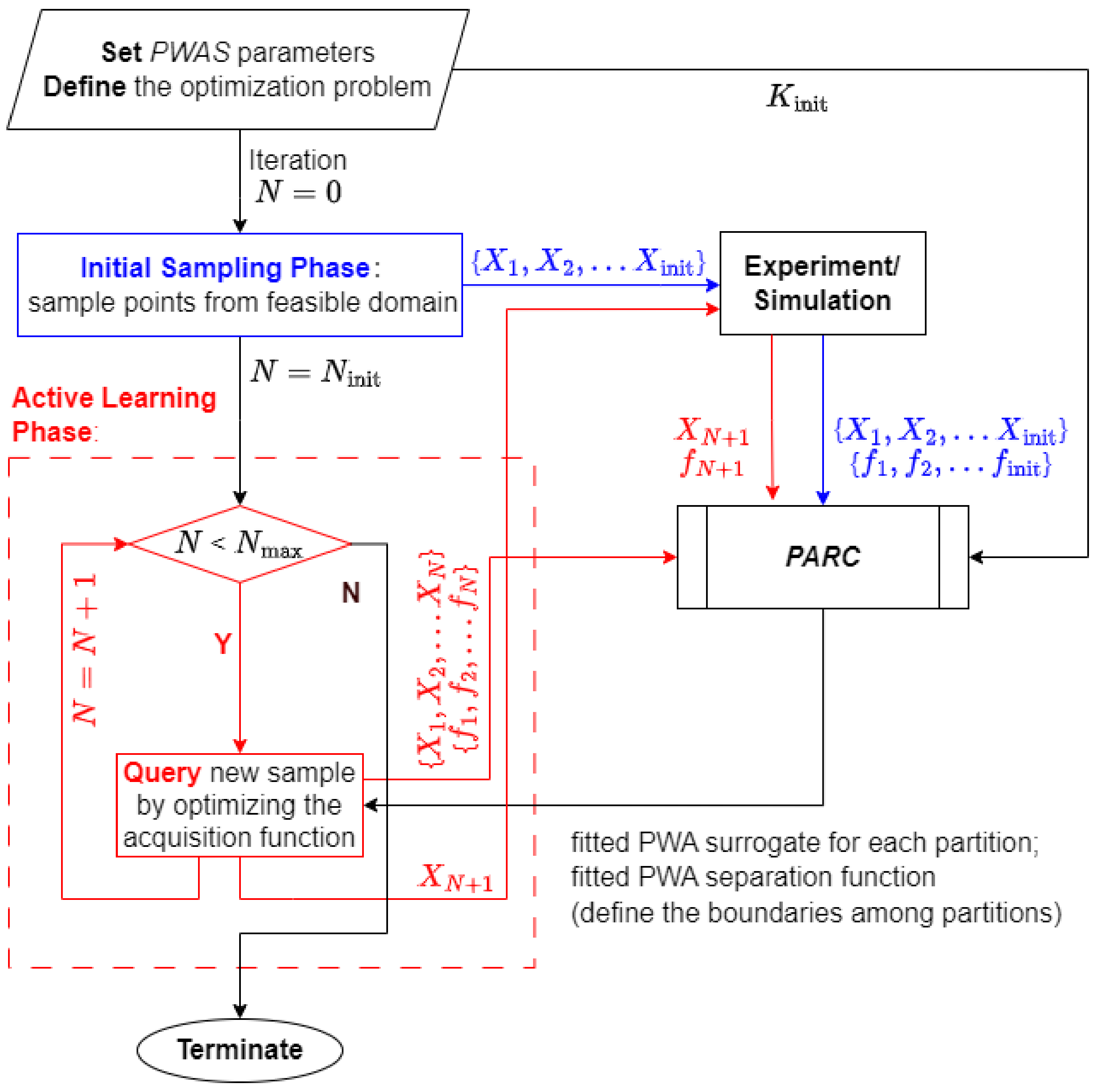
x_i can be integer, categorical, continuous

e.g. number of reactors, solvent type, temperature

trial-and-error
factorial design
high-throughput experimentation
model-based methods
data-driven methods
...

