A framework for fully autonomous design of materials via multiobjective optimization and active learning

Tyler Chang[1], Jakob Elias[1], Stefan Wild[2], Santanu Chaudhuri[1,3] and Joseph Libera[1]

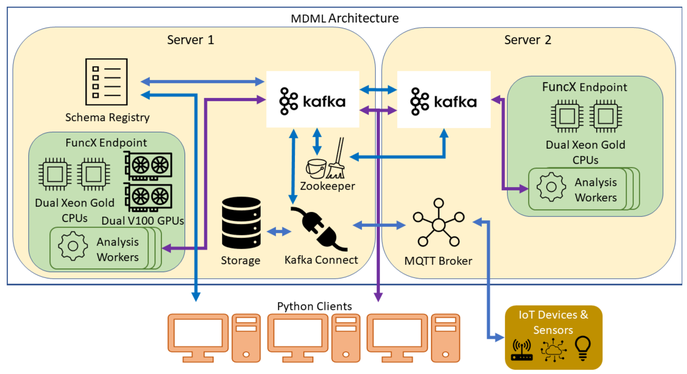
[1]*Argonne National Laboratory* [2]*Lawrence Berkeley National Laboratory* [3]*University of Illinois Chicago*

Our (Big) Goals

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

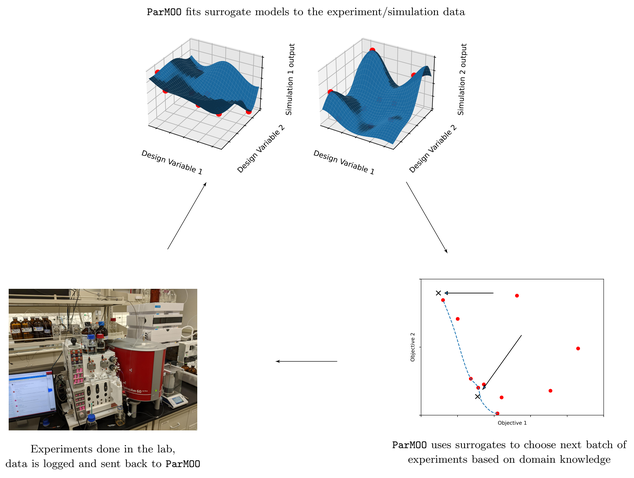
Streaming data from multiple sources

- How to collect and analyze data?
- MDML is a platform for streaming, analyzing, and logging experiment and simulation data



Model-based optimization

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

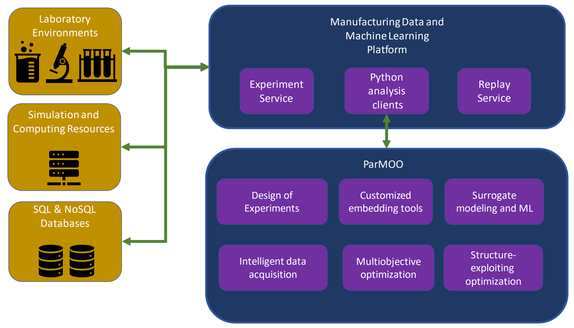


Experiments done in the lab, data is logged and sent back to ParMOO

ParMOO uses surrogates to choose next batch of experiments based on domain knowledge

Our framework and software stack

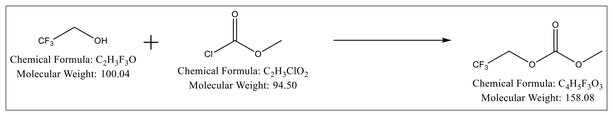
- MDML gives us access to *heterogeneous data from laboratory sources*
- ParMOO gives us modular/customizable *modeling, embedding, and solvers*



Build & deploy custom solvers for computational and experimental problem!

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

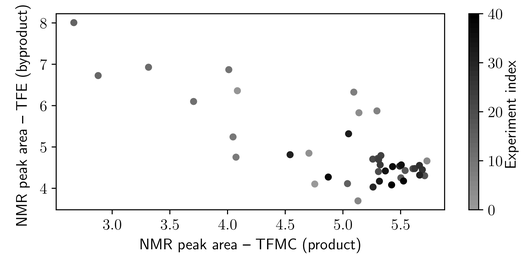
Design variables and bound constraints for experiment

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

Experiment Results

- **Want to maximize TFMC production at high temperatures**
- High temperatures trigger a side-reaction and produces byproduct (TFE)
- 15-pt Latin hypercube, Gaussian RBF surrogate, L-BFGS-B optimizer
- 3 scalarizations per batch, sorted by temp
- evaluated batch on CFR, TFMC and TFE peaks recorded by NMR

Results after 41 experiments steered by our solver



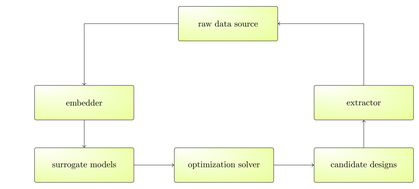
Get this code!

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

Next Steps

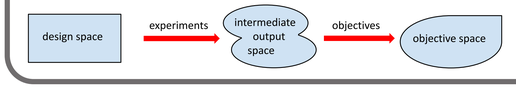
Need to handle more complex design spaces:

- Generative AI for embeddings
- Trust-region descent methods
- Subspace iterations
- Custom surrogate models
- Structure-exploiting optimizers



(Top) using custom embedders to optimize in latent space

(Bottom) exploiting problem structure using composite objectives



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.