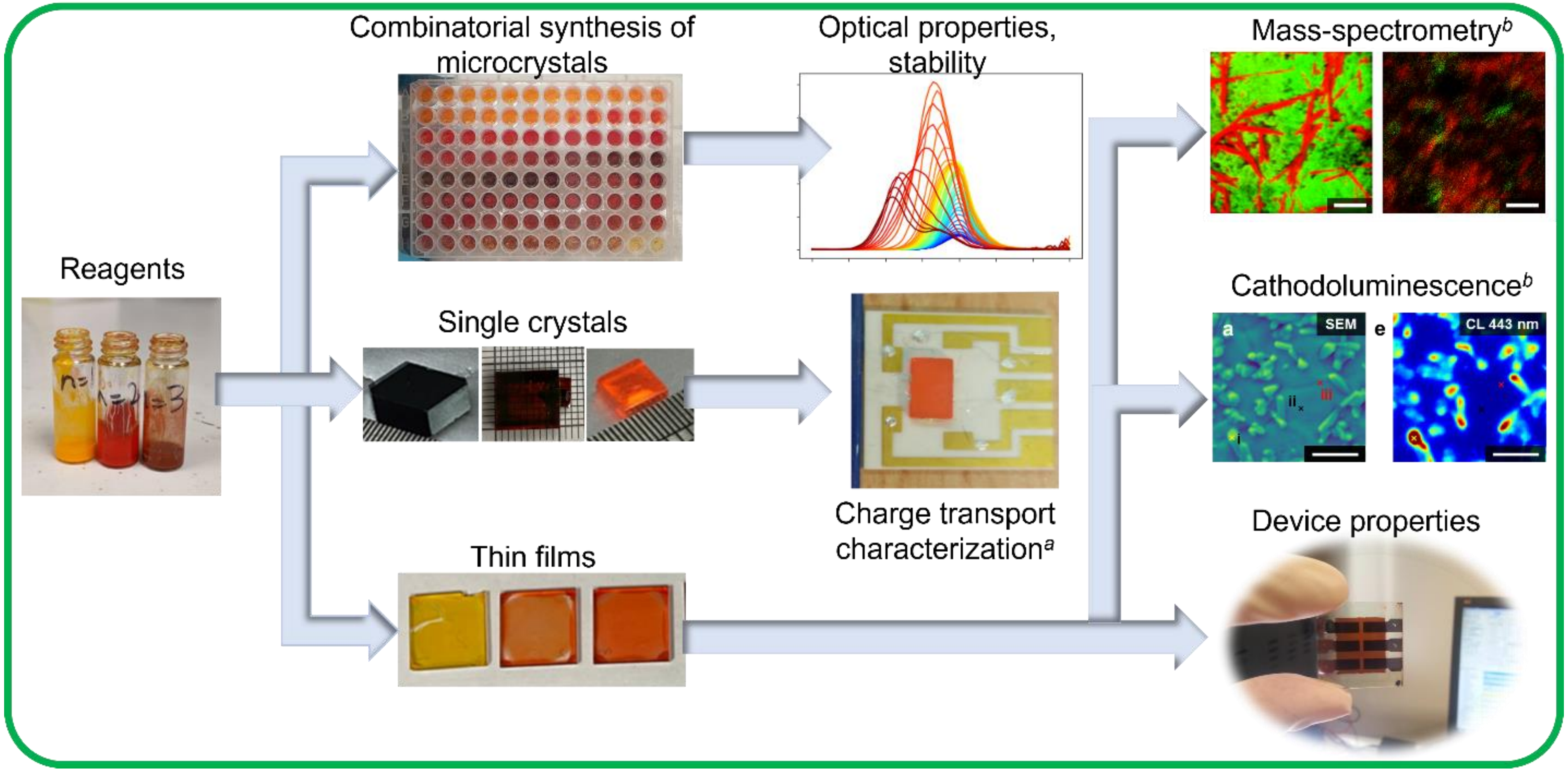
THE UNIVERSITY of TENNESSEE KNOXVILLE

Multifidelity and Multiobjective Bayesian Optimization

Sergei V. Kalinin

What is A Workflow?