- Performing chemical and physical experiments is often time-consuming and costly
- it is important to plan experiments efficiently to gather pertinent data with a **small** number of required experiments
- **Goal**: develop effective experimental planning strategies

PIECEWISE AFFINE SURROGATE MODEL



Set *PWAS* parameters
Define the optimization problem

Iteration
$N = 0$

Initial Sampling Phase:
sample points from feasible domain

$N = N_{\text{init}}$

Active Learning Phase:

$N < N_{\max}$

Query new sample by optimizing the acquisition function

Experiment/ Simulation

PARC

Terminate

fitted PWA surrogate for each partition;
fitted PWA separation function
(define the boundaries among partitions)

PWAS:
Global optimization with mixed variables using Piecewise Affine Surrogates

PARC:
a block-descent algorithm to fit PWA surrogate

Why piecewise affine (PWA) function as surrogate model:

- Allow discontinuities (categorical variables)
- Have direct MILP reformulation (solved by efficient MILP solvers)

Exploration models: max-box & hamming distance (MILP reformulation)

Acquisition function: PWA (exploitation) + Exploration function

Initial Sampling Phase:

- Box constraints only: Latin Hypercube Sampling (LHS)
- Linear constraints with integer and/or categorical variables:
  - Try LHS first and discard any infeasible samples; if not sufficient,
  - Then, solve a MILP problem to sequentially generate samples

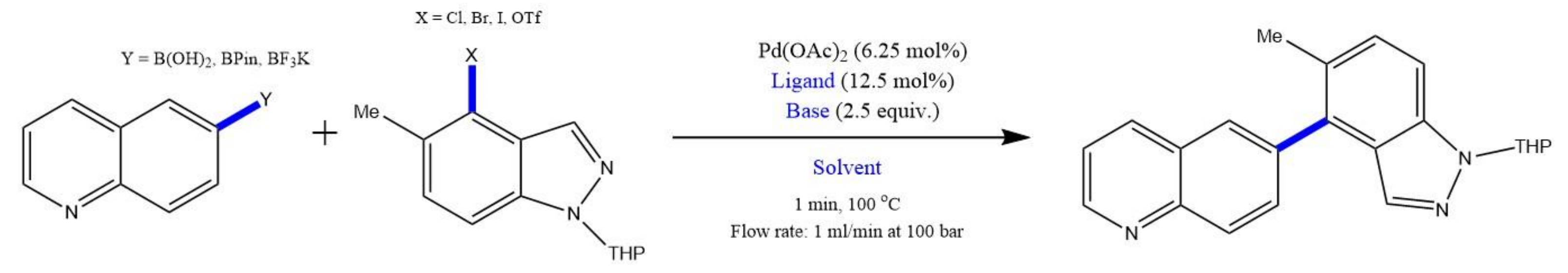
Active Learning Phase:

- Adaptively update/refit the surrogate function (PARC)
- Incorporated distance-based exploration function
- Solve a MILP problem to sequentially generate samples

