

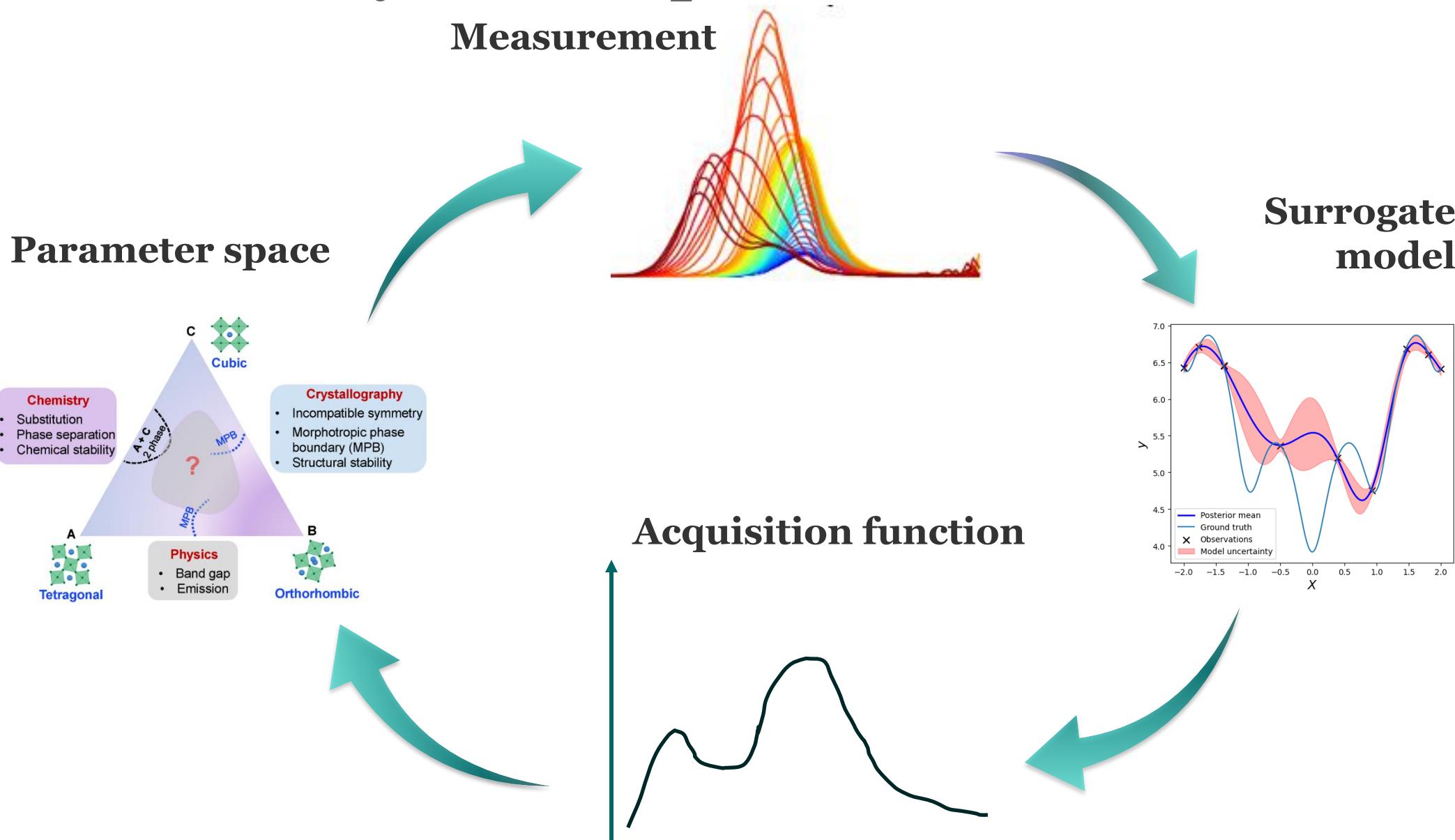
Lecture 26: Multifidelity and Multiobjective BO

Sergei V. Kalinin

This and that

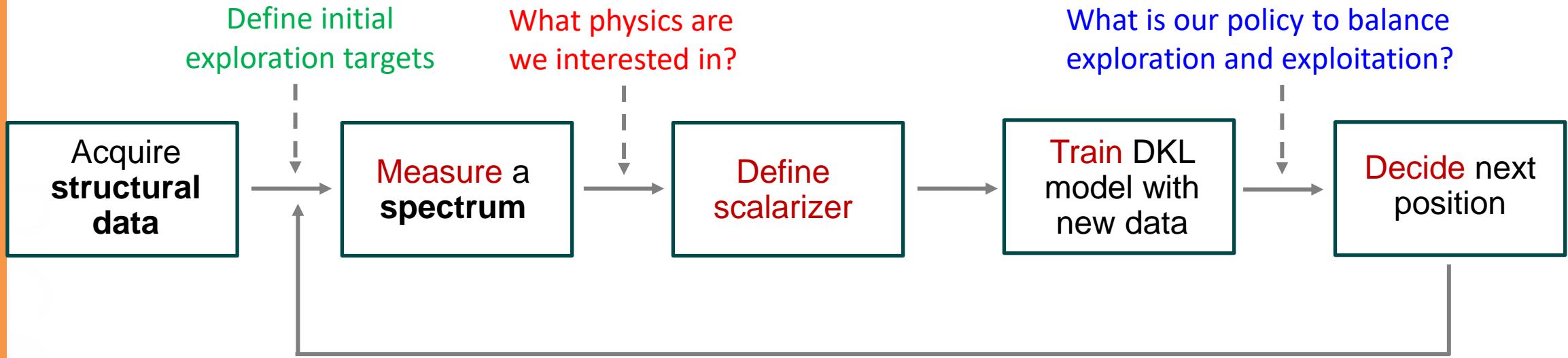
- Thursday – Q&A with Sherryl
- Final: Pick a project:
 - BO of 2D Ising model
 - Learning in latent space
 - Making hypotheses

Classical Bayesian Optimization



Implicitly, we postulate that **cost** and **gain** are constant along the iterative cycle

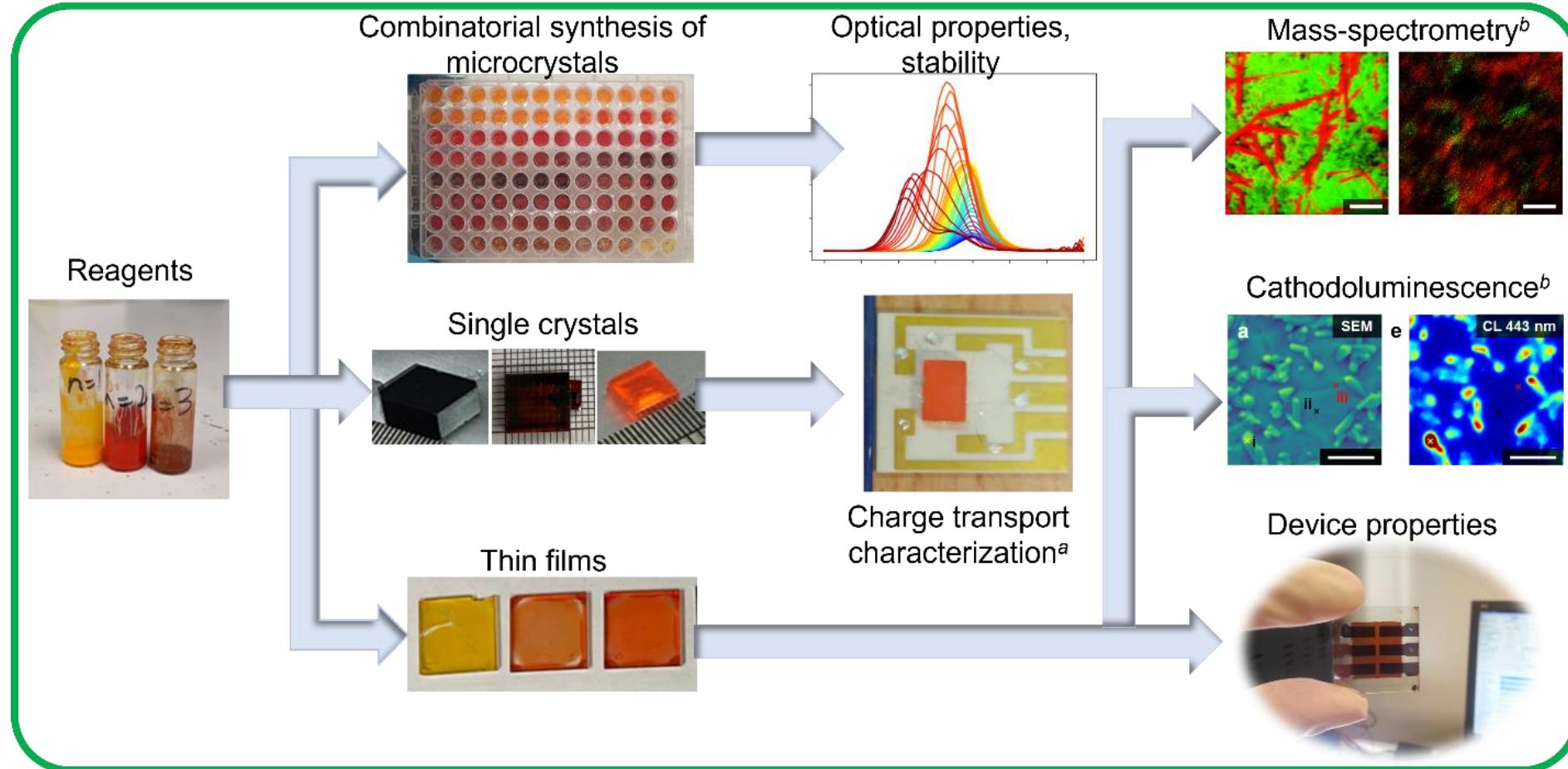
hAE Myopic Workflows



Key concepts:

- Reward (objective)
- Cost
- Value

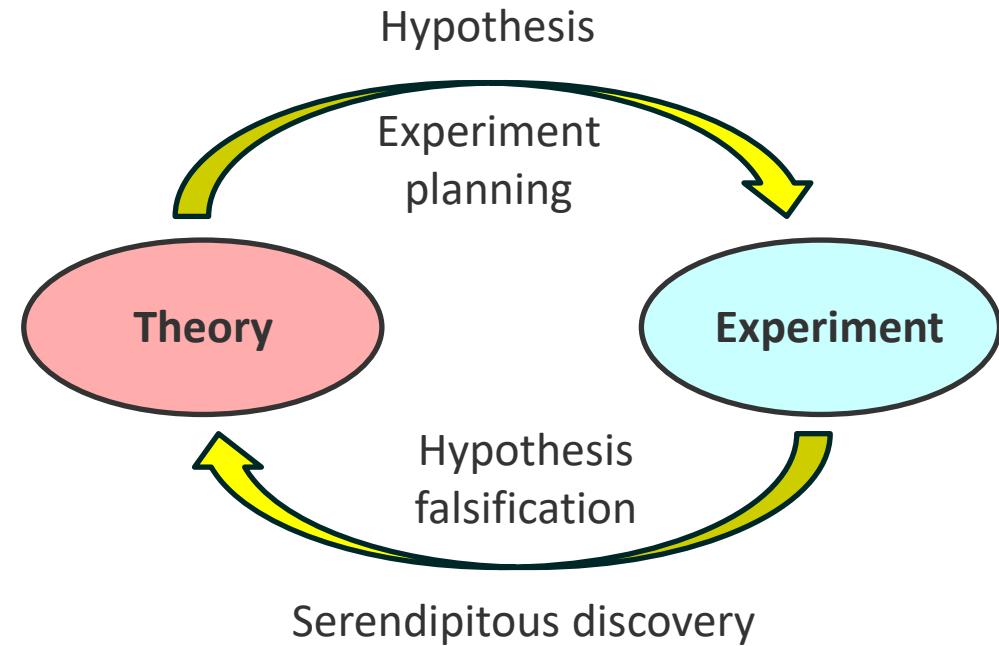
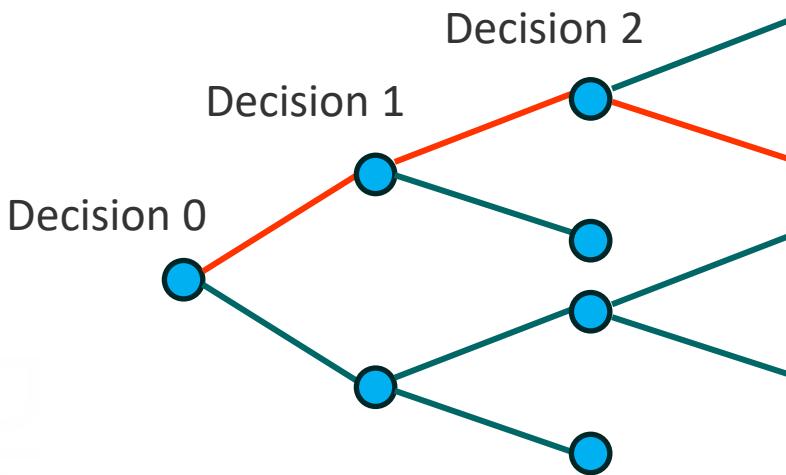
What is A Workflow?



Workflow: ideation, orchestration, implementation

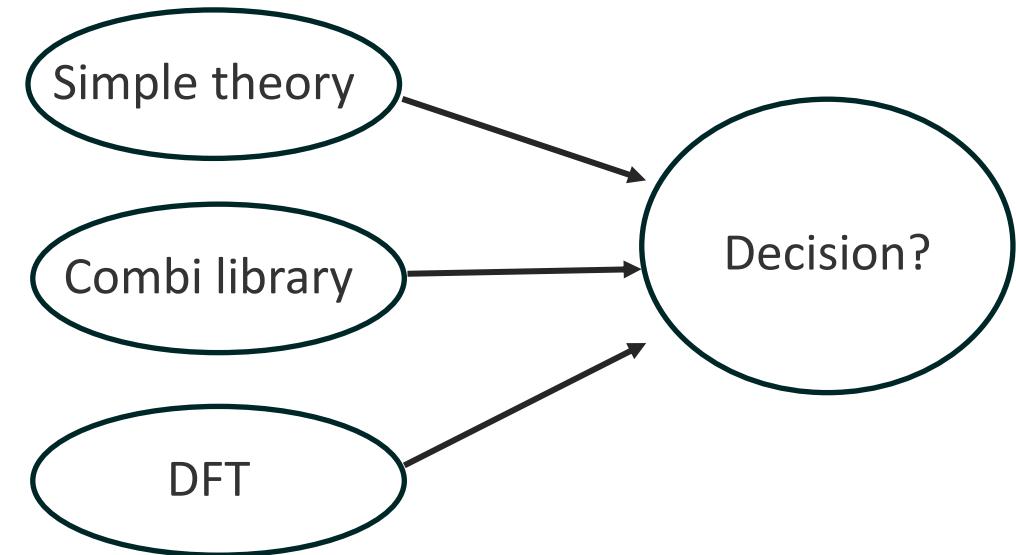
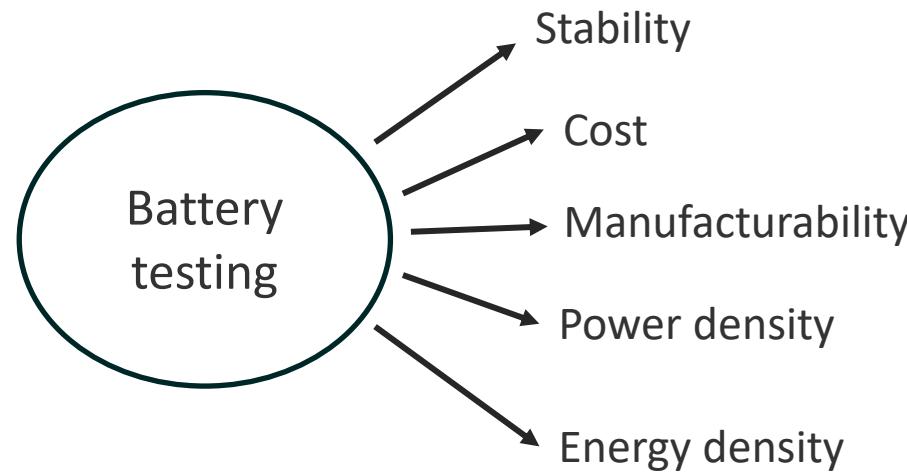
- Domain specific language
- Dynamic planning: latencies and costs
- Reward and value functions
- Are they optimal?
- Can we design them better?
- Can they be changed dynamically?

What do we hope to achieve?



- **Experiment is a combinatorial space of opportunities:**
→ Investing only in scaling of throughput is only a linear improvement
- **Science is a cycle between theory-driven hypothesis generation and experiment:**
→ We need to accelerate all elements of the cycle
- **Experimental and computational tool development:**
→ Currently constrained by human paradigm

The real world is more complex!



1. We need to balance multiple functionalities
2. Integrate multiple sources of data and make decisions considering costs, latencies, and beliefs

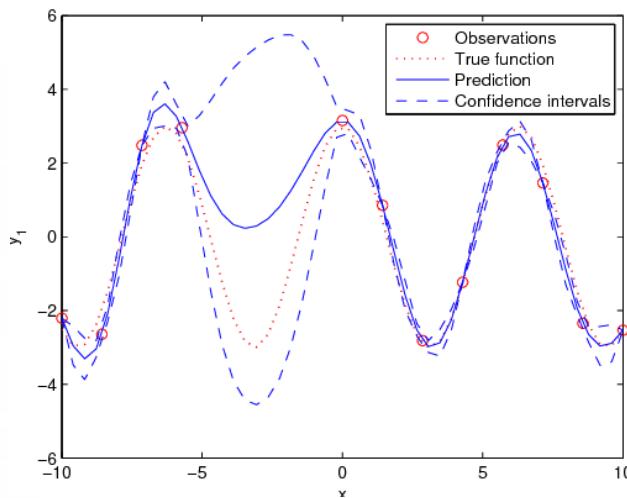
Multi-Objective Bayesian Optimization

Multi-objective Optimization:

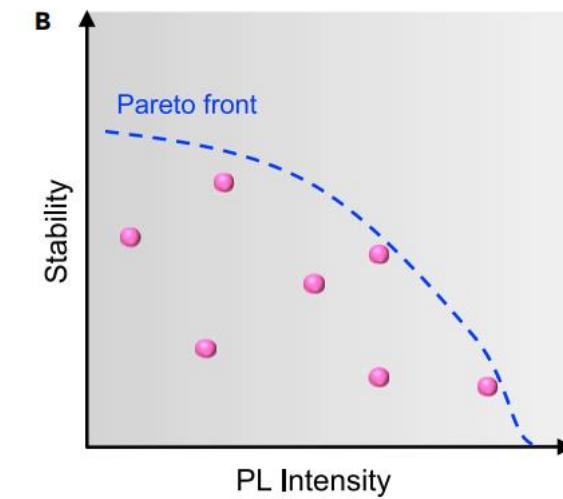
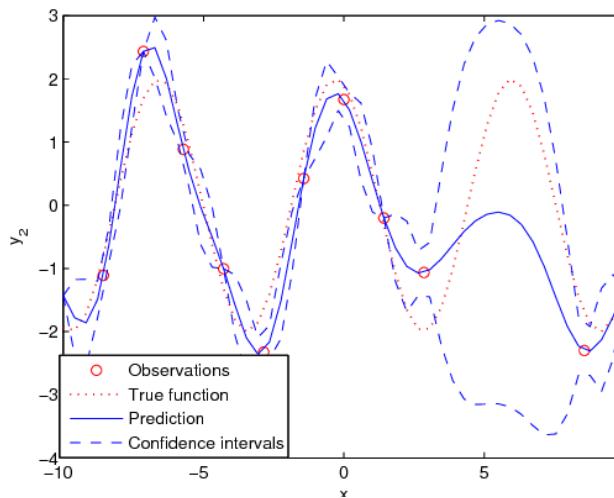
$$\min f(\mathbf{X}) = [\min f_1(\mathbf{X}), \min f_2(\mathbf{X}), \dots, \min f_n(\mathbf{X})] \text{ s.t } \mathbf{X} \in \mathbb{R}$$

Multi-objective Bayesian Optimization: $\min f(\mathbf{X})$ where $f(\mathbf{X})$ is expensive to evaluate

(a) Output 1 (y_1)

