

exploring BO for chemical reaction optimisation

Austin Mroz

Eric & Wendy Schmidt AI in Science Postdoctoral Fellow

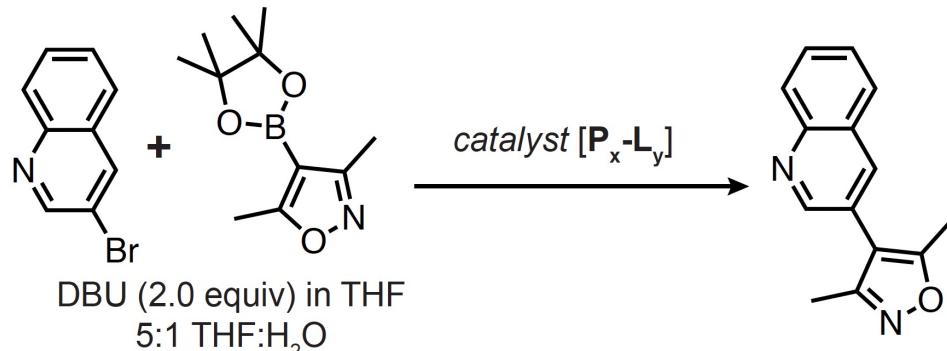
github.com/austin-mroz

a.mroz@imperial.ac.uk

February 2026



Suzuki-Miyaura cross-coupling reaction



parameter

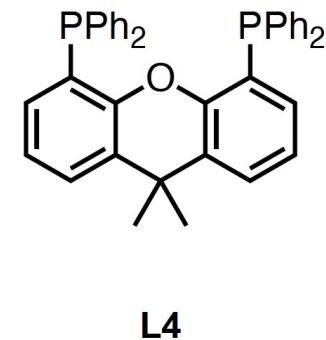
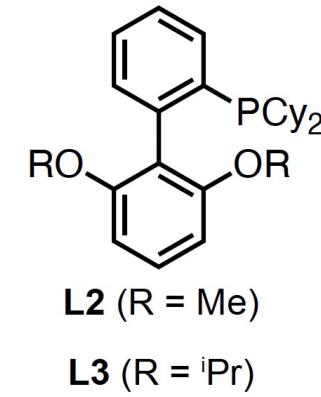
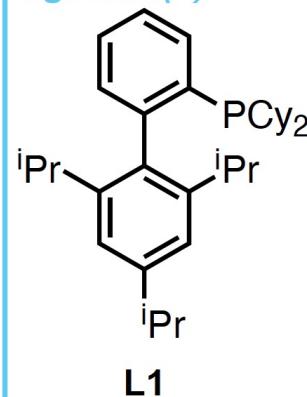
catalyst $[P_x-L_y]$

catalyst loading

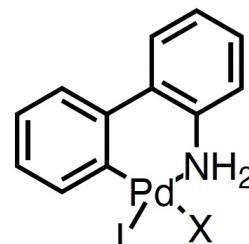
temperature

residence time

ligands (L)



precatalyst scaffolds (P)



P1 (X = OMs)

P2 (X = Cl)

L5 PCy₃

L6 PPh₃

L7 P^tBu₃

key questions

explore **chemical representation strategies** with



- **How do different catalyst encodings affect optimization?**
 - What are the conceptual differences between OHE, fingerprints, and Mordred descriptors, and why might they lead to different BO performance?
 - How does this change the BO behavior and the quality/speed of the optimization?
- **How does the choice and number of initial experiments affect:**
 - the speed at which yield improves, and
 - the best yield reached within ~200 experiments?
- **How would you extend to a multi-objective problem?**
 - If you also want to maximize TON, how would you change the objective / targets in BayBE?
 - What kind of multi-objective strategy would you use, and how would you interpret a Pareto front of yield vs TON?

explore **domain & sampling strategies** with



- Are the chosen input bounds (temperature, catalyst loading, residence time) and categorical options (catalysts) chemically sensible?
- **How should you seed BO with initial experiments?**
- How many random initial experiments should you generate?
- **How do different choices of this initial random set affect:**
 - the quality of the first surrogate model, and
 - the subsequent BO trajectory?
 - Can you implement and diagnose the full BO loop?
- **Which regions of the domain (temperatures, residence times, catalyst loadings, catalysts) does BOFire eventually favor?**
- **How would you extend to a multi-objective problem?**
 - If you also want to maximize TON, how would you change the objective / targets in BayBE?
 - What kind of multi-objective strategy would you use, and how would you interpret a Pareto front of yield vs TON?

multi-objective BO