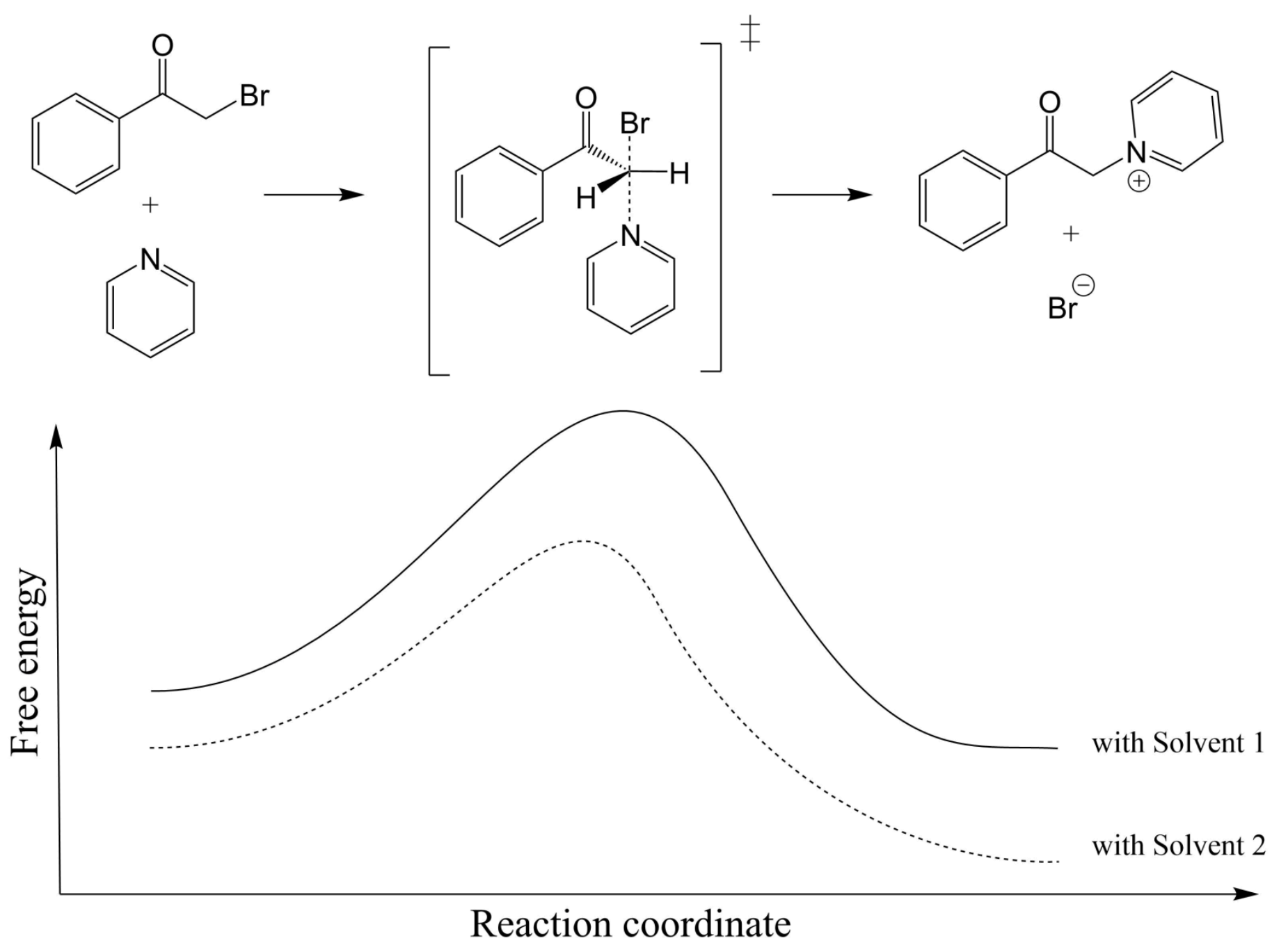
CASE STUDIES

PROBLEM DESCRIPTION

1. Reaction condition optimization (Suzuki-Miyaura cross-coupling)



2. Solvent design (Menschutkin reaction)



Optimization variables	# options
Aryl halide (X)	4
Boronic acid derivative (Y)	3
Base	7
Ligand	11
Solvent	4
Total # of possible combinations	3,696