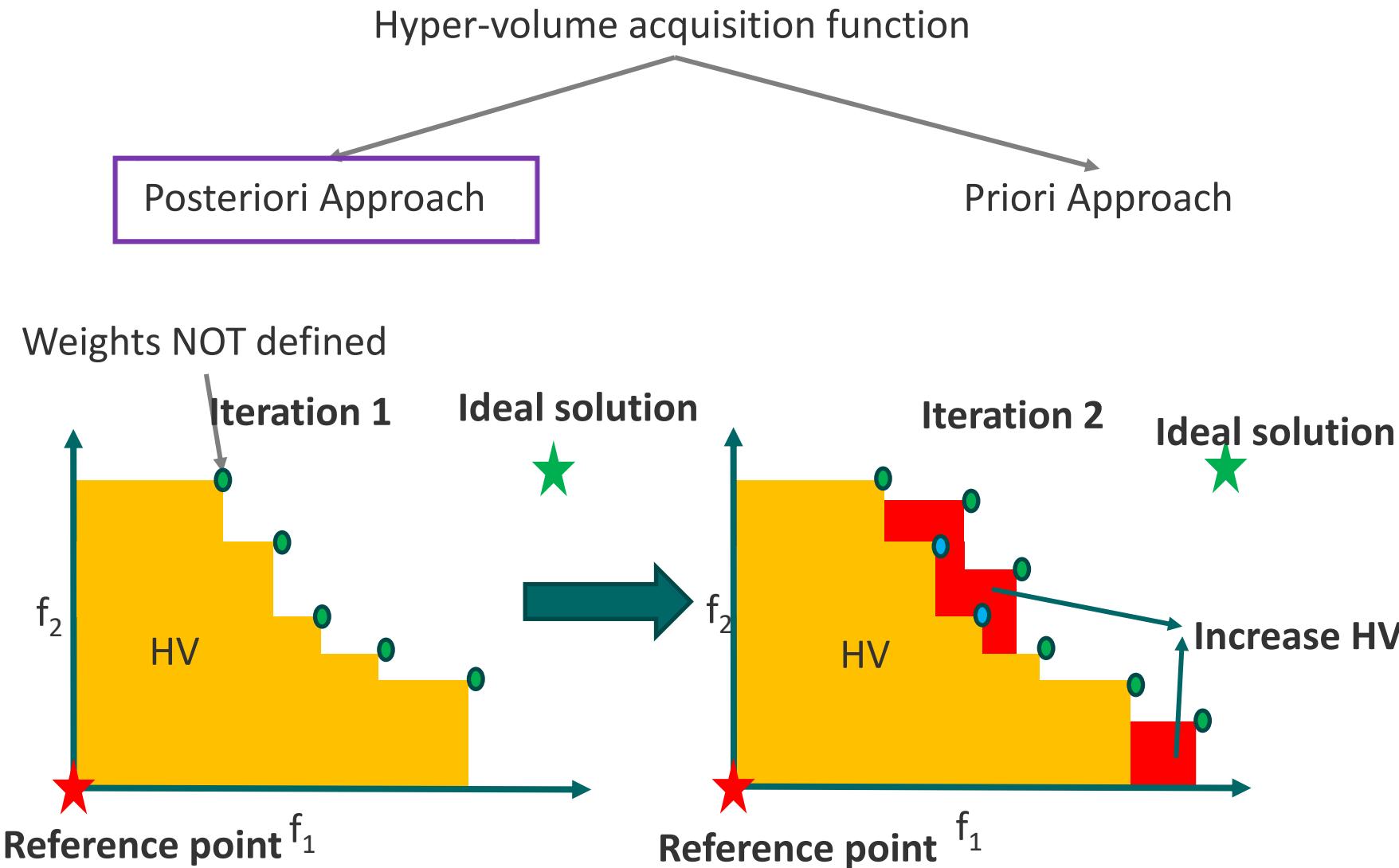


(b) Output 2 (y_2)

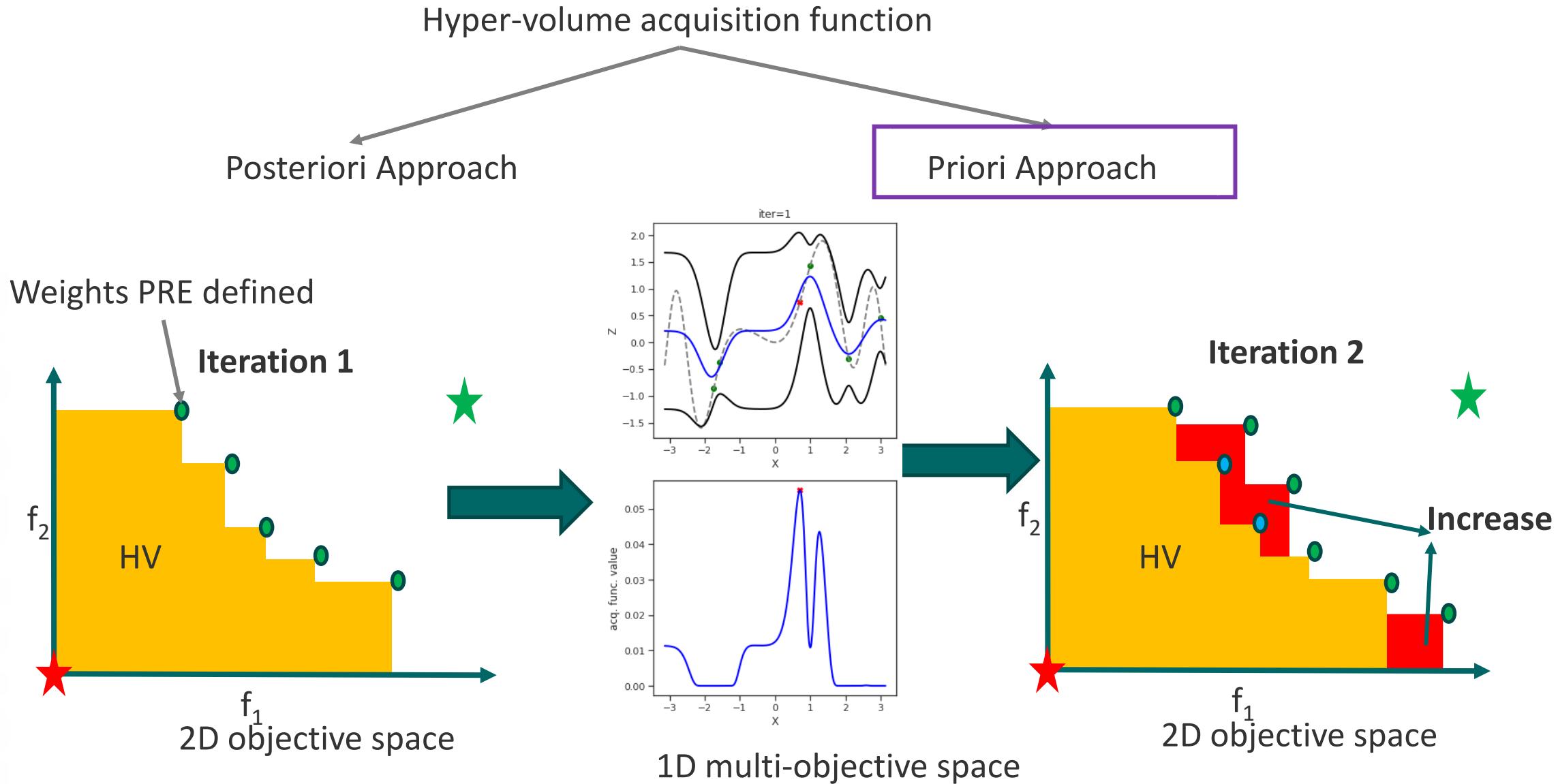


Multi-output Gaussian Process

Multi-Objective Bayesian Optimization

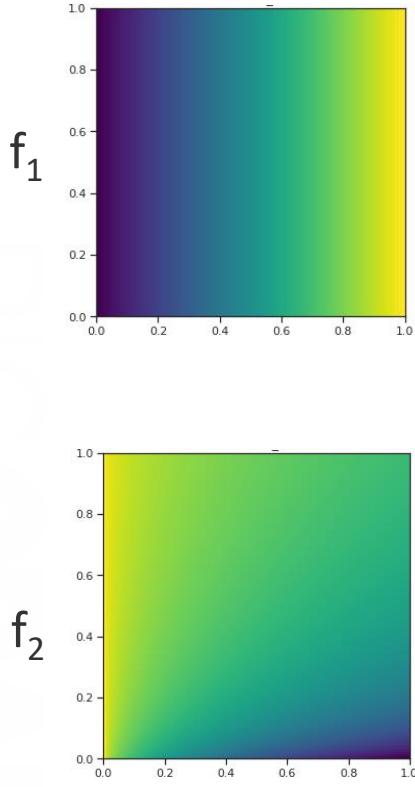


Multi-Objective Bayesian Optimization

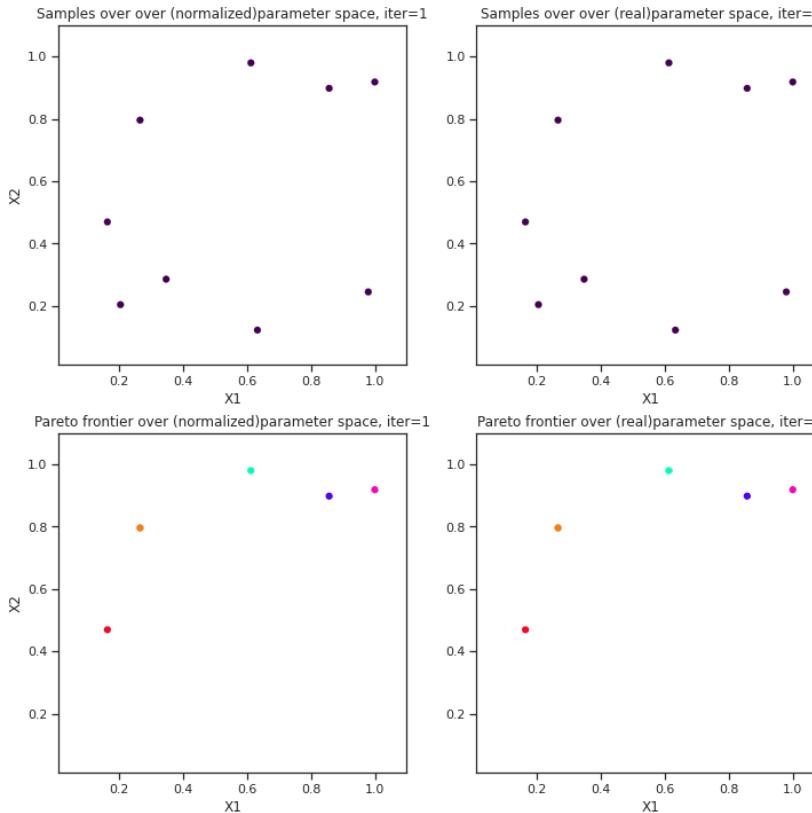


Case Study 1

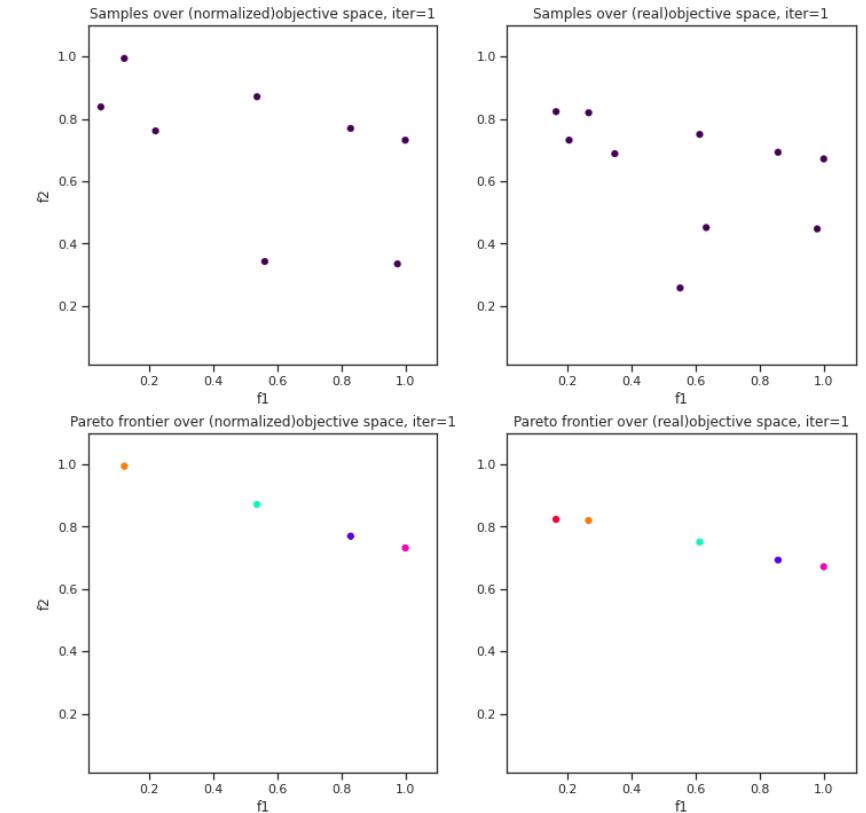
Ground truth



Building Pareto over design space



Building Pareto over objective space

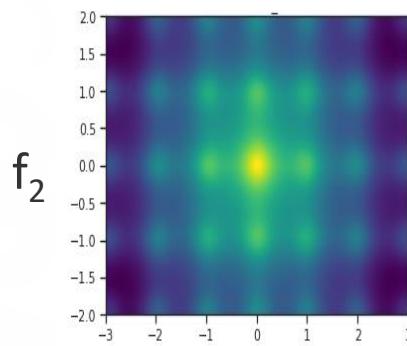
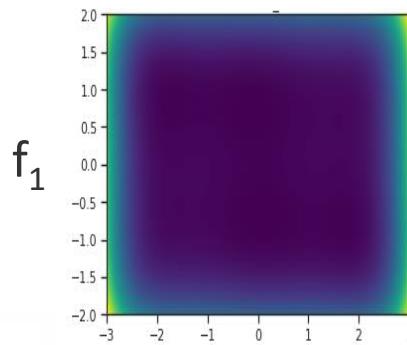


Numerical Test Problems: ZDT1

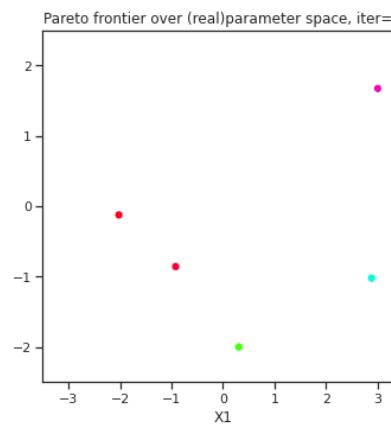
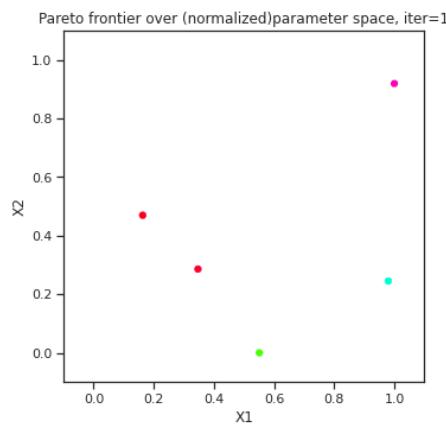
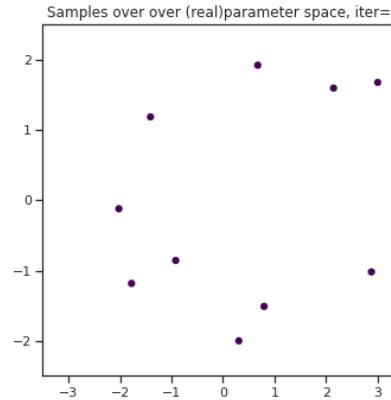
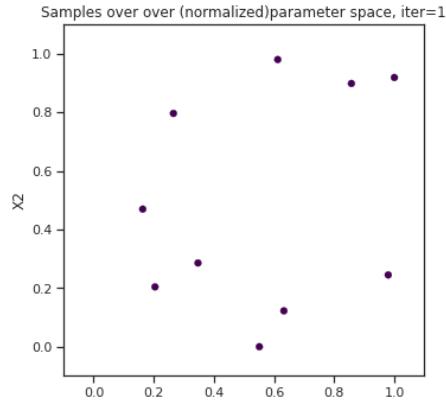
Acquisition Function: qEIHV (Posteriori); Batch_size: 4, Max BO sampling: 50 x Batch_size

Case Study 2

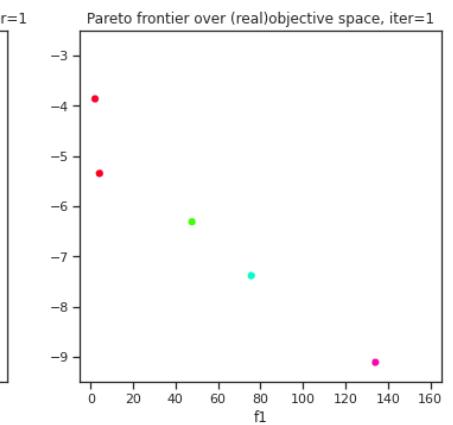
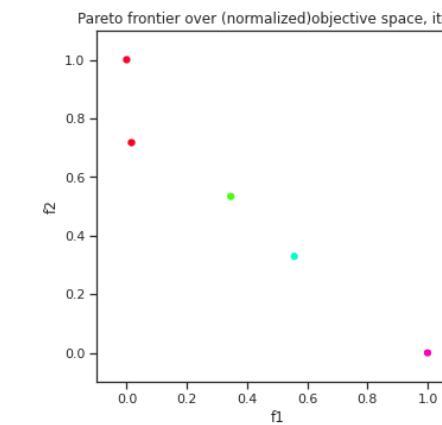
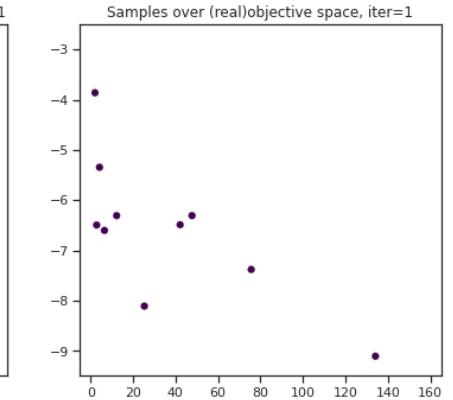
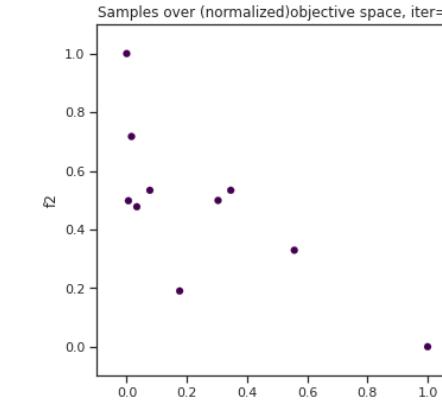
Ground truth



Building Pareto over design space



Building Pareto over objective space

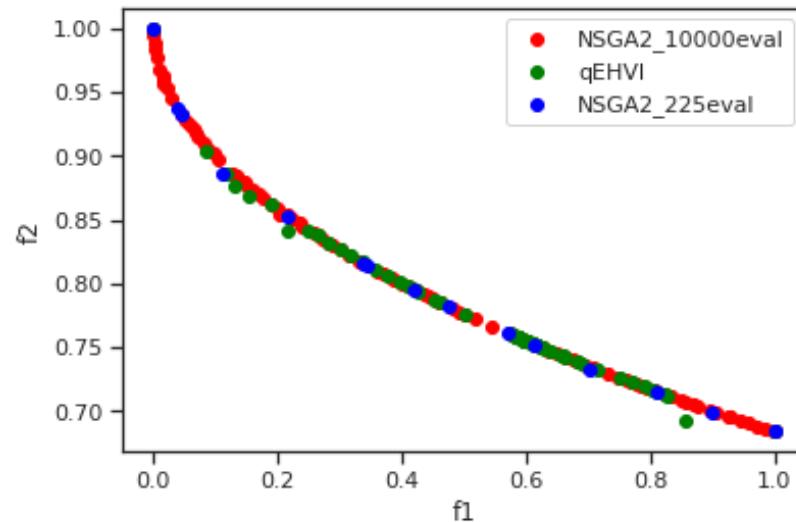
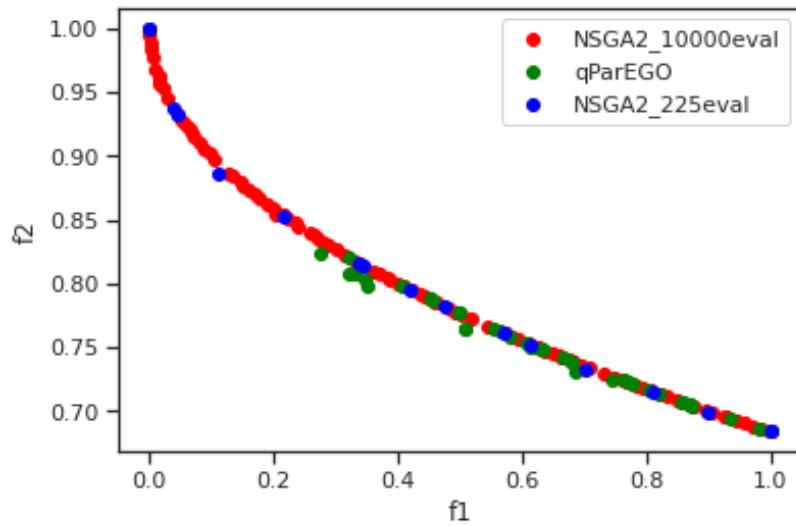