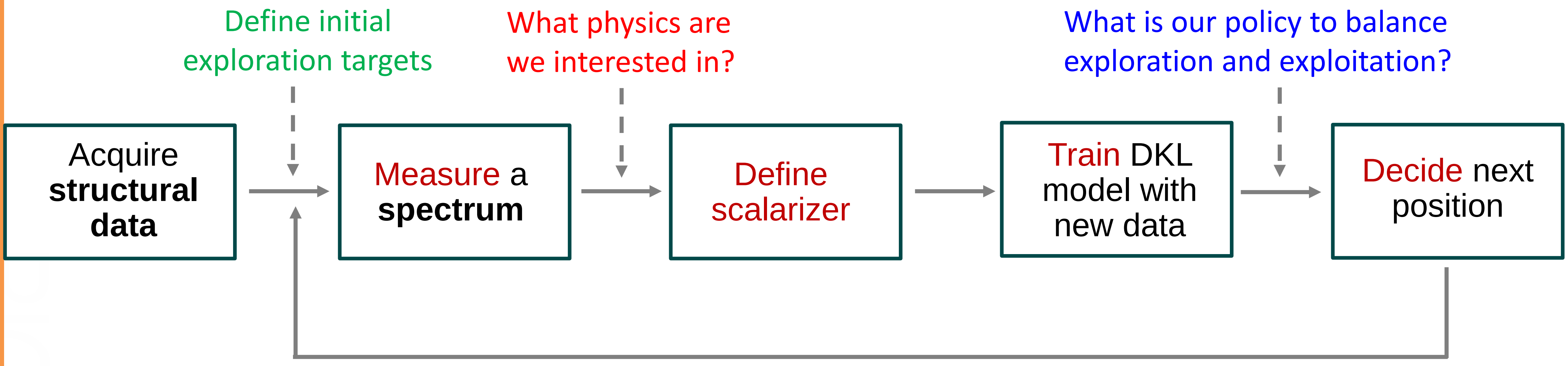


- **Workflow:** ideation, orchestration, implementation
- Domain specific language
- Dynamic planning: latencies and costs
- Reward and value functions

Designed in academia and adopted by industry

- Are they optimal?
- Can we design them better?
- Can they be changed dynamically?