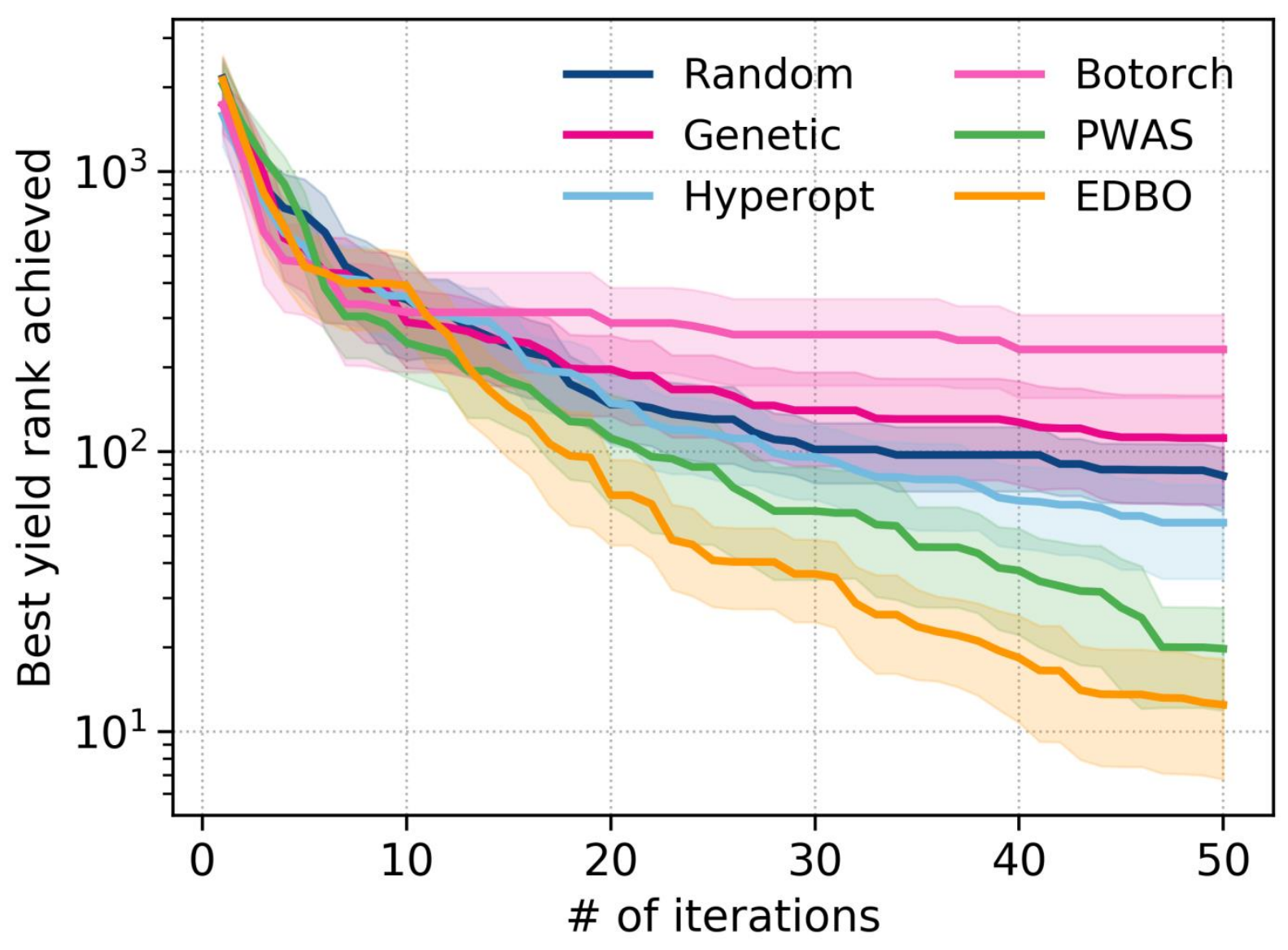
1. Fully categorical; **Max. yield**
2. Mixed integer and categorical
   - With linear constraints
   - **Max. reaction kinetics**