Numerical Test Problems (Non-Physics): 6-Hump Camel Back – Inversed Ackley's Path (6HC-IAP)

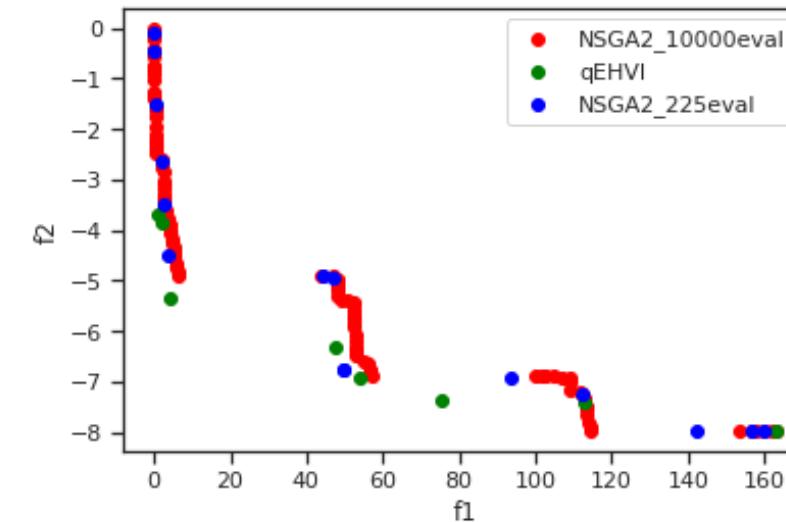
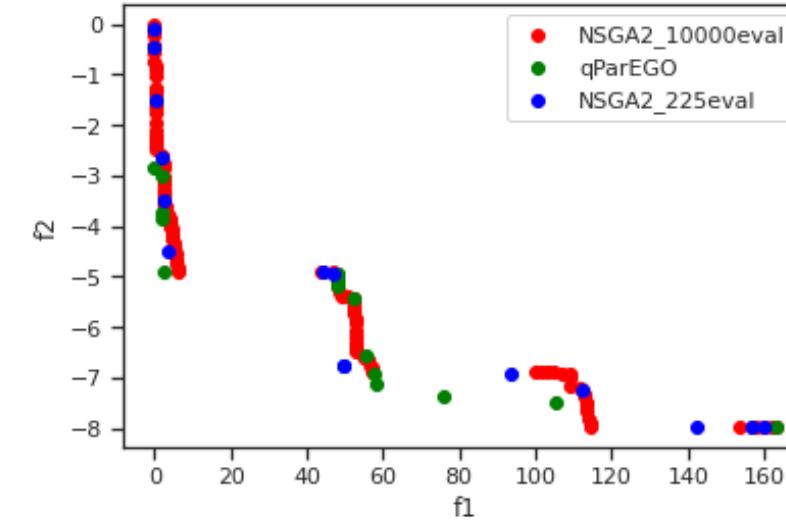
Acquisition Function: qParEGO (Prior); Batch_size: 4, Max BO sampling: 50 x Batch_size

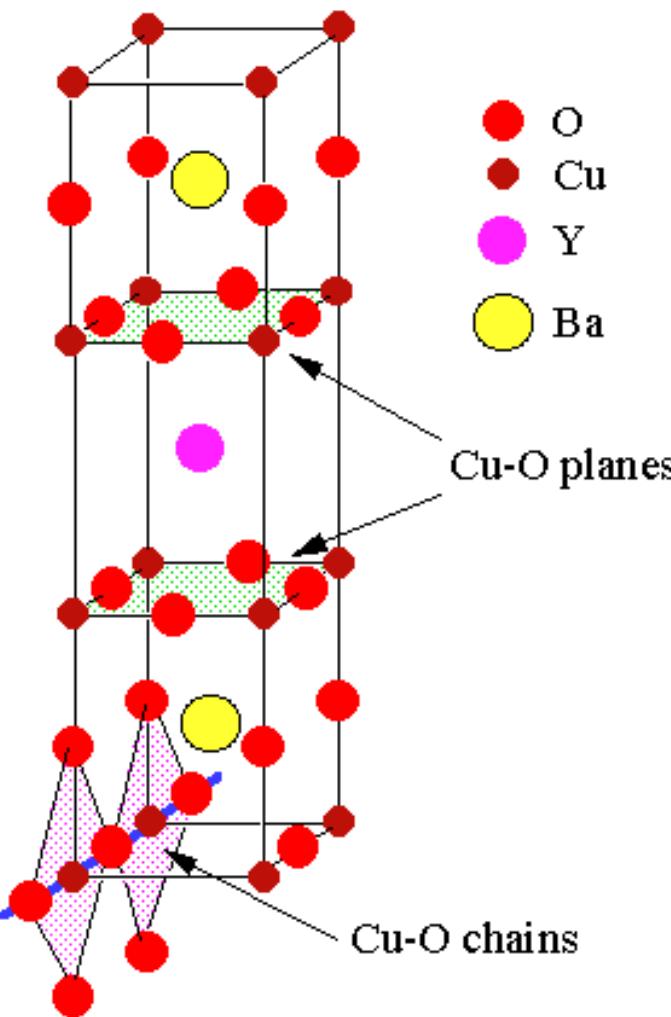
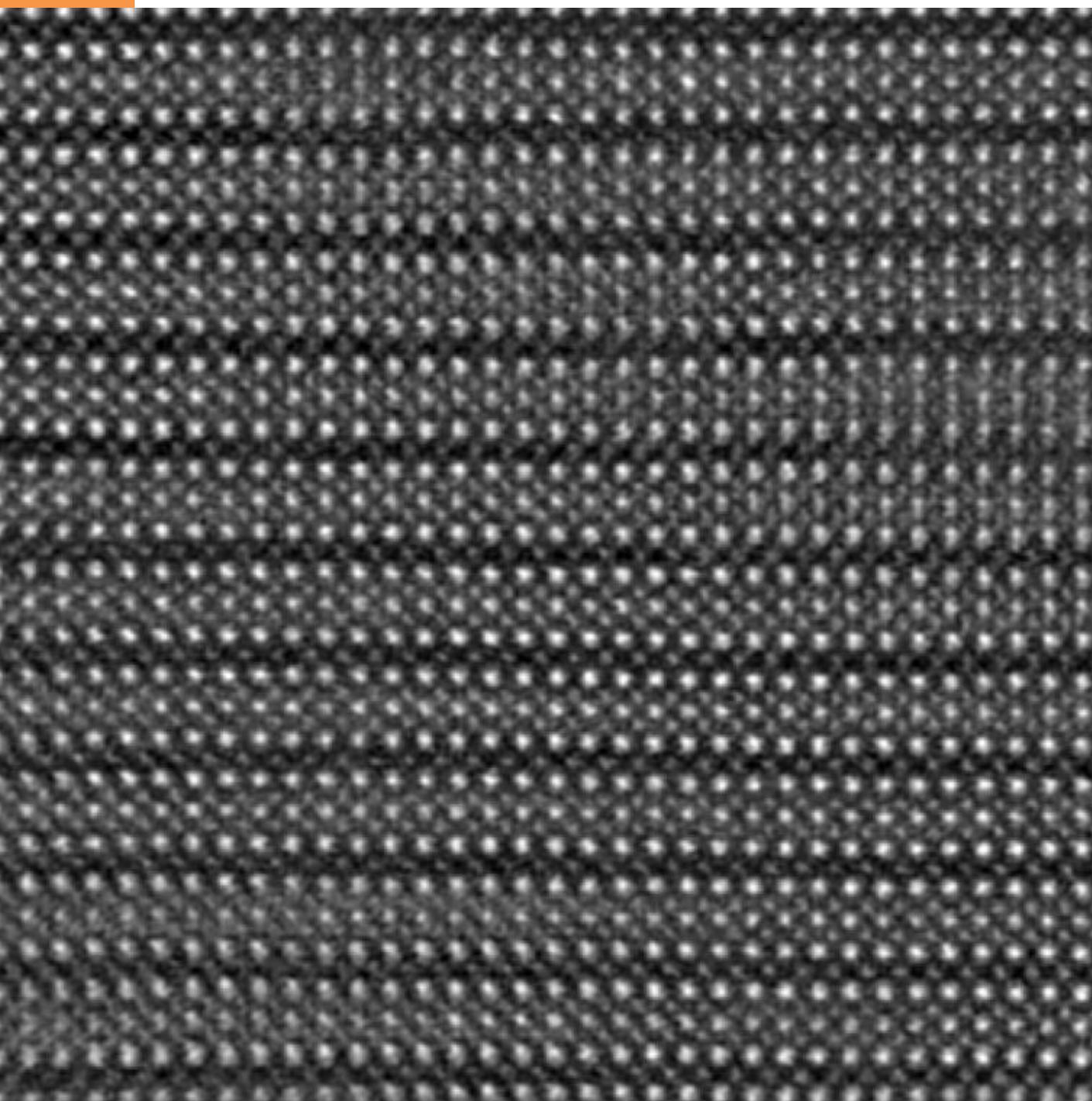
Case Study 1 and 2

ZDT1



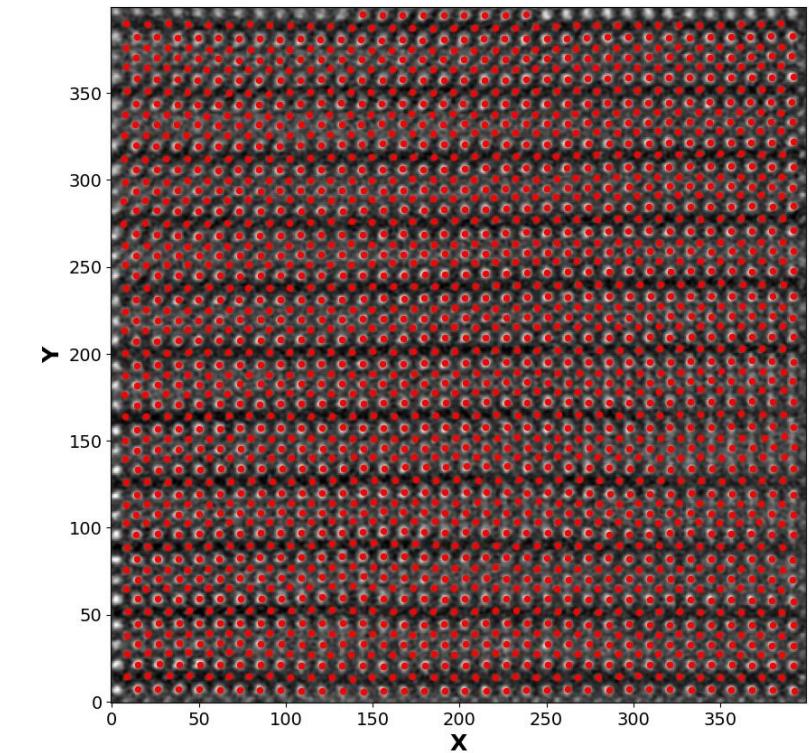
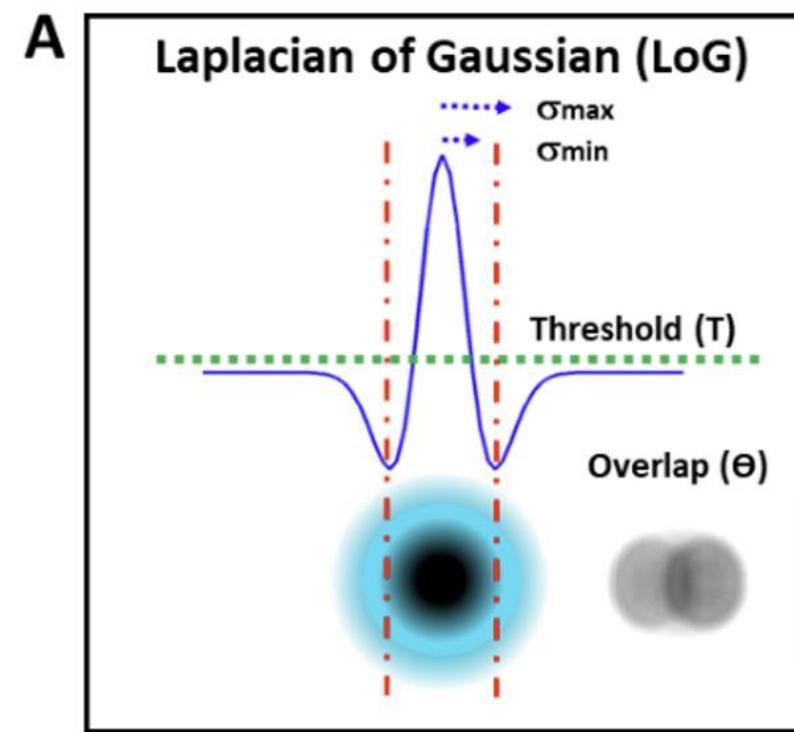
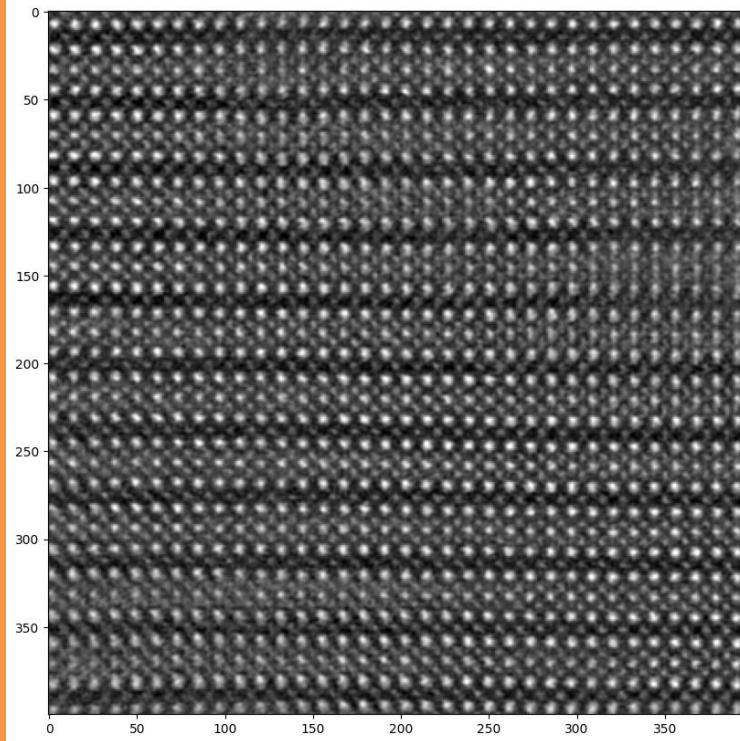
6-Hump Camel Back – Inversed Ackley's Path (6HC-IAP)





- Radiation damage on oxygen sublattice
- Formation of antiphase defects
- Bending of Cu-O layers

Data by A. Goyal and H. Zhang



$\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ thick film, doped with Dy_2O_3 nanoparticles,
Zhang, Y.; Rupich, M.; Solovyov, V.; Li, Q.; Goyal, A. Dynamic behavior of reversible oxygen
migration in irradiated-annealed high temperature superconducting wires. *Scientific Reports*
2020, 10 (1), 14848.

$\sigma_{\min} = (0.1, 3.0)$,

$\sigma_{\max} = (1.0, 6.0)$

Threshold = (0.01, 0.4)

Overlap = (0.01, 0.5)

Quality Function: Objective 1

Quality Count (QC), is defined as the normalized difference between the number of atoms found by Laplacian of Gaussian (LoG) and the number of atoms found defined by Oracle:

$$QC = \frac{LoG \text{ blobs} - Oracle}{Oracle}$$

Error Function: Objective 2

Any atom with a cumulative interaction value less than DS (summation of distances) with its four nearest neighbors will be regarded as not having physical significance, and thus, categorized as an error within this context.

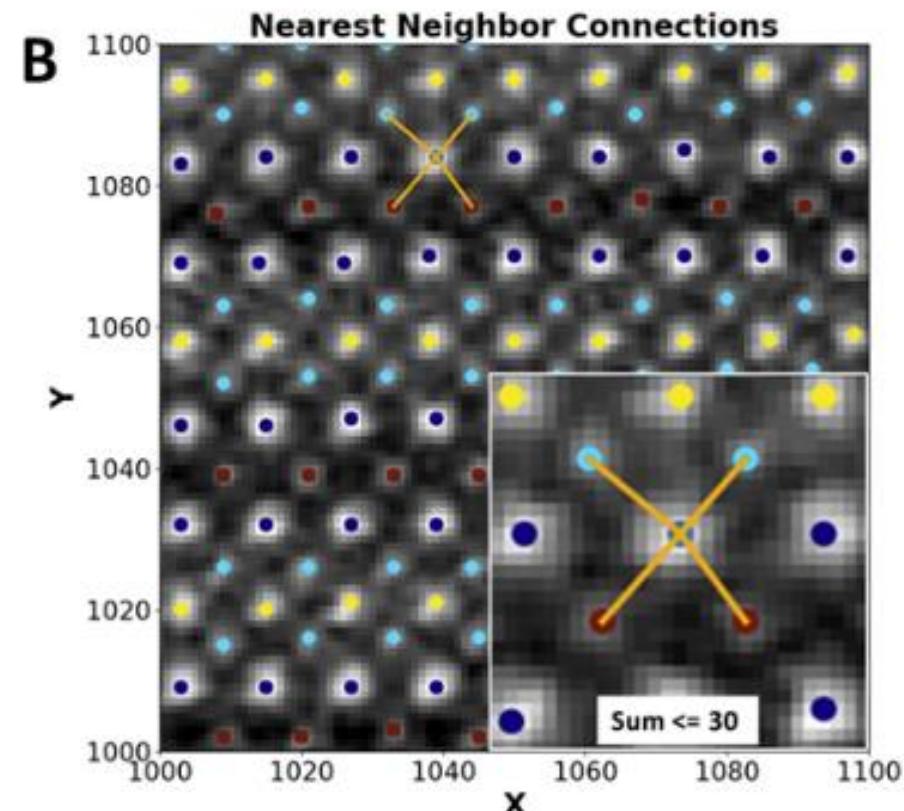
$$ER = \frac{\# \text{ atoms with cumulative interaction value less than } DS}{Oracle}$$

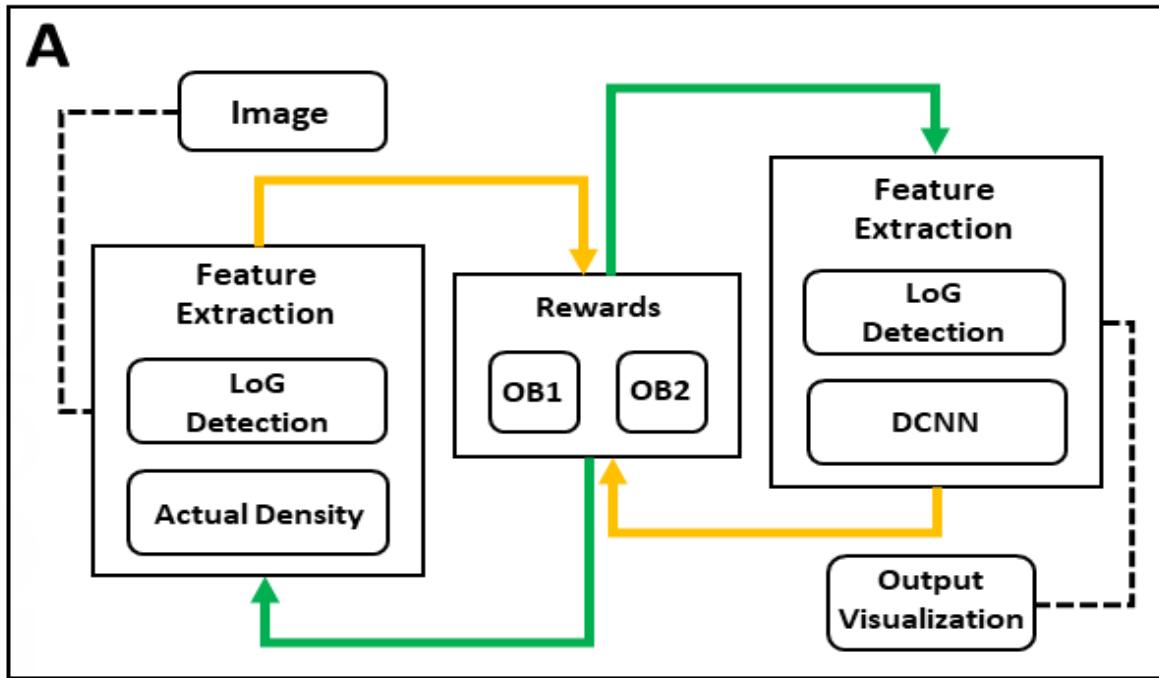
Oracle can be:

- Number of atoms via **physical criterion** – i.e. number of atoms in the structure with respect to stoichiometry

Benchmark:

- Number of atoms via **DCNN**

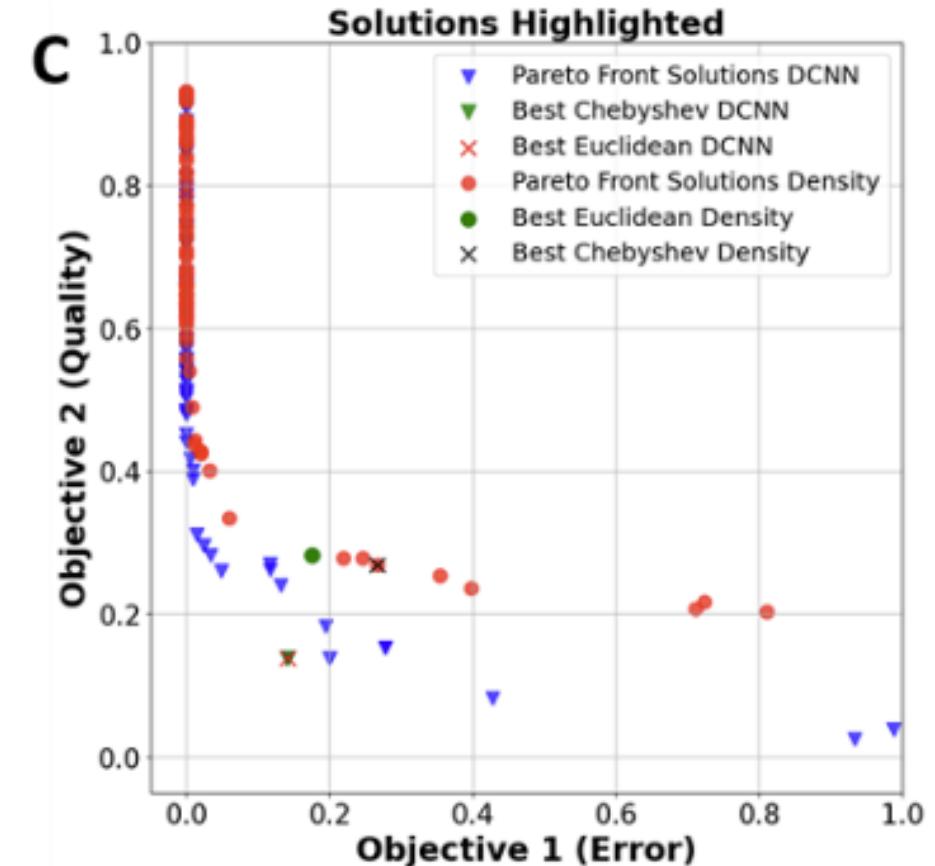




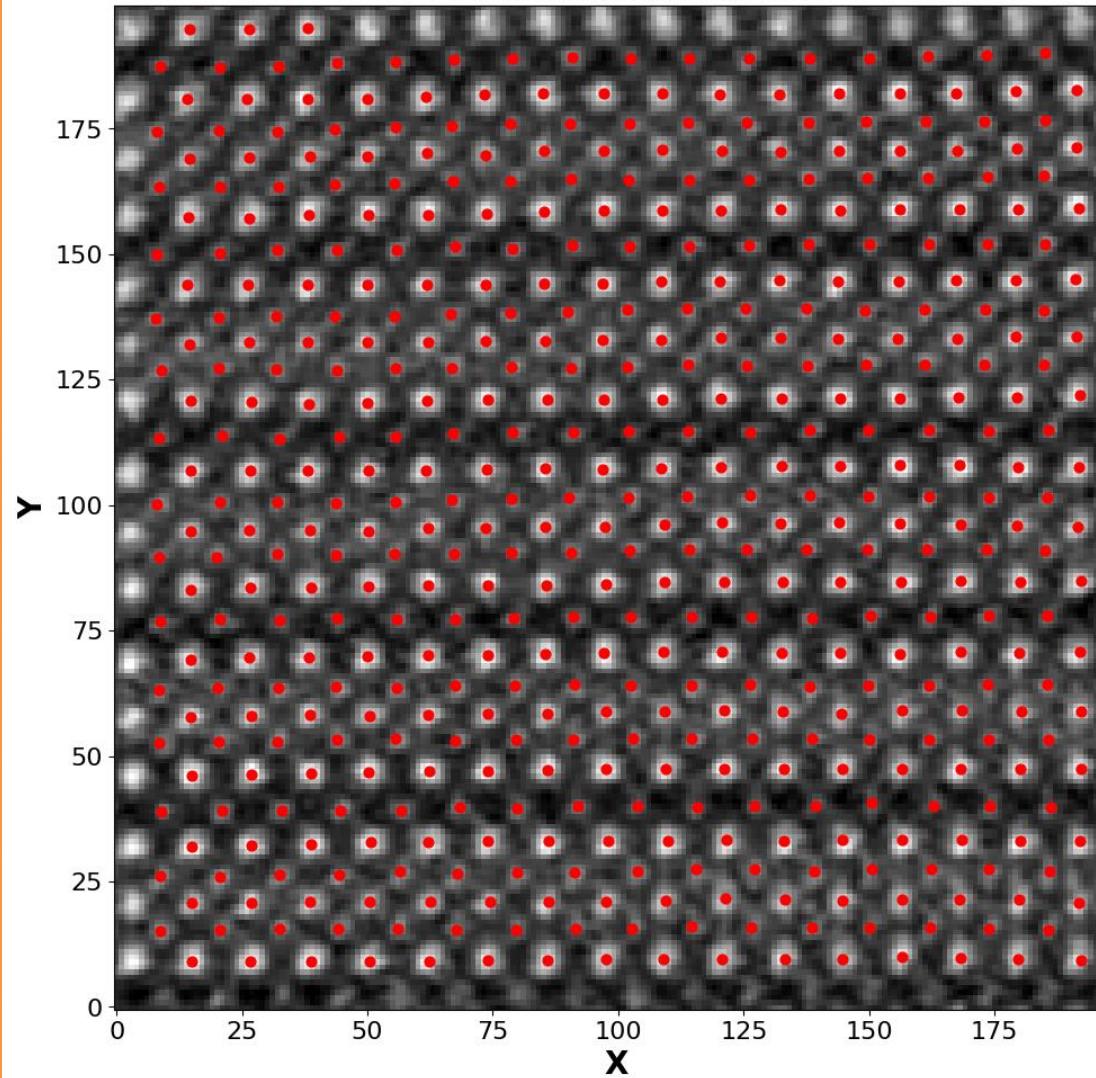
Both side of the reward should be minimized

Quality Function: Objective 1

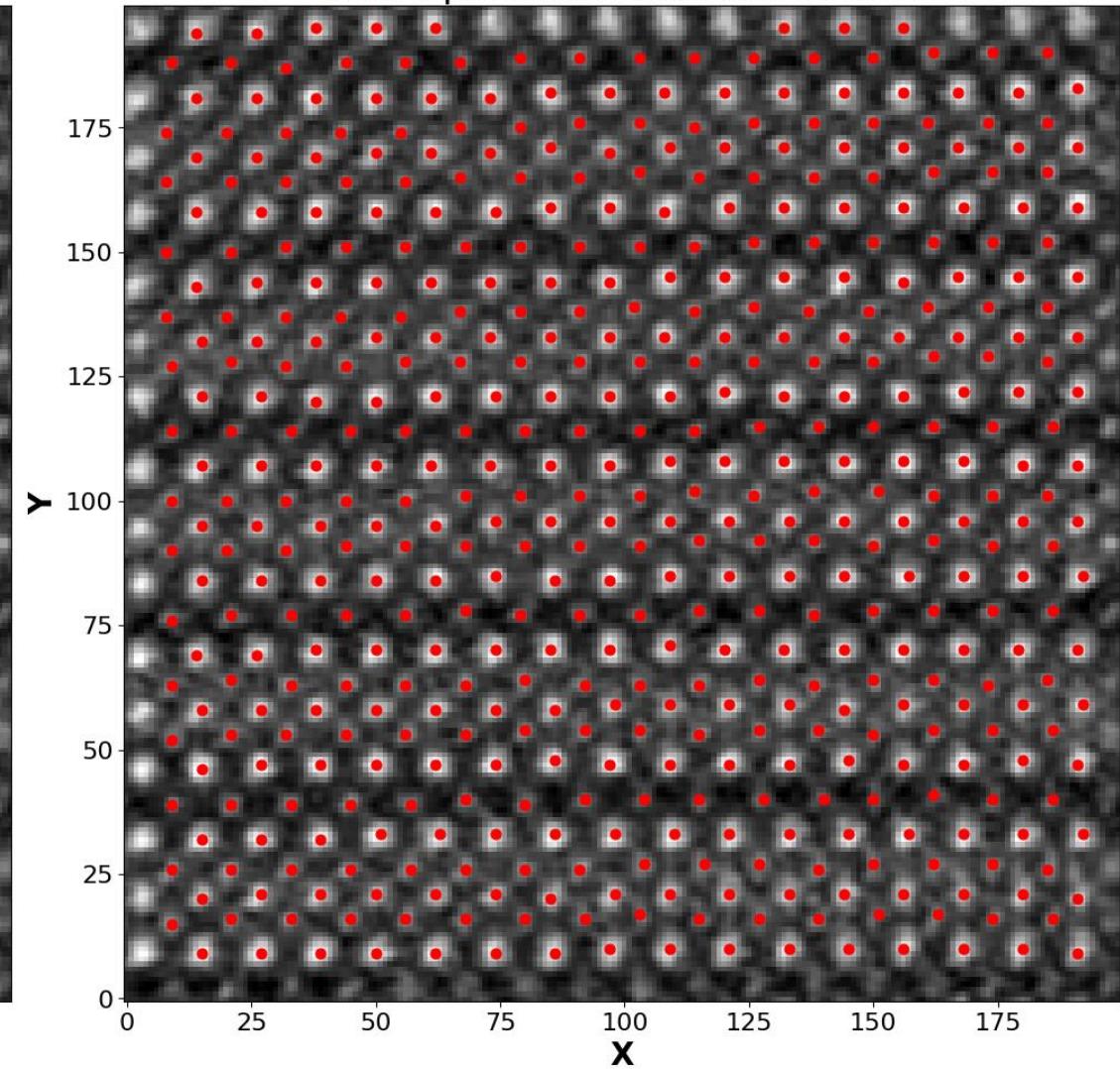
Error Function: Objective 2

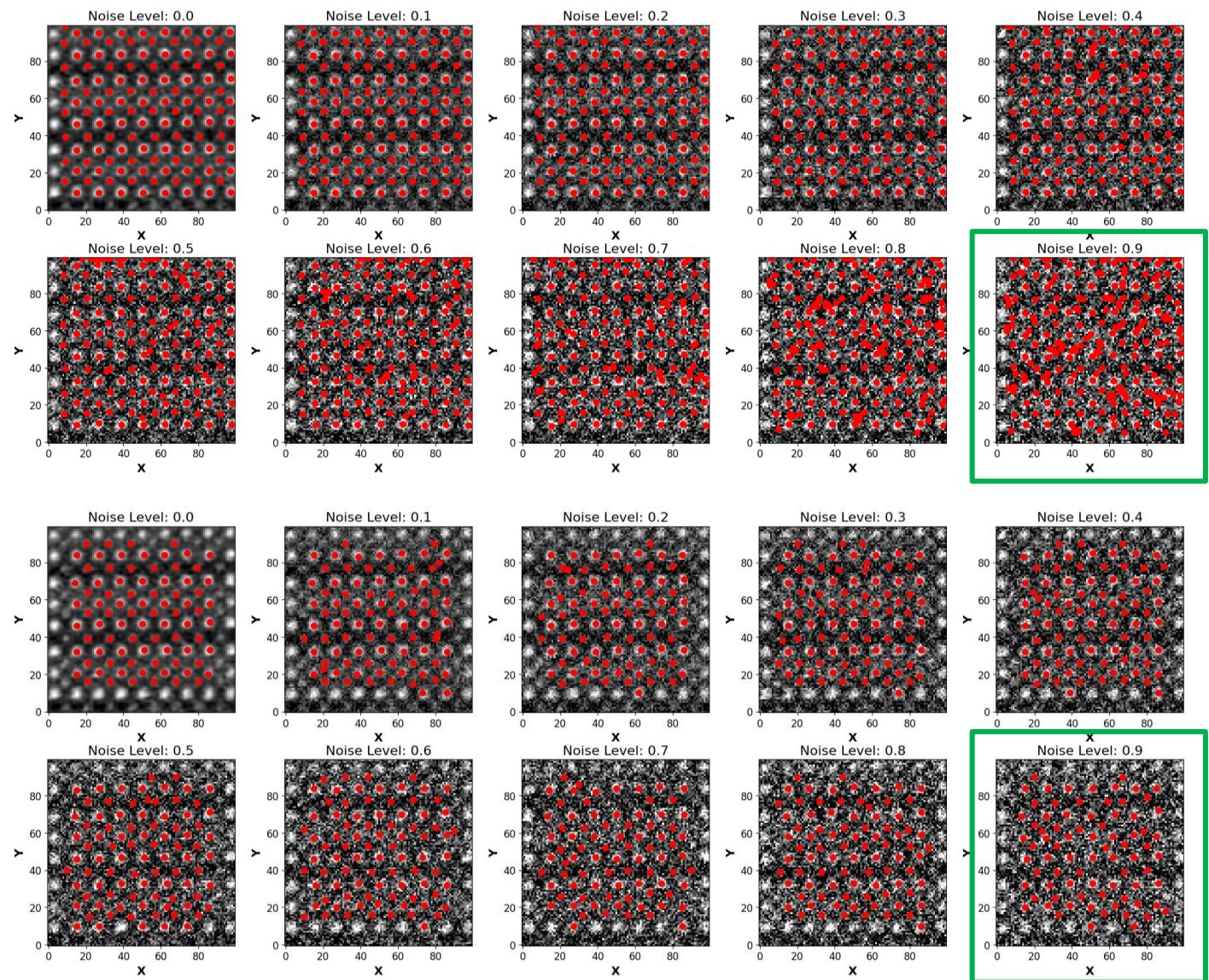
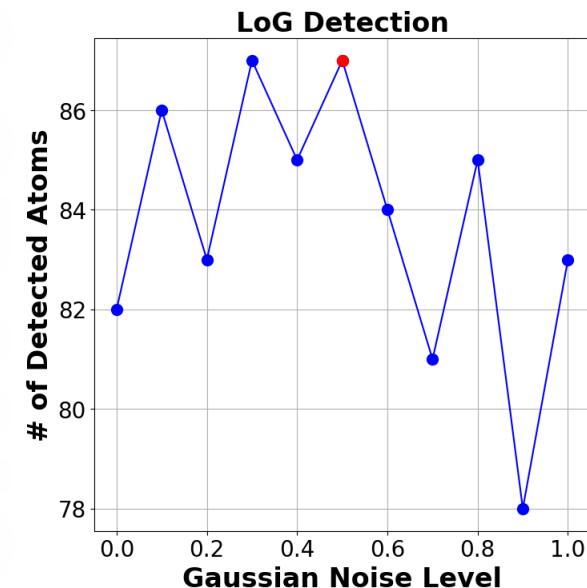
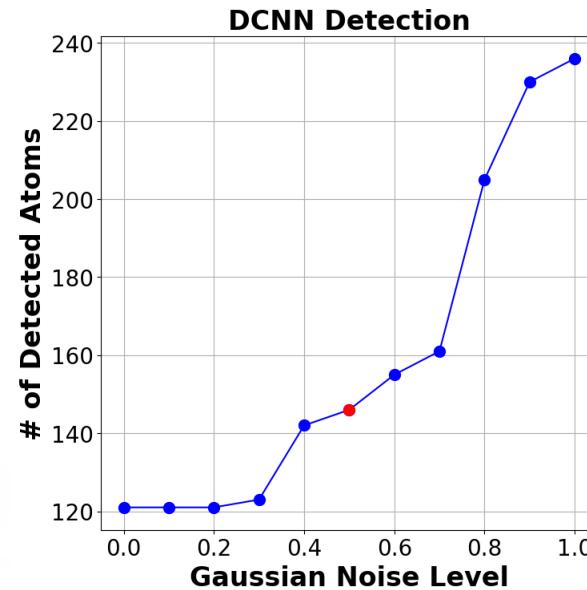


NN Detected Atoms



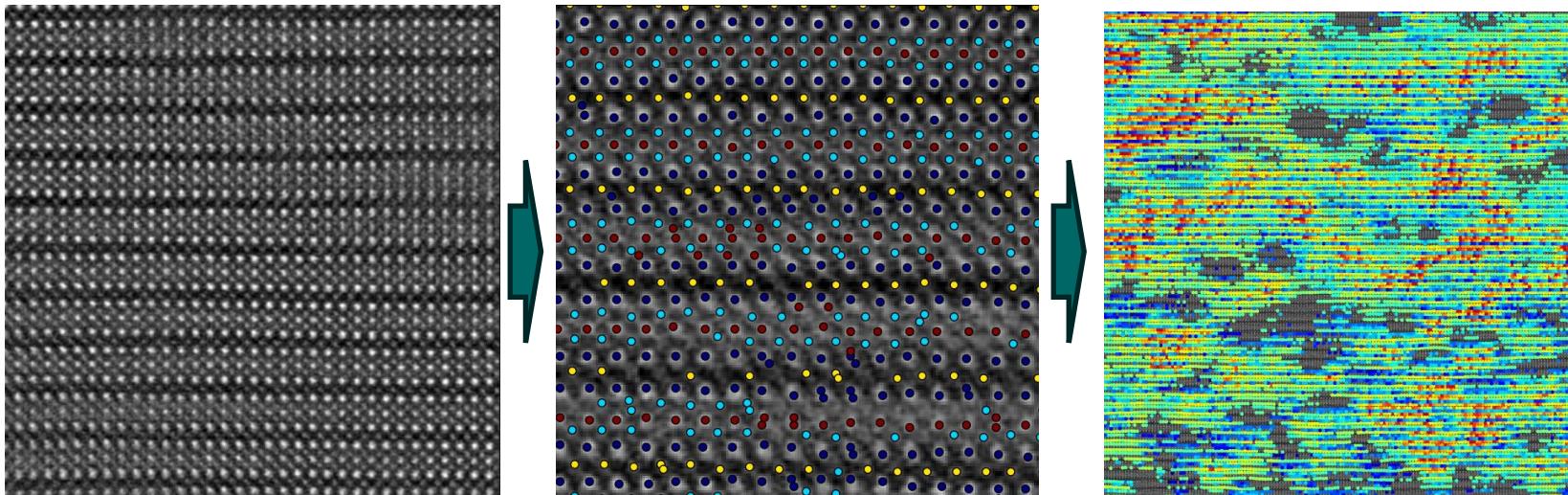
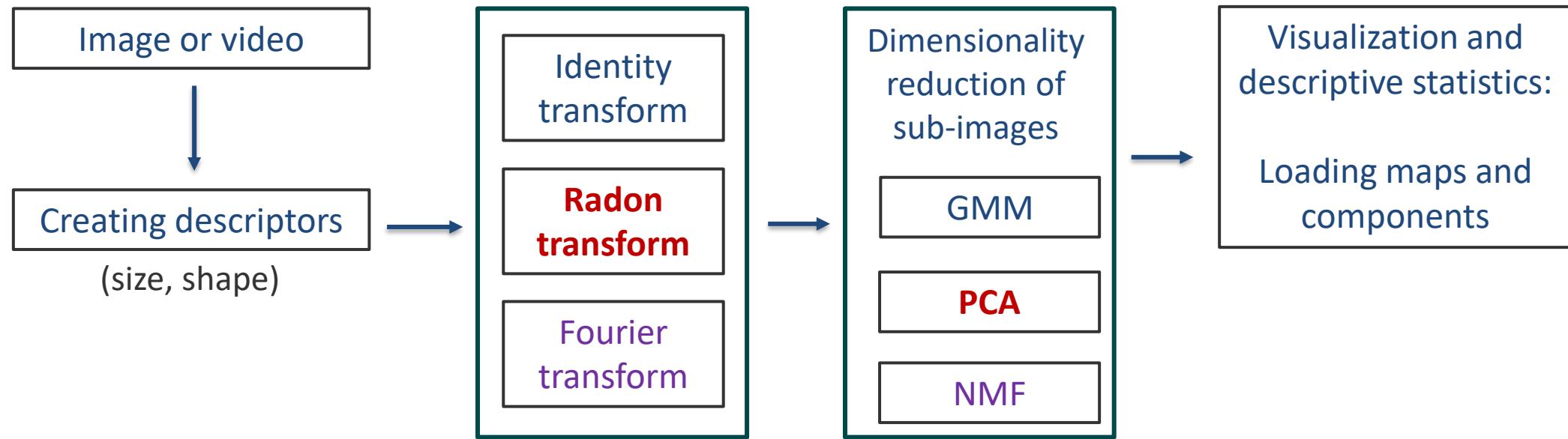
LoG-Optimized Detected Atoms





In contrast to the DCNN, the LoG detection exhibits a much lower variability in the number of detected atoms across noise levels, maintaining a relatively consistent count.

