# A framework for fully autonomous design of materials via multiobjective optimization and active learning: challenges and next steps

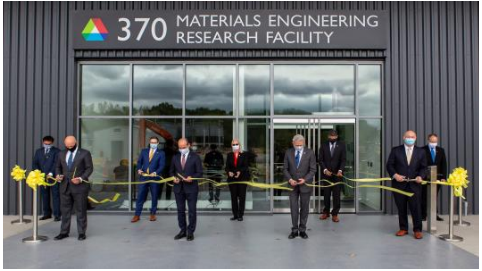
Tyler Chang[⋆a], Jakob Elias[b], Stefan Wild[a→c], Santanu Chaudhuri[b,d] & Joe Libera[b]

[a]Mathematics and Computer Science Division, Argonne National Laboratory
[b]Applied Materials Division, Argonne National Laboratory
[c]Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory
[d]Department of Civil, Materials, and Environmental Engineering, University of Illinois Chicago

ICLR 2023 Workshop on ML4Materials (May 2023)

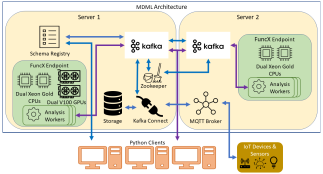
# Big Picture

The Argonne Material Engineering Research Facility (MERF):

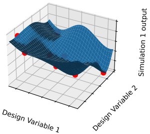


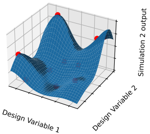

Accelerate experimentation → production pipeline with ML:

←




→



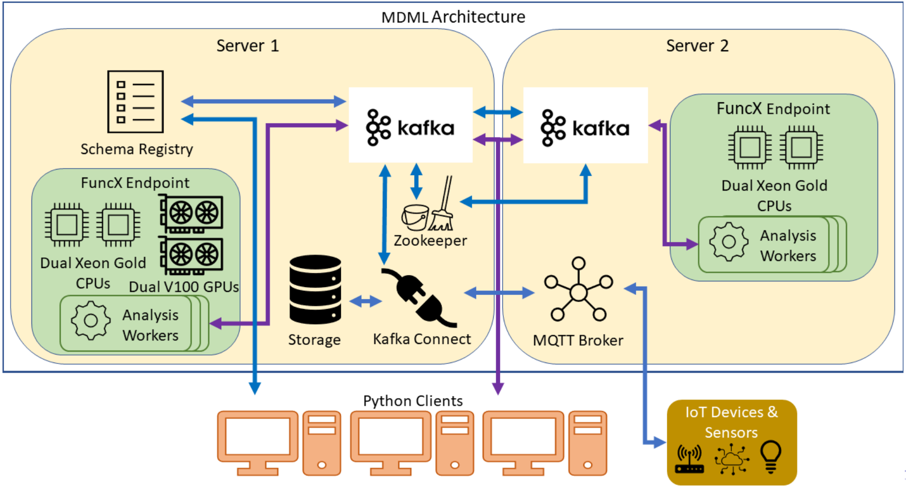



# Our (Big) Goals

- Design a software framework for *self-driving labs*
- Accelerate discovery via intelligent experimentation
- Democratize lab-work by building open-source tools

# Streaming data from multiple sources

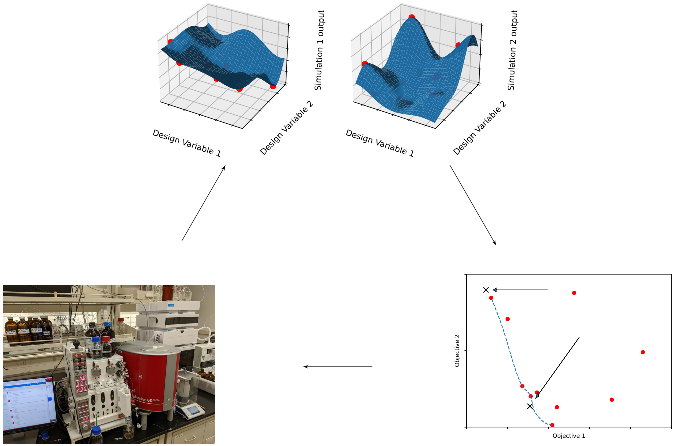
- How to collect and analyze data?
- MDML is a platform for streaming, analyzing, and logging experiment and simulation data (used at the MERF)



# Model-based optimization

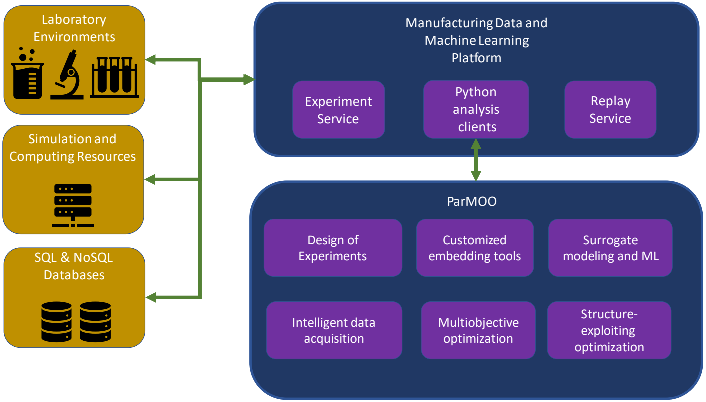
- ParMOO (multiobjective optimization) library is used to implement an active learning loop

ParMOO fits surrogate models to the experiment/simulation data



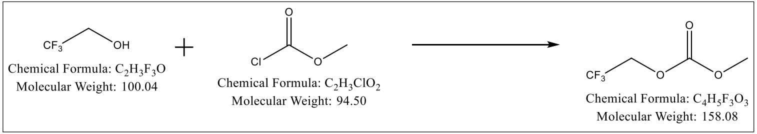
# Our framework and software stack

- MDML gives us access to *heterogeneous data from laboratory sources*
- ParMOO gives us modular/customizable *modeling, embedding, and solvers*



# Example: TFMC Manufacturing Conditions

Optimize the production of TFMC via a known reaction...