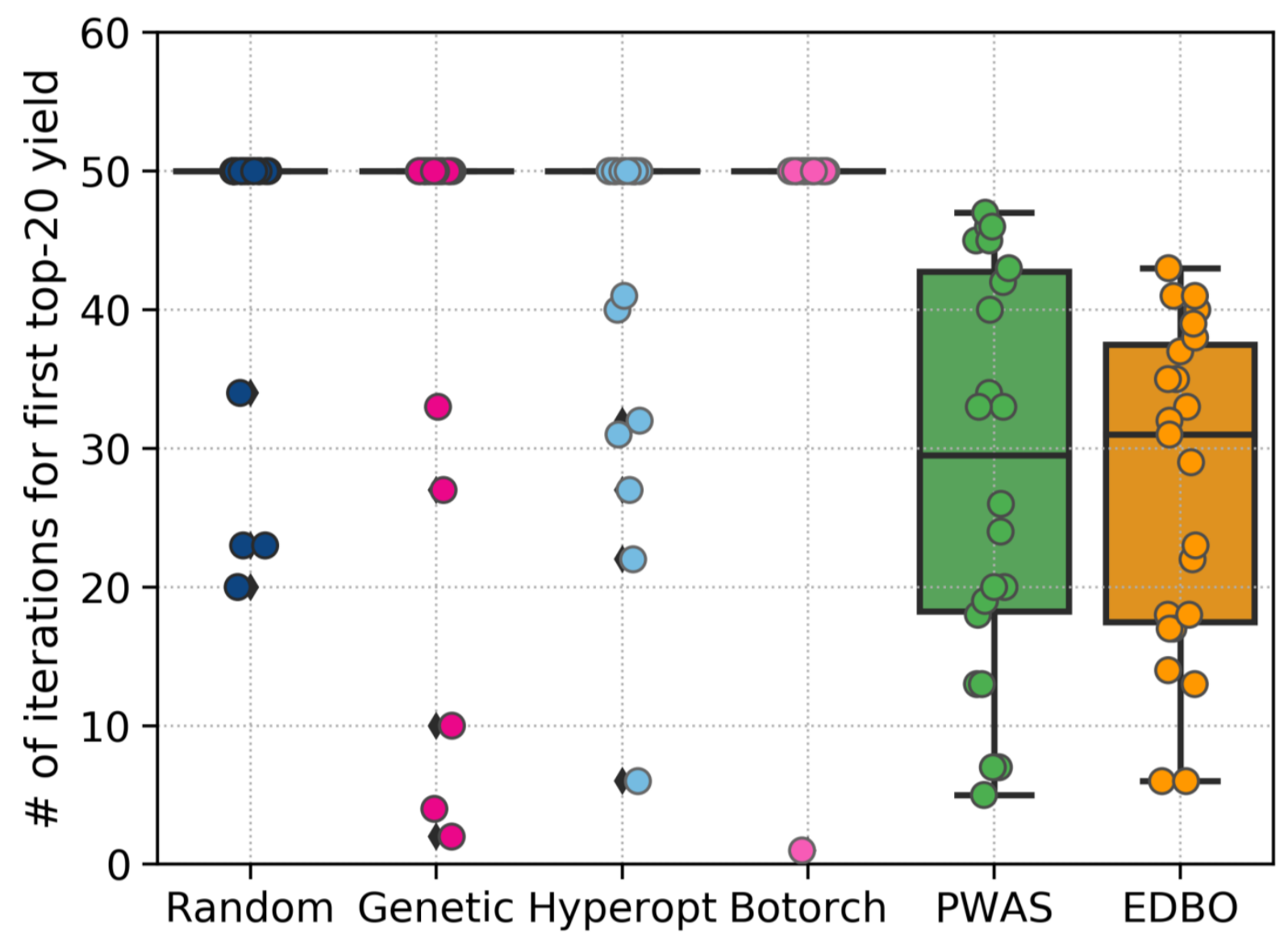
Compare PWAS with genetic algorithm and three BO variants

Number of functional group types	46 (integer)
Number of auxiliary variables introduced for chemical feasibility	1 (categorical) and 7 (binary)
Number of inequality/equality design constraints	115 (linear) / 5 (linear)