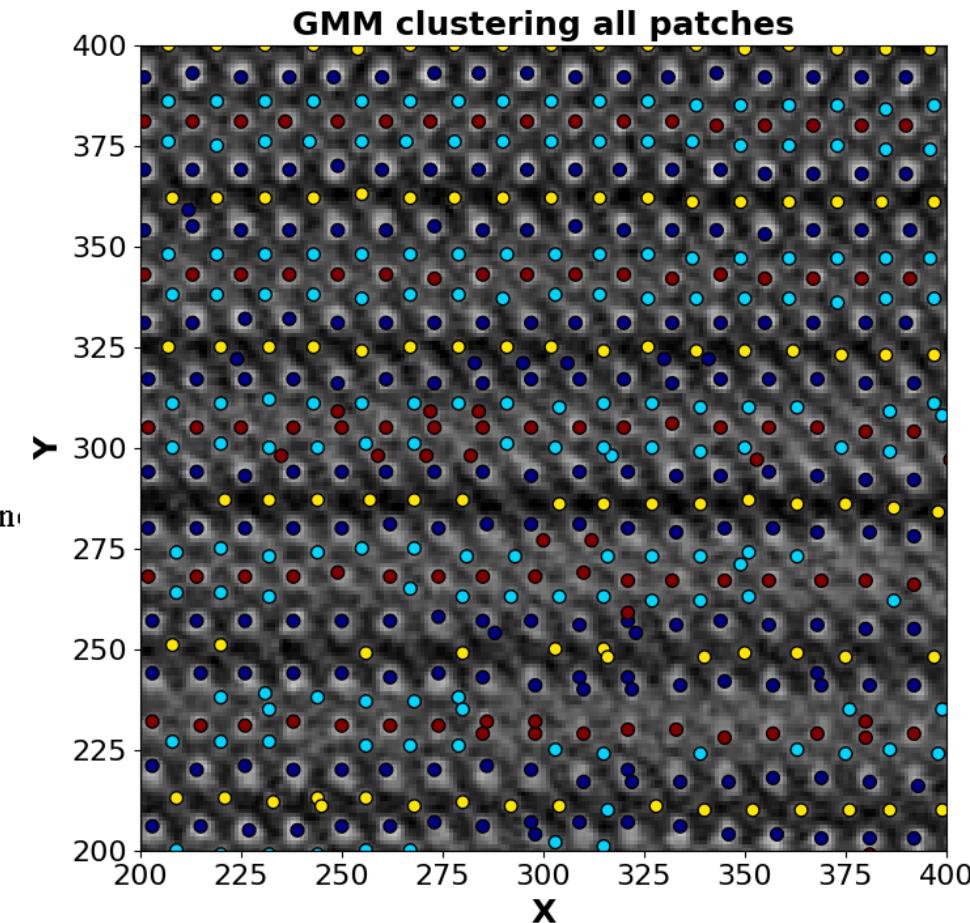
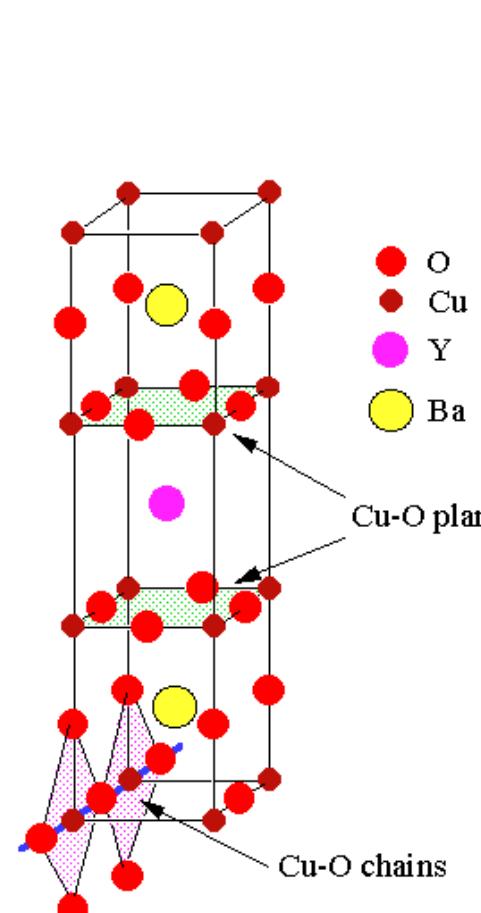
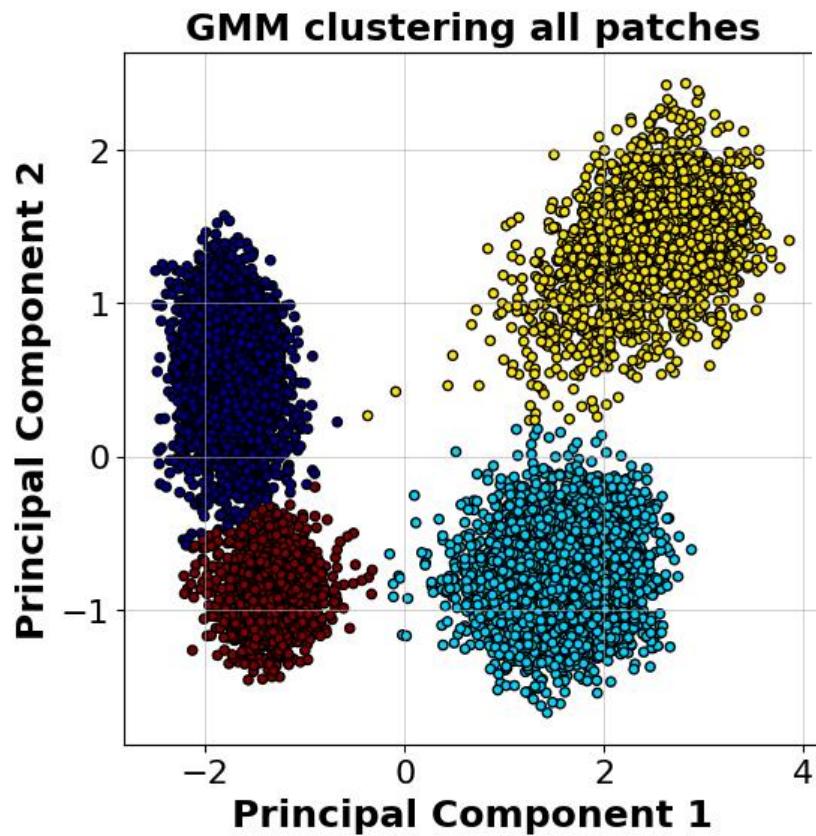
Example of analysis pipeline



Object discovery

- Large number of iterative steps
- Workflows with constant human oversight
- Non-myopic reward

1: Perform GMM clustering on all patches (LoG-optimized)

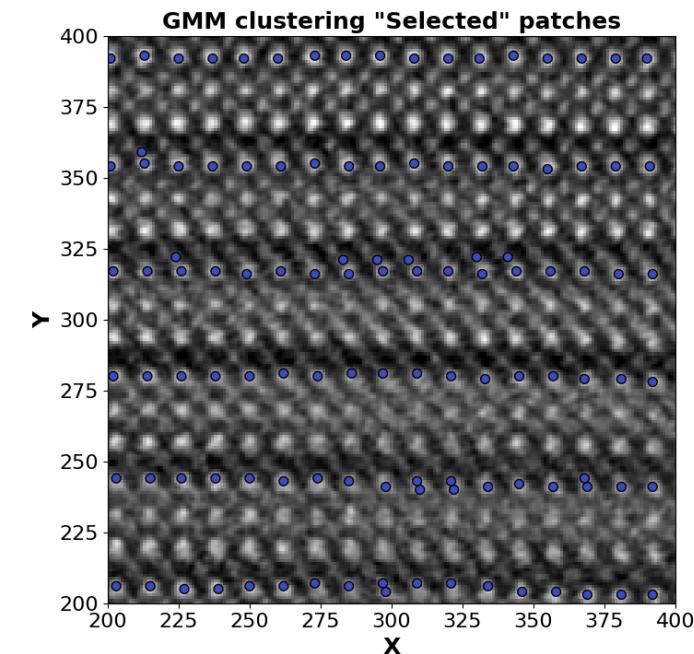
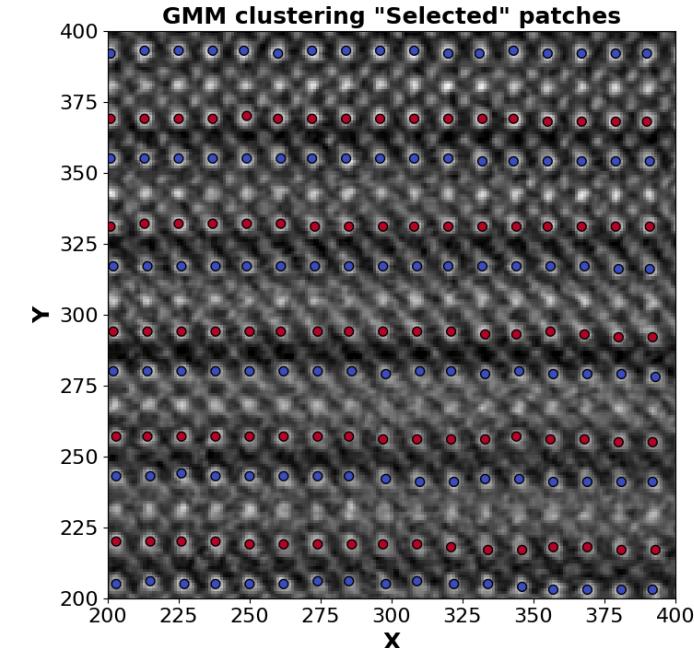
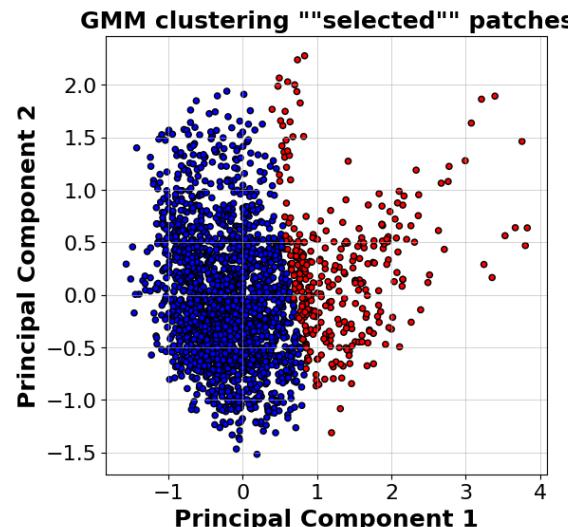
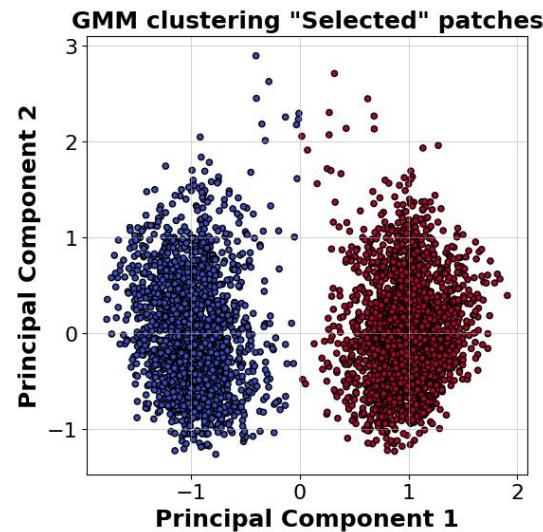


Well-defined clustered
scattered on the image
CuO plane : yellow
CuO chain : cyan
Ba: Dark blue
Y: Red

2: Perform GMM clustering
on patches (LoG-
optimized) **centered on all
Ba atoms**

3: Perform GMM clustering
on patches (LoG-
optimized) **centered on
one type of Ba atoms**

Here we are seeking the accurate values of
Threshold and covariance type to observe if
there is a variation in just one type of Ba?

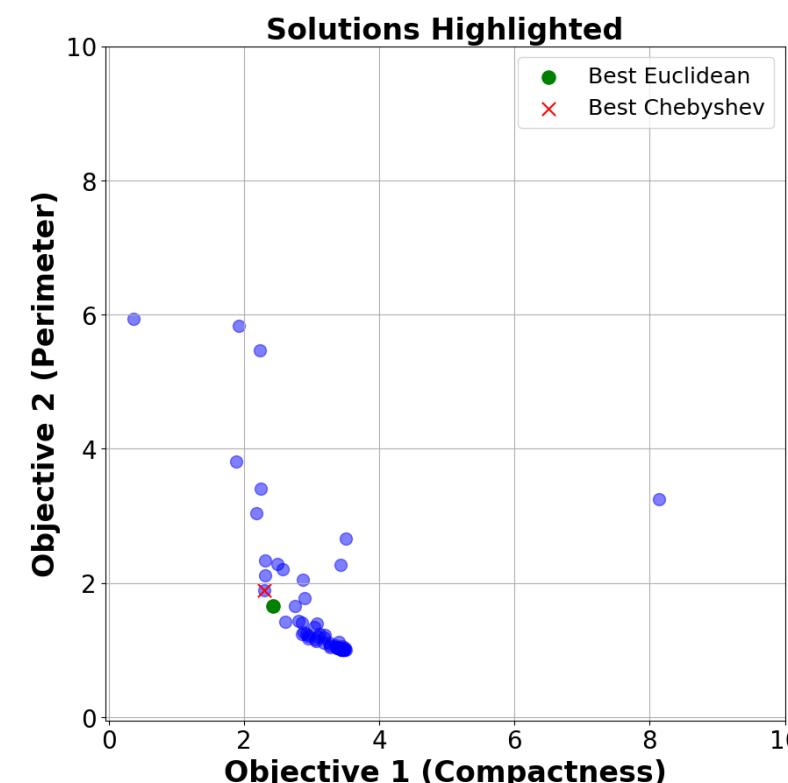


Threshold = (0.0, 1.0)

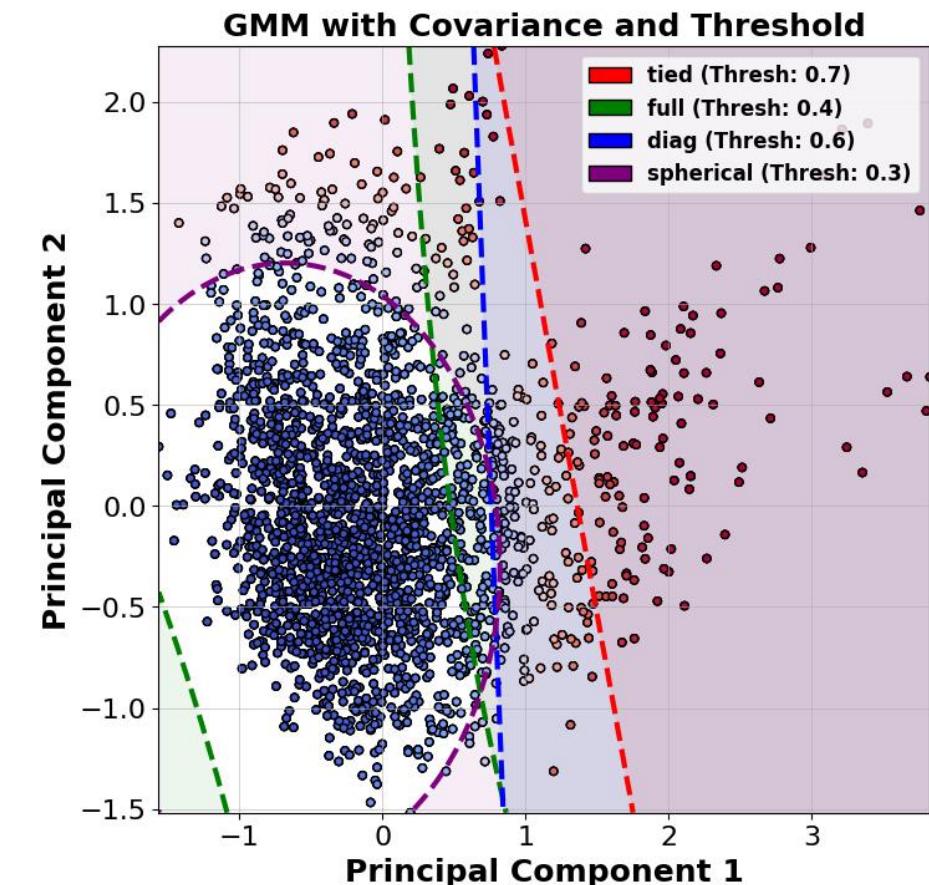
Covariance Type: tied, full, diag, spherical

Both side of the reward should be minimized

Objective 1:
Compactness of the amorphous area
Objective 2:
Perimeter of amorphous area



GMM clusters based on only one type of Ba atoms, introducing some variety, which it can be differentiated by different values of threshold and covariance type in GMM clustering

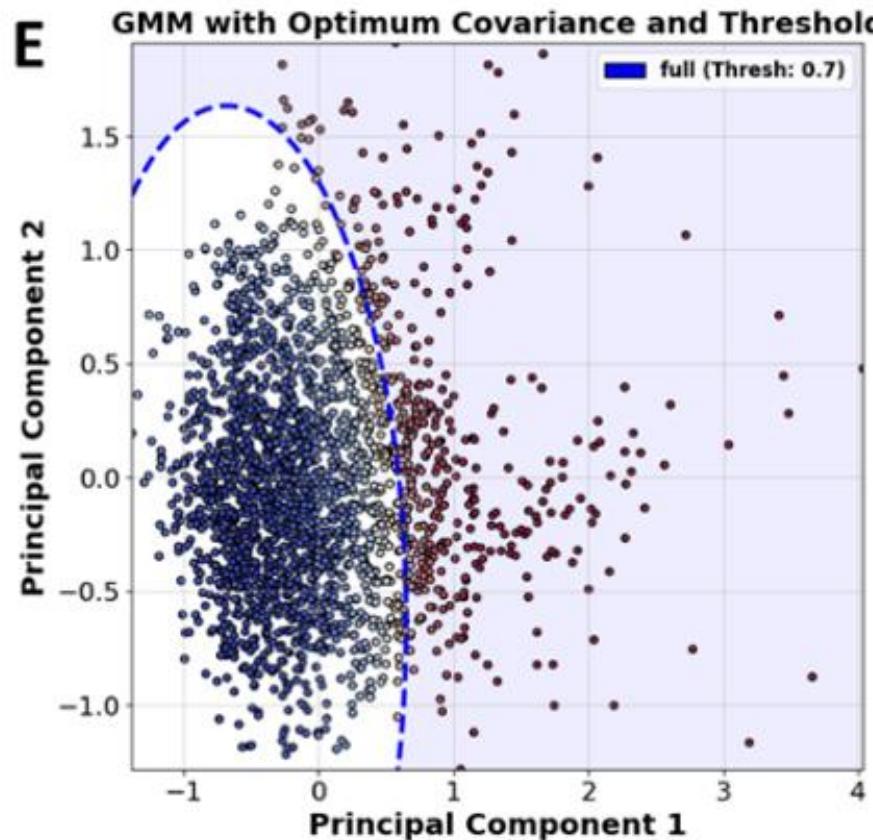


Objective 1:

Compactness of the amorphous area

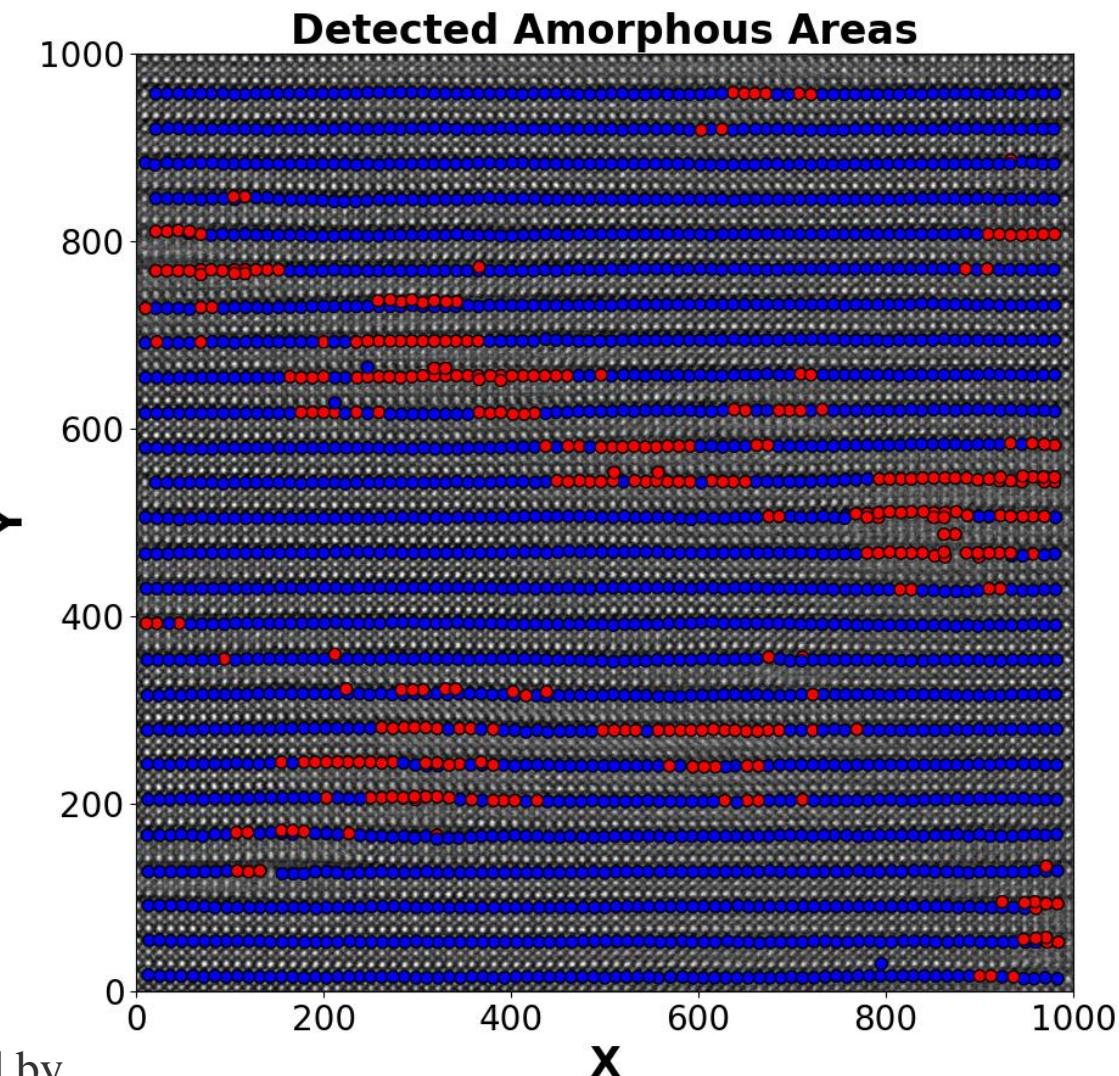
Objective 2:

Perimeter of amorphous area



Optimal threshold and covariance type achieved by
MOBO for GMM clustering,

1000*1000 pixels



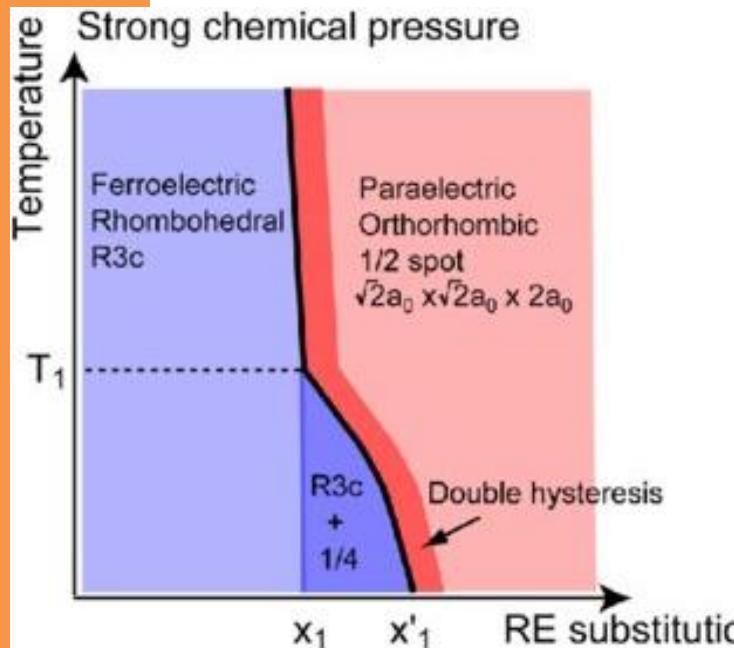
Uncovered amorphous areas in the substrate

So far, we set a single “fixed expensive model” for evaluation

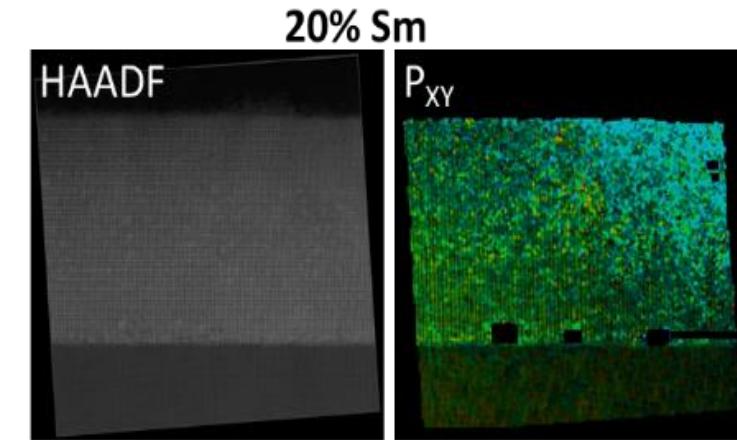
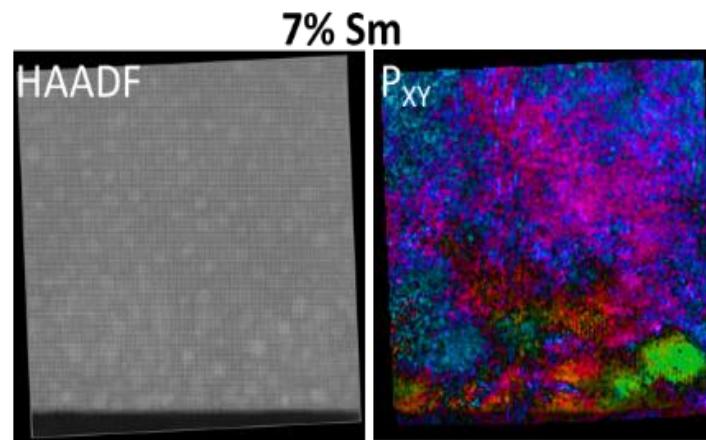
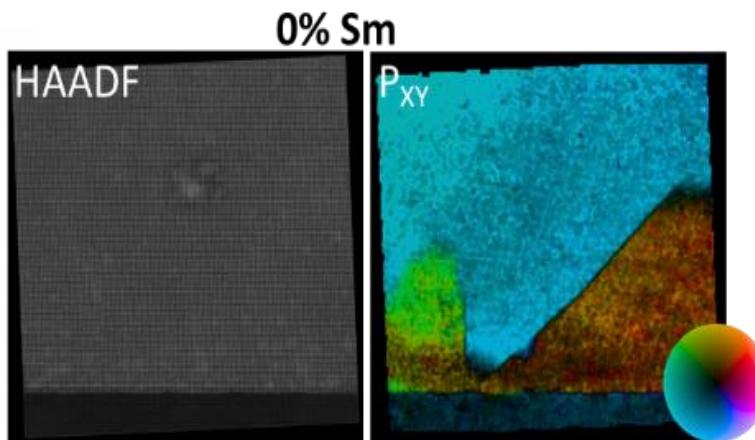
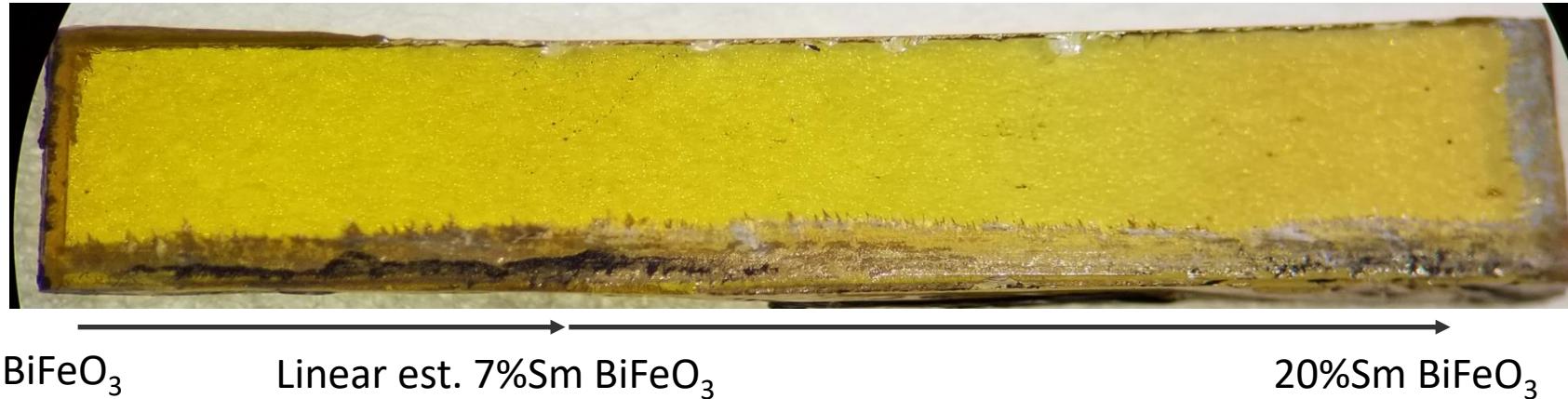
Sometimes, the “expensive model” can be too expensive for even BO... however, a cheaper proxy “lesser accurate” model can be available.

How can we utilize both ? Time for multi-fidelity !!

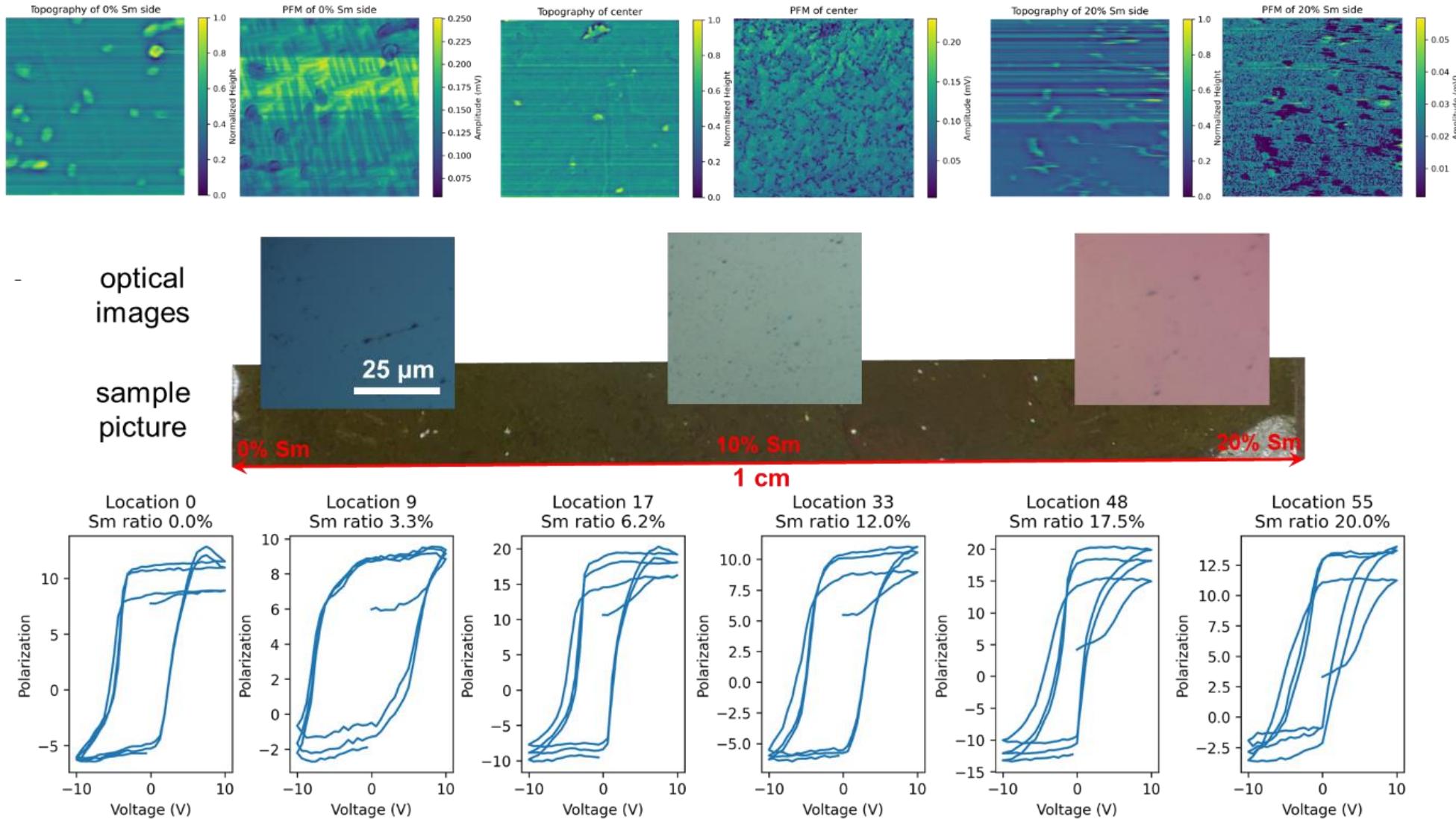
Combi Libraries: Expensive Measurements



Sample by I. Takeuchi, UMD
Phase diagram by N. Valanoor et al.

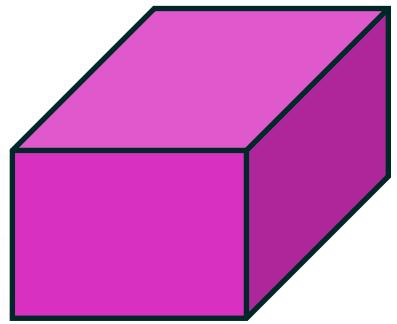


Combi Libraries: Cheap(er) Measurements

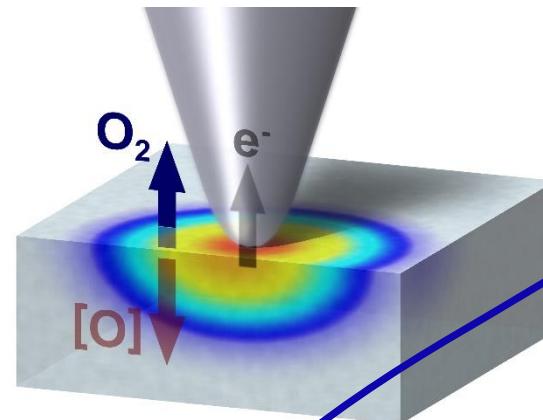


Multiple fidelity cycles are possible

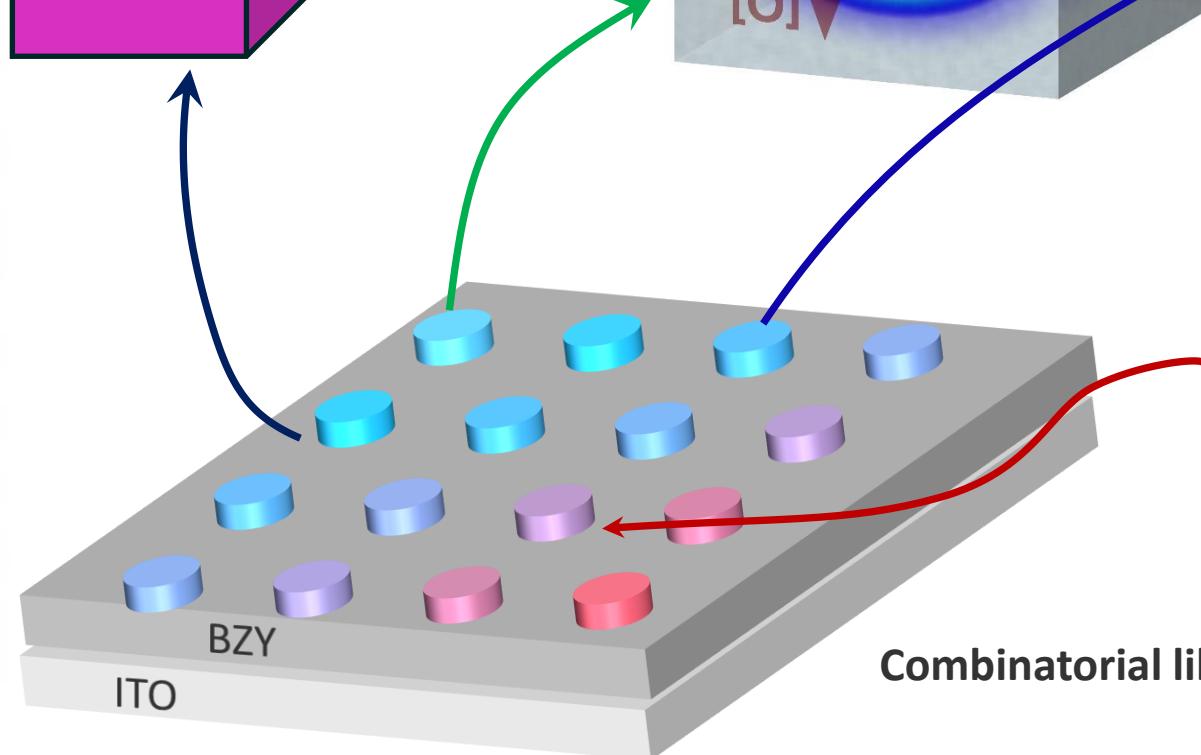
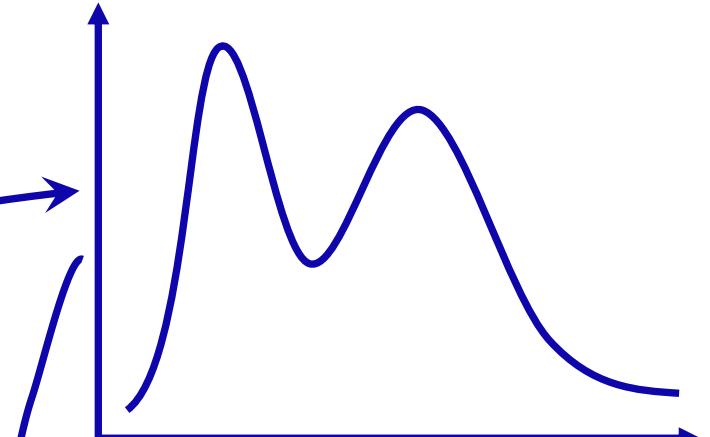
High-fidelity measurements:
IS on bulk ceramics