Trifluoroethanol

Methyl chloroformate

2,2,2-trifluoroethyl
Methyl carbonate

Solvent: acetonitrile

Base: N,N-Diisopropylethylamine

... in LabVIEW automated CFR
and measured via NMR

Left to right:
PC running LabVIEW
CFR and feed
NMR spectroscope

# Problem definition

- **Want to maximize TFMC production at high temperatures**
- High temperatures trigger a side-reaction and produces byproduct (TFE)
  - **Minimize TFE production**

*Design variables and bound constraints for experiment:*

| Parameter | Lower bound | Upper bound |
|---|---|---|
| Temperature (degrees C) | 40 | 150 |
| Reaction time (seconds) | 60 | 300 |
| Equivalence ratio (no units) | 0.9 | 2 |

# Challenges and Solver Settings

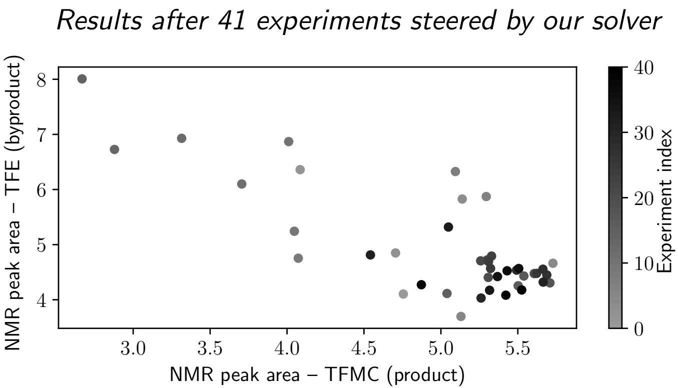
Solve a small problem on an extremely limited budget:

- 3 variable, 2 objective problem
- Experiments take about 10 mins and we have limited supply of reagent materials
- About 50 experiments max

Solver settings:

- 15-pt Latin hypercube, Gaussian RBF surrogate, L-BFGS-B optimizer
- 3 scalarizations per batch, sorted by temp (to speedup reaction)
- evaluated batch on CFR, TFMC and TFE peaks recorded by NMR

# Experiment Results

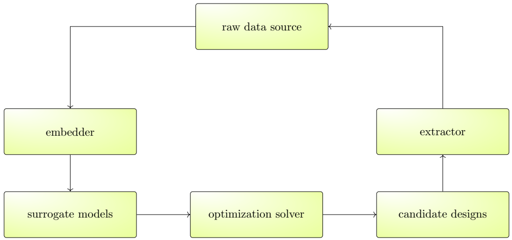


# Next Steps

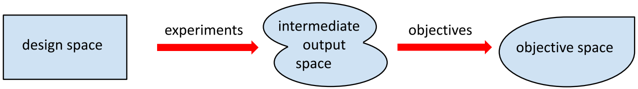
Need to handle more complex design spaces:

- Custom embeddings (including generative AI)
- Trust-region descent / subspace iterations

- Custom surrogate models
- Structure-exploiting optimizers




(Top) using custom embedders to optimize in latent space



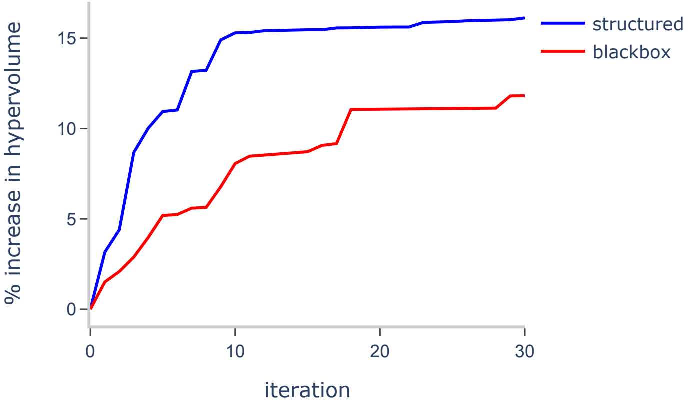

(Bottom) exploiting problem structure using composite objectives

# A sneak peek!

We created a surrogate problem based on this data to explore next steps!

`https://github.com/parmoo/parmoo-solver-farm`
$\rightarrow$ `cfr-material-design-2022`

5 variable (2 categorical), 3 objectives (1 cheap) problem



Email: `tchang@anl.gov`

We love open source!

```
git clone https://github.com/parmoo/cfr-materials
pip install REQUIREMENTS.txt
```

This material was based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics and SciDAC programs under Contract Nos. DE-AC02-05CH11231 and DE-AC02-06CH11357 and by the Argonne LDRD program.