RESULTS HIGHLIGHTS

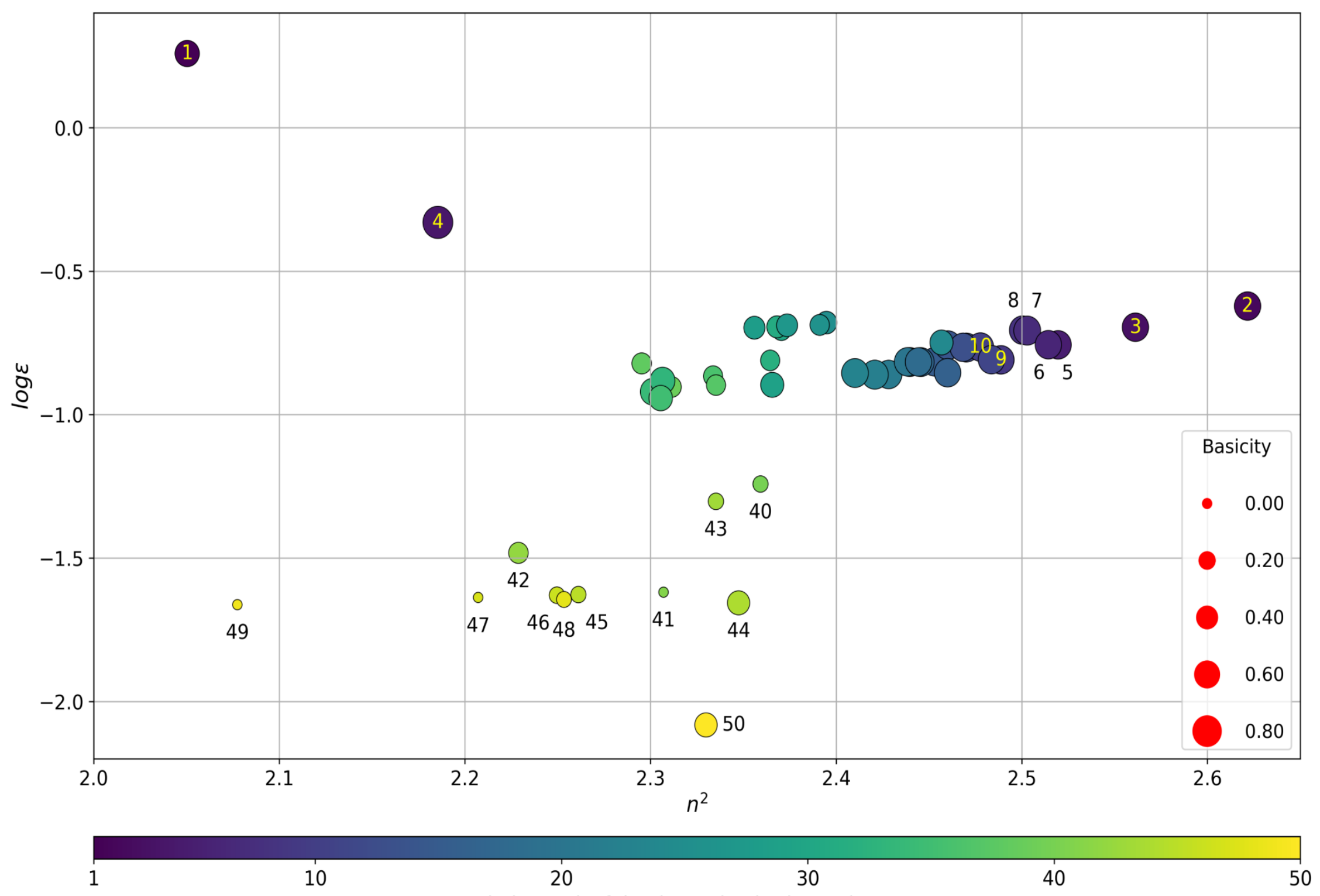
1. **Reaction condition optimization** (Suzuki-Miyaura cross-coupling)





2. Solvent design

- The dielectric constant (ϵ) is found to be the predominant factor influencing reaction kinetics
  - Align with the established results: favour polar aprotic solvents
- PWAS can identify feasible solvents with satisfactory performance
- PWAS can learn correlations between solvent properties and reaction rates and offer valuable insights.



CONCLUSIONS AND FUTURE WORK

- Addressed the experimental planning problems with discrete and mixed variables, subject to linear equality/inequality constraints
- Demonstrated the effectiveness of mixed-integer surrogates and acquisition function (PWAS)

Future Work:

- Extend the framework to handle nonlinear constraints
- Integrate exploration strategies in PWAS to BO methods
- Implement and integrate with automated/autonomous lab