Low-fidelity measurements:
ESM on combinatorial library



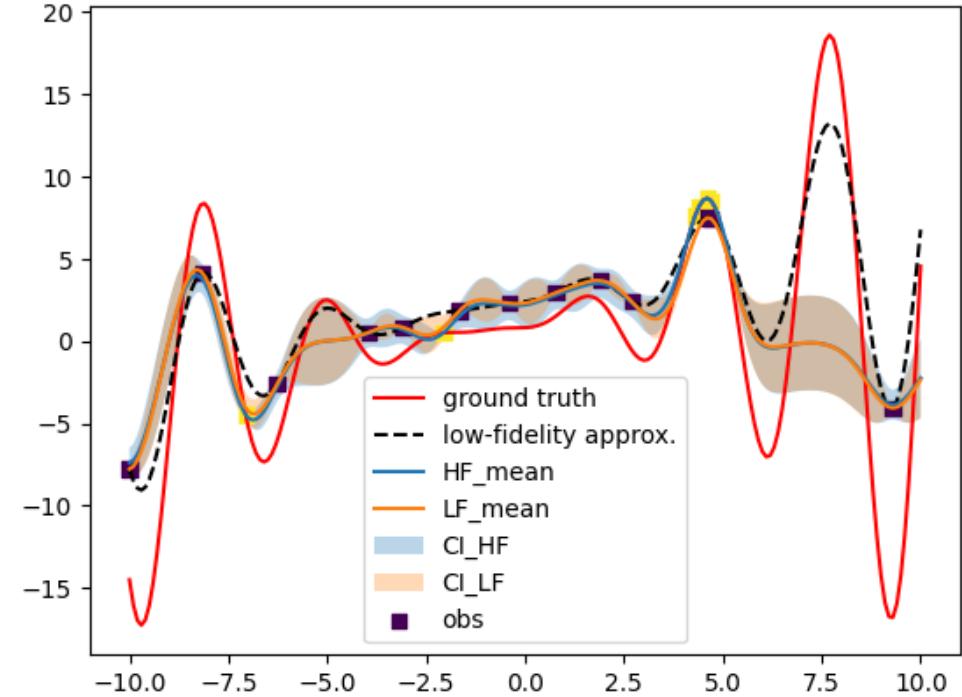
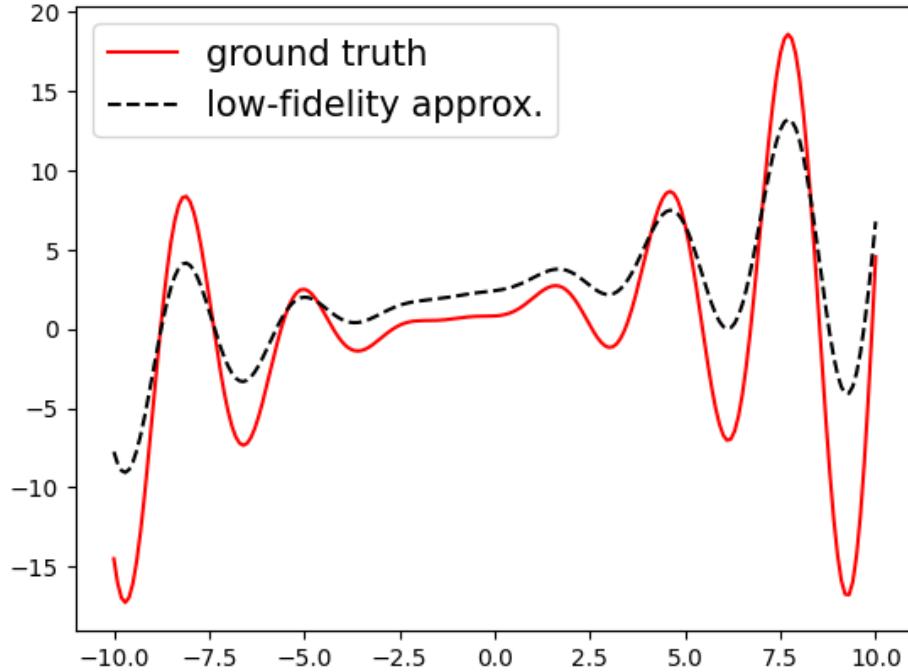
Medium-fidelity measurements:
Distribution of relaxation times on caps



Combinatorial library

Bayesian optimization

Multifidelity Gaussian Processes



$$y(x) = f^T \beta + z(x)$$

Polynomial Regression

$$z(x) \sim GP(0, \sigma^2 R(x^i, x^j) * K_I(f^i, f^j))$$

$$R(x^i, x^j) = \exp\left(-\sum_{k=1}^d \theta_k (x_k^i - x_k^j)^2\right)$$

Fidelity kernel

$$K_I(f^i, f^j) = \exp(-\delta |f^i - f^j|)$$

$$\delta > 0$$

hyperparameter

Multifidelity acquisition functions

$$\max_{X,f} U(f(X,f)|MFGP)$$

Acquisition value of x, given HF

$$\Delta EI_h(x^*) = EI(x^*) - 0 = EI(x^*)$$

Same as EI for standard BO

$$a_h = U(f(X|f=2|MFGP))$$

Acquisition value of x, given LF

$$\Delta EI_l(x^* | Y_l(x_l^*)) = EI(x^*) - EI(x^* | Y_l(x_l^*))$$

Further improvement with LF samples

$$l_h = |a_h - U(f(X|f=1|MFGP))|$$

Multi-fidelity acquisition function

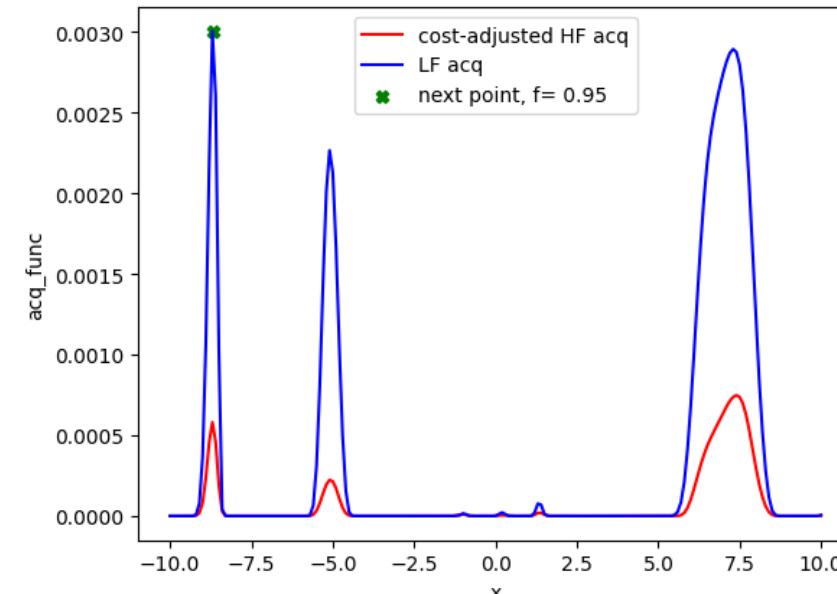
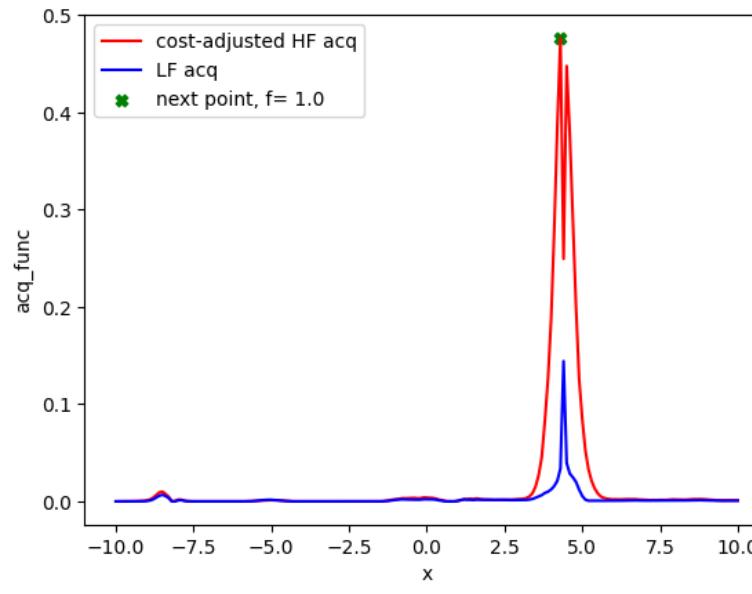
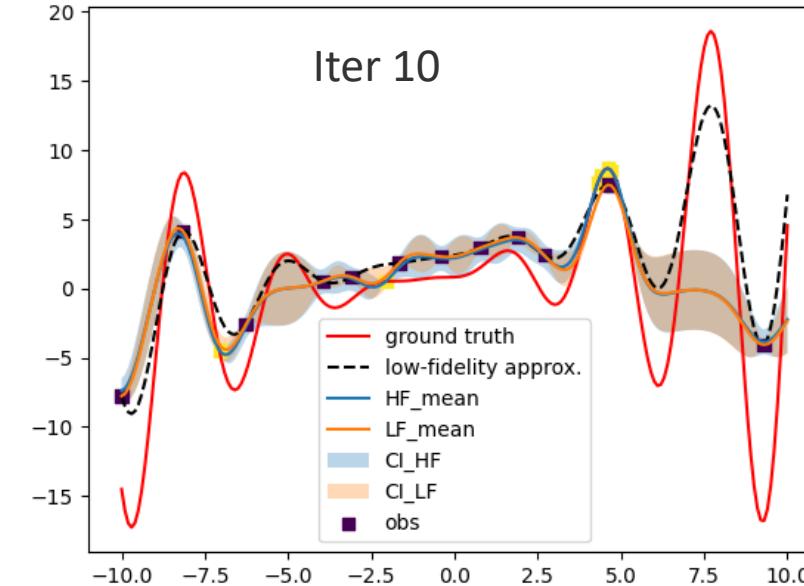
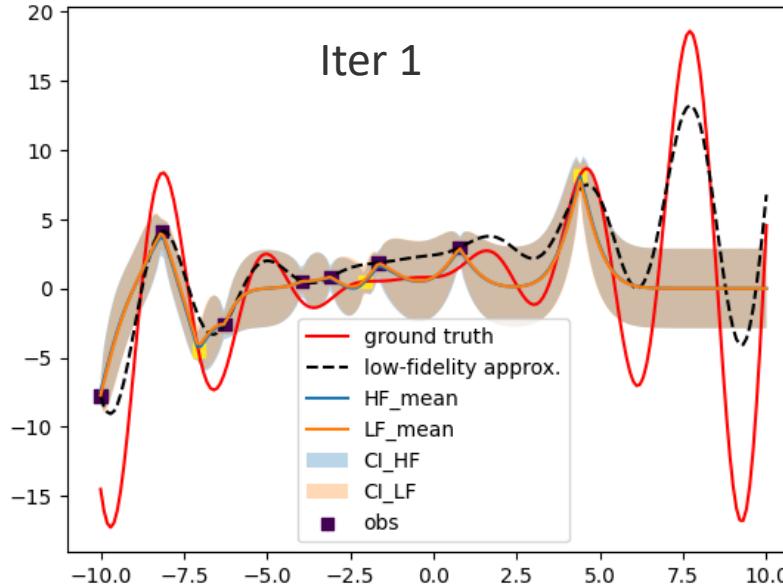
$$U(f(X,f)|MFGP) = \begin{cases} \frac{a_h}{C} & \text{if } f = 2 \\ l_h & \text{if } f = 1 \end{cases}$$

Cost-ratio: Can be derived from model complexity and domain knowledge

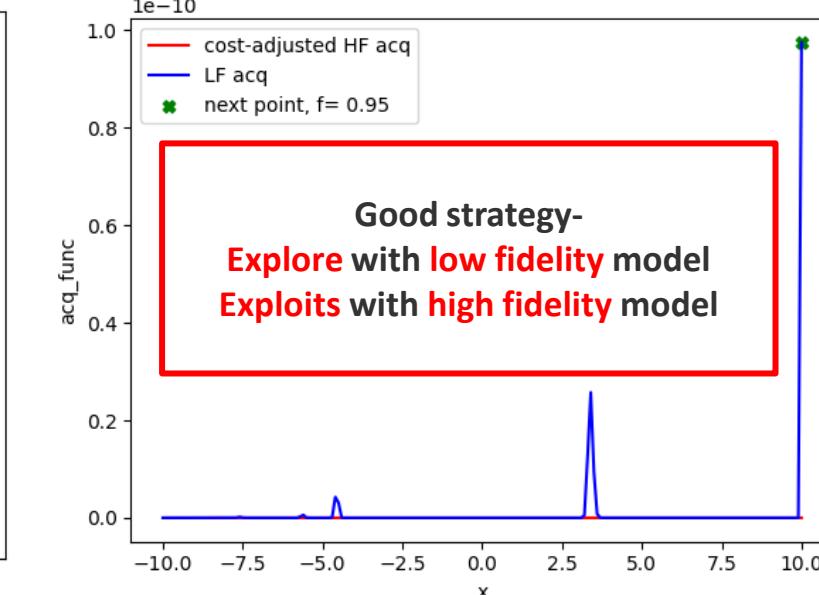
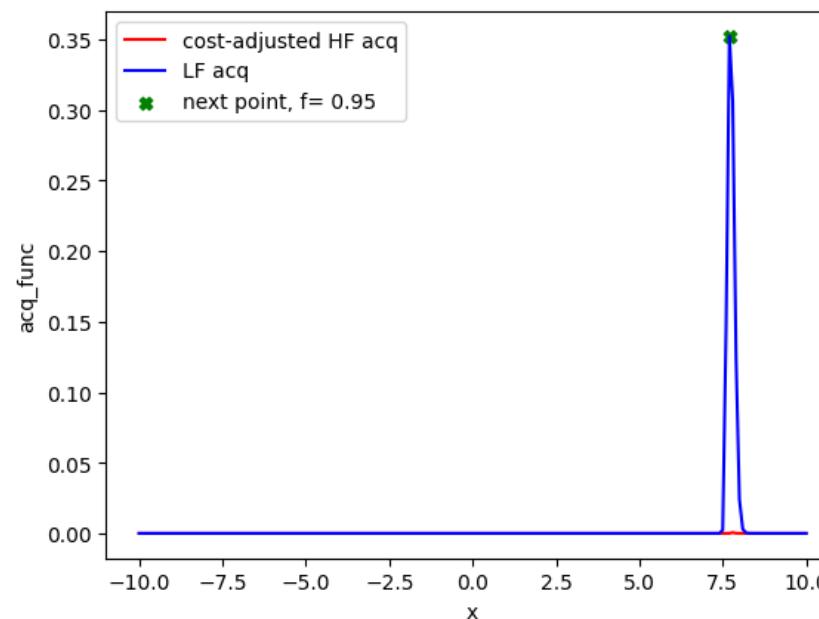
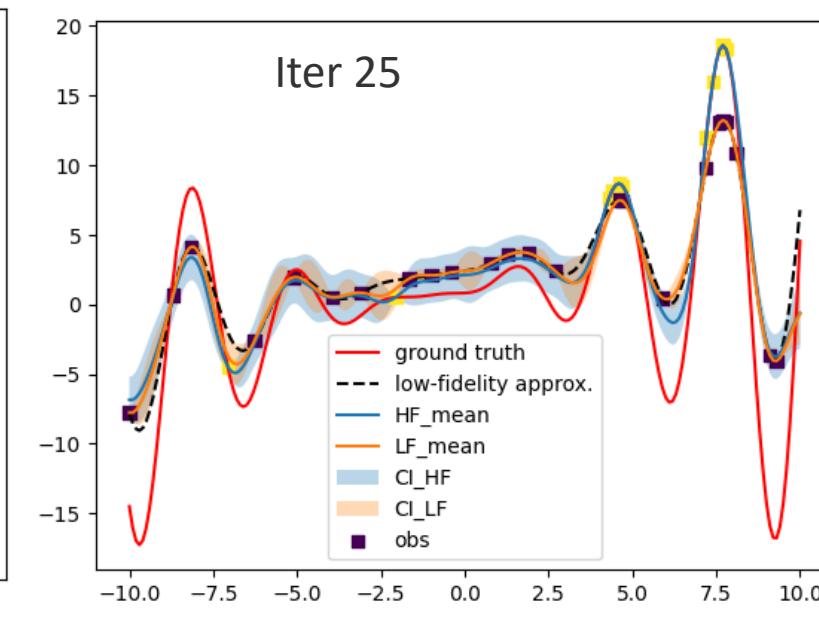
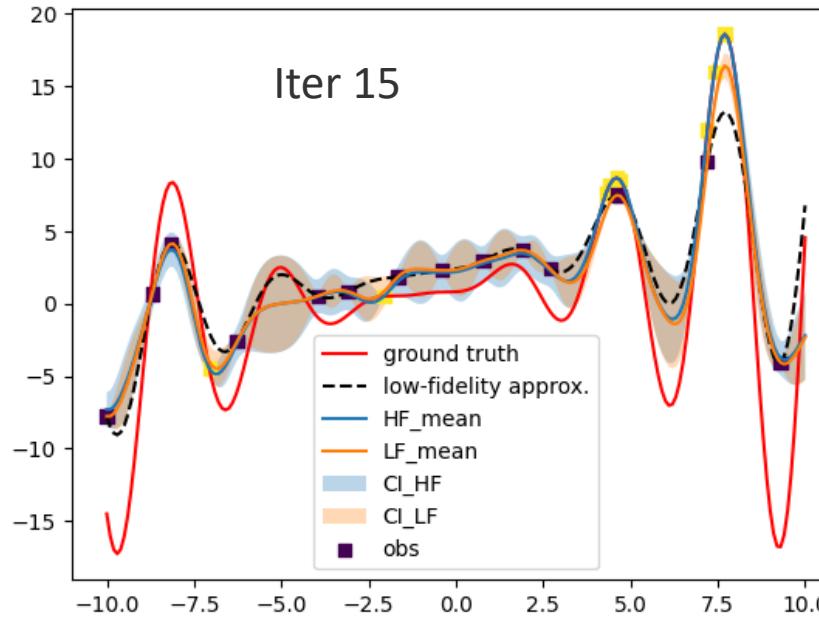
Similar trend

Shu, L., Jiang, P. & Wang, Y. A multi-fidelity Bayesian optimization approach based on the expected further improvement. *Struct Multidisc Optim* **63**, 1709–1719 (2021).