Bringing Human into the Loop

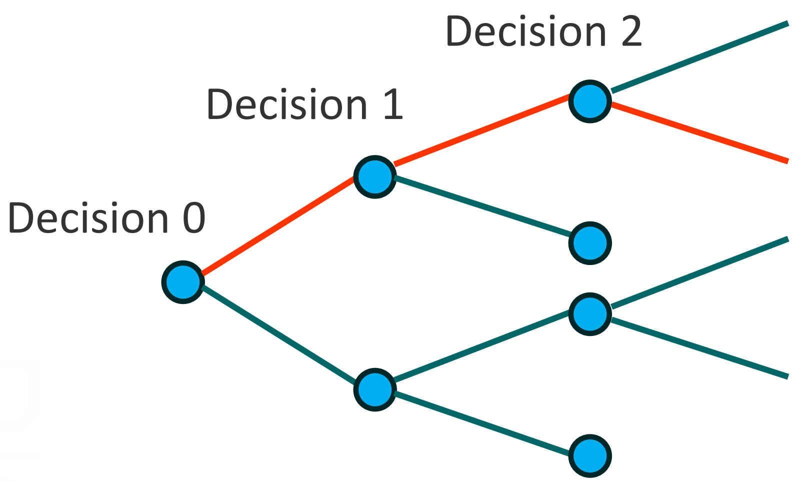


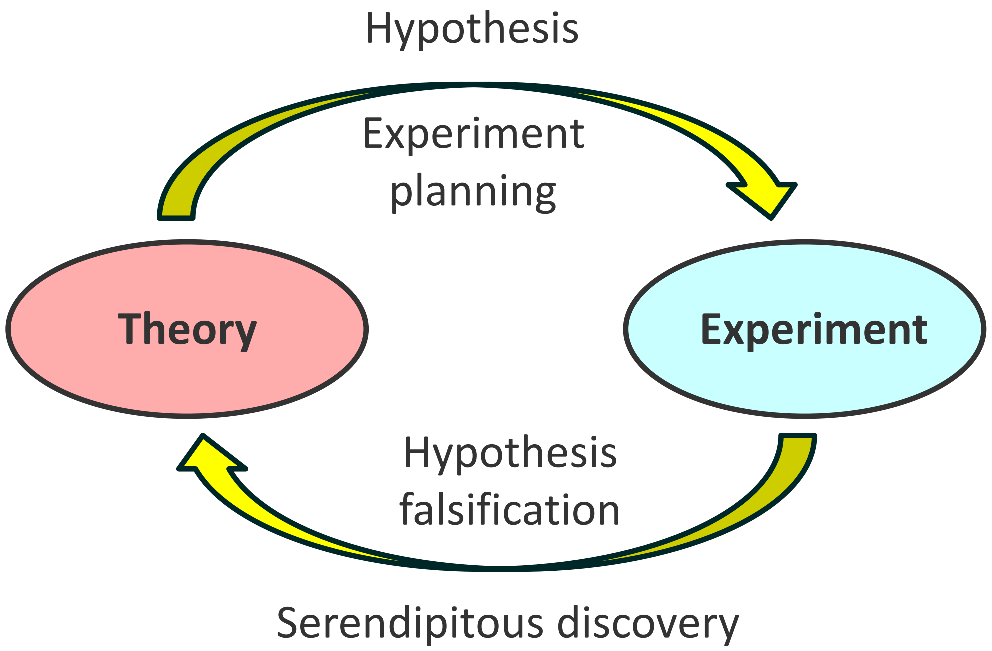
Key concepts:

- **Scalarizer:** (any) function that transforms spectrum into measure of interest. Can be integration over interval, parameters of a peak fit, ration of peaks, or more complex analysis

- **Experimental trace:** collection of image patches and associated spectra acquired during experiment. Note that we collect spectra, not only scalarizers

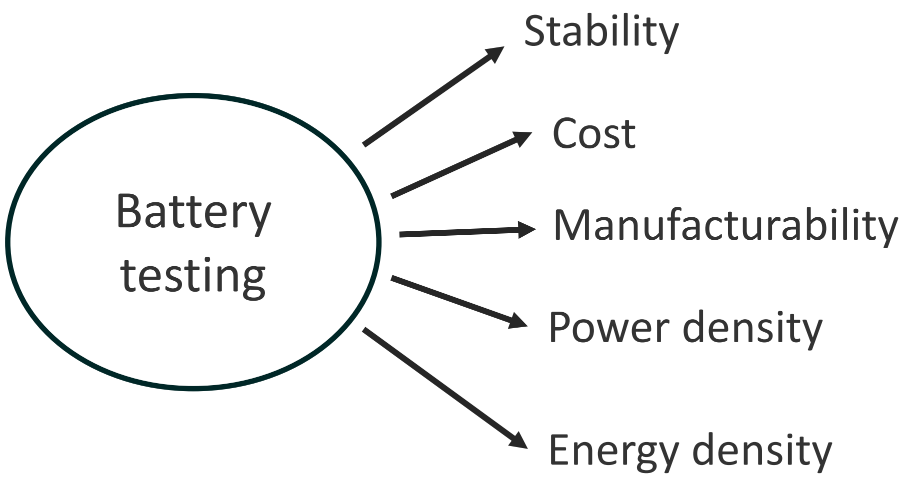
What do we hope to achieve?