Multifidelity GP



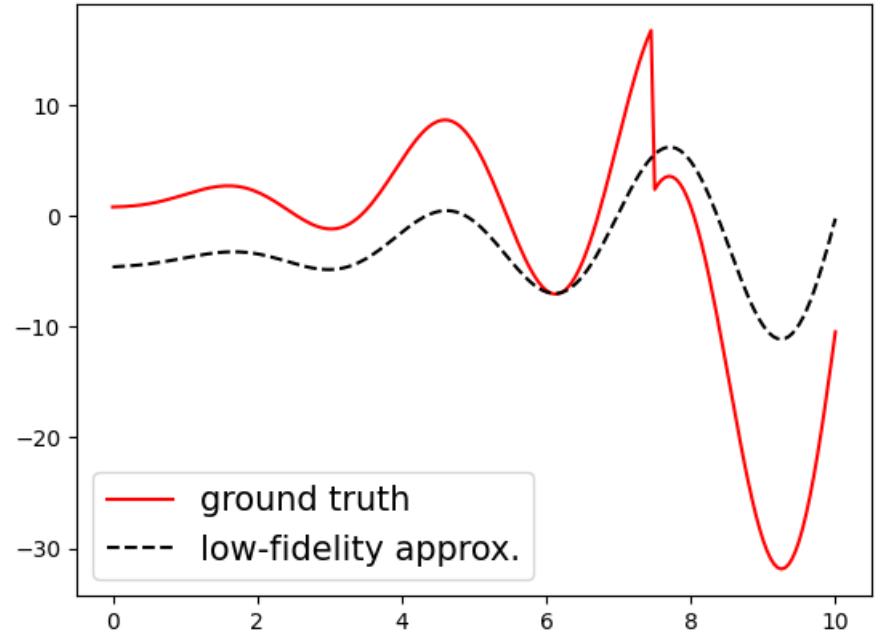
Multifidelity GP



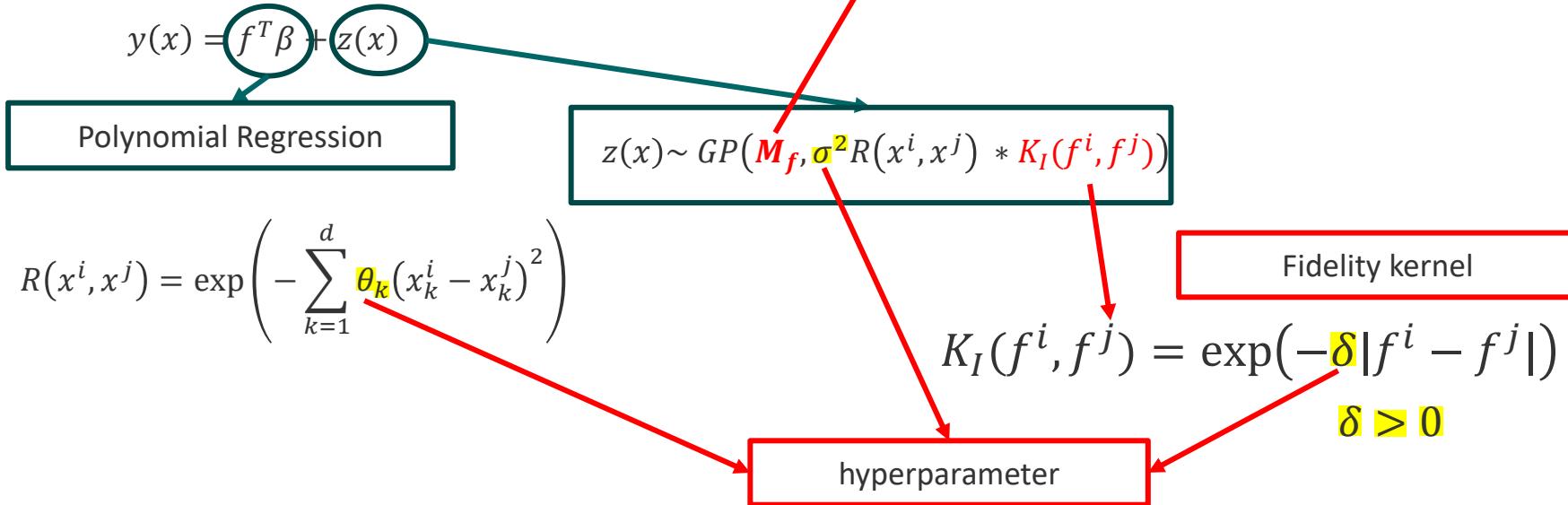
Potential applications:

- 1. 2D Ising Model:** Low-fidelity (20x20) → High-fidelity (60x60)
- 2. Hybrid perovskites:** Low-fidelity (cheap measurements) → High-fidelity (expensive measurements)
- 3. Material Synthesis:** Low-fidelity (combi library) → High-fidelity (PLD synthesis)
4. and many more ...

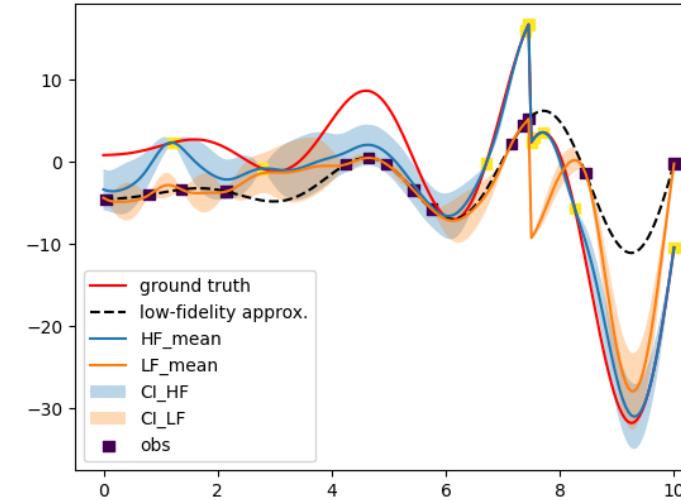
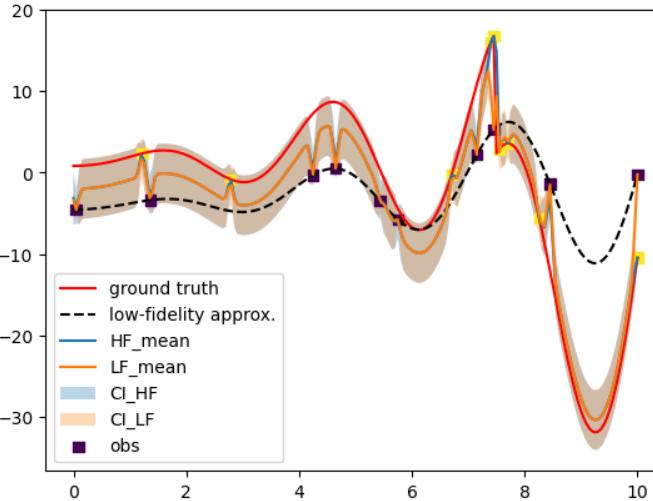
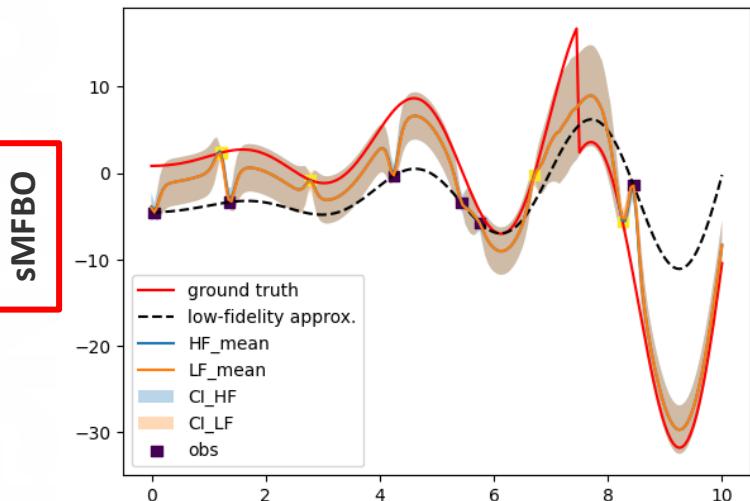
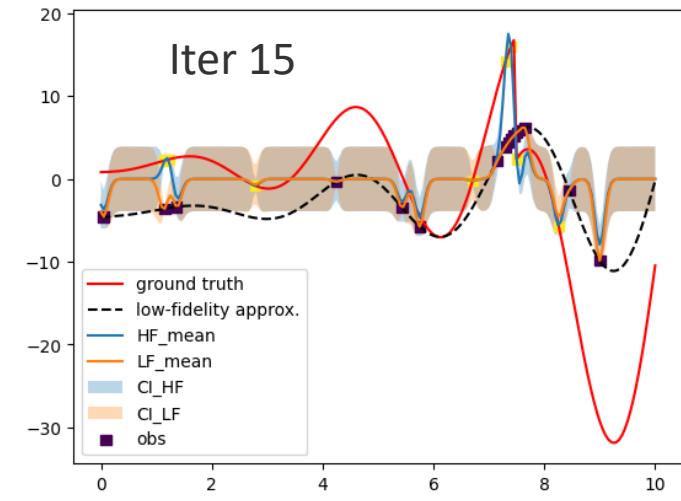
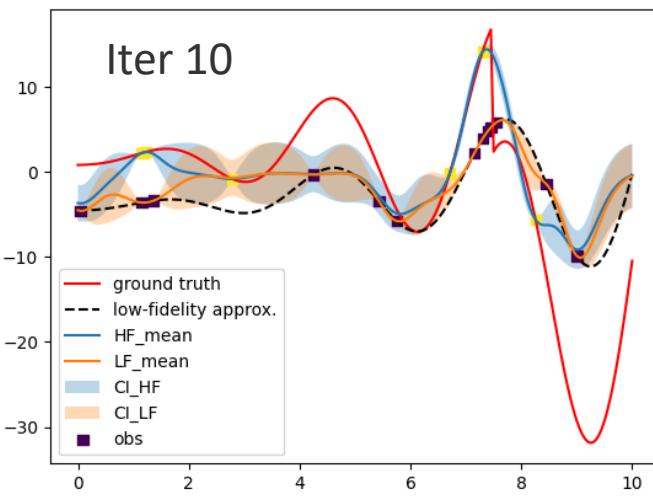
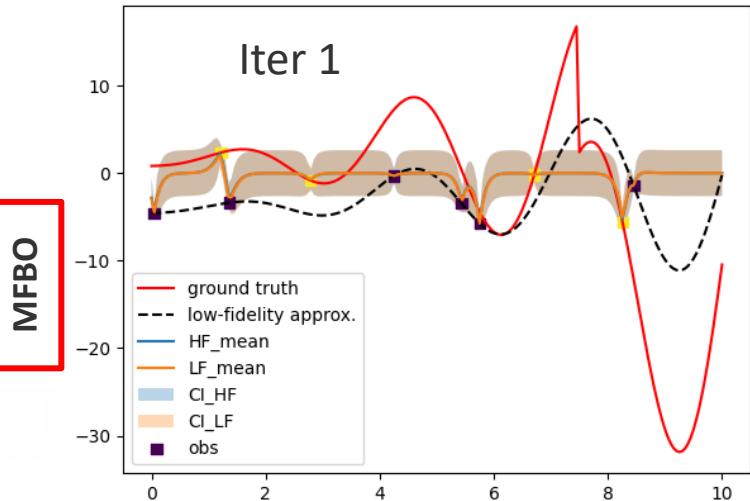
Structured Multifidelity GP



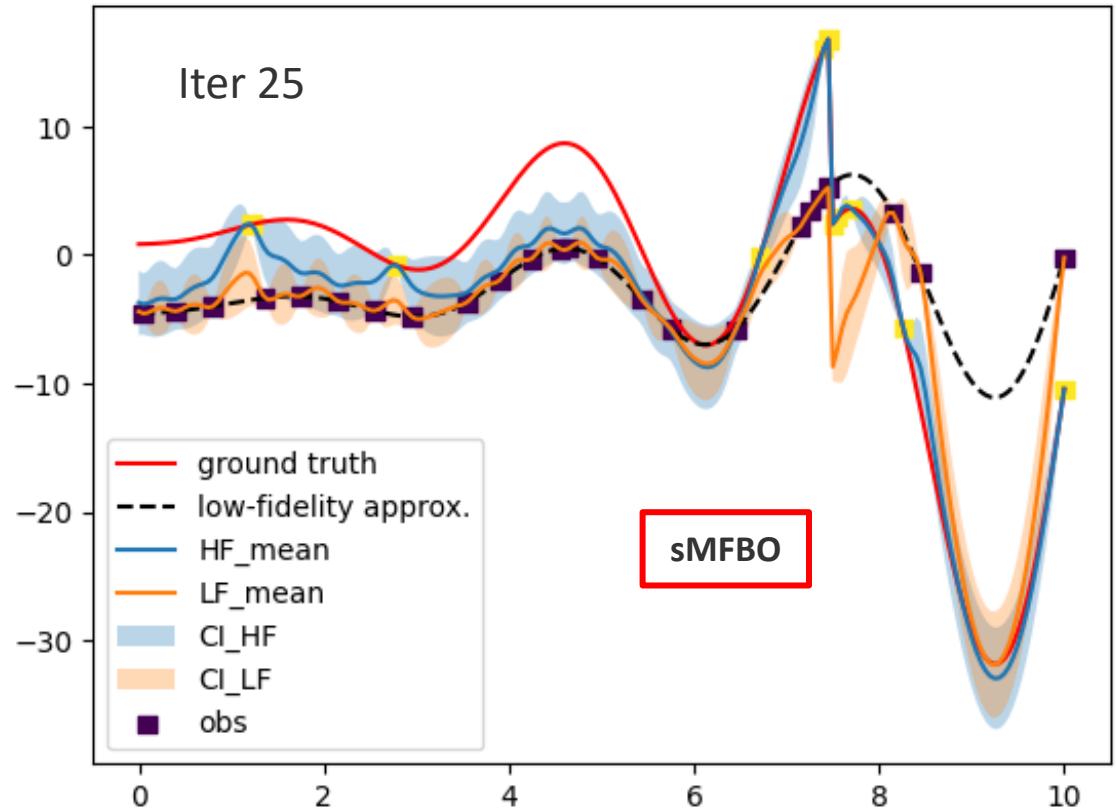
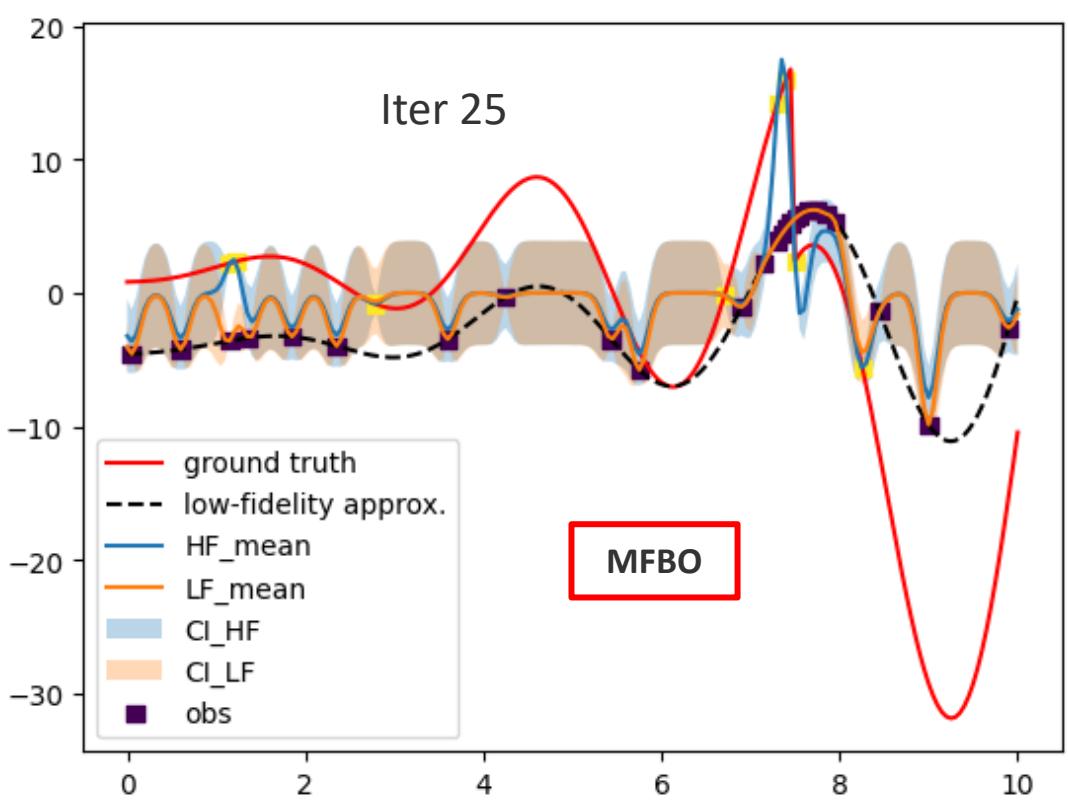
Mean function

$$M_f(x, a, b, c) = \begin{cases} f_1(x, a), & x < c \\ f_2(x, b), & x \geq c \end{cases}$$
$$a = \text{Unif}[5, 10], b = N(0, 1), c = N(15, 2)$$


Structured Multifidelity GP



Structured Multifidelity GP



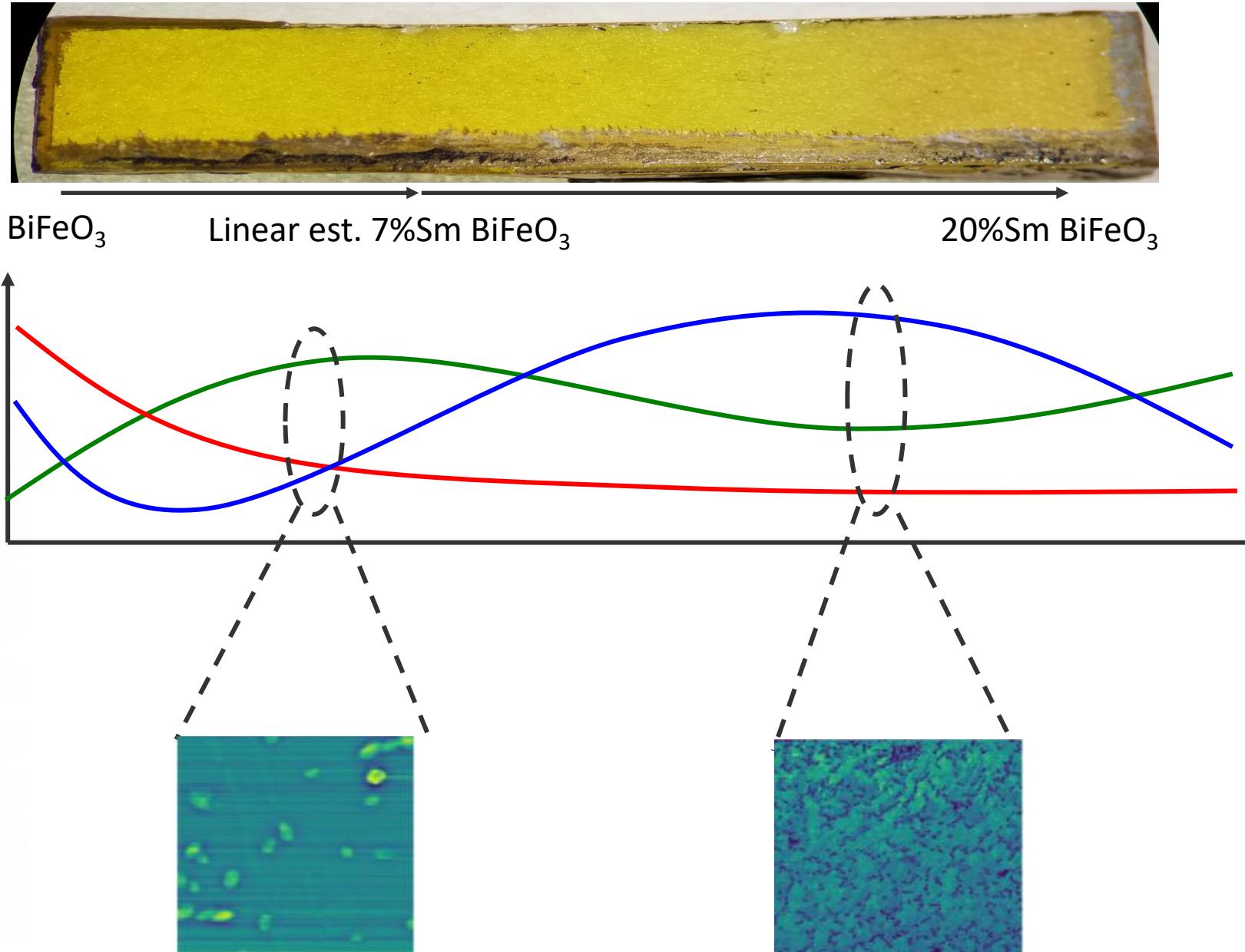
1. Multifidelity **structured** GP:

- We have the easy to evaluate function with probabilistic model and expensive to evaluate function
- The easy function is a proxy for expensive one and has some correlative relationship to it
- We create policy that balances evaluation costs

2. Multitask GP:

- We have multiple observables in different spaces
- And common **latent model** that emits them
- Can find minima in the expensive space suggested by cheap(er) function

General combinatorial library exploration



Compositional library
(can be 1D or 2D,
encoding from binary to
quaternary diagrams).
Composition $c(x)$ or $c(x,y)$
is assumed to be known

Non-observable latent
variables that represent
materials functionality. These
form GP or sGP as a function
of concentration (and via it,
space)

The latent variables emit the
observational data in the form
of images or spectra (via
second GP or decoder)

Experimental Instantiations

1. **Scenario I:** Data in full (microRaman across the combinatorial library, or grid measurement of topography or domains by PFM).
 - a. Can use the simple VAE or GMM to find latents (or even PCA)
 - b. However, VAE or GMM will not capture the spatial effects in $sGP(c(x,y))$
2. **Scenario II:** Active learning with one high dimensional imaging/spectroscopy method.
 - a. Normal GP/sGP/HL if measured property is scalar (if we have good scalarizer for image/spectra)
 - b. If it is active learning on images/spectra we **do not have way to do it.**
3. **Scenario III:** Active learning if we have full low dimensionality proxy data and active learning for low dimensional data. This is multifidelity GP and sGP
4. **Scenario IV:** Active learning when we have full high dimensional proxy data and use active learning for another high dimensional data (use Raman results to select places for STEM or PFM)
5. **Scenario V:** Co-navigation between 2 active high-dimensional data sets (meaning that measurements that emit from latents are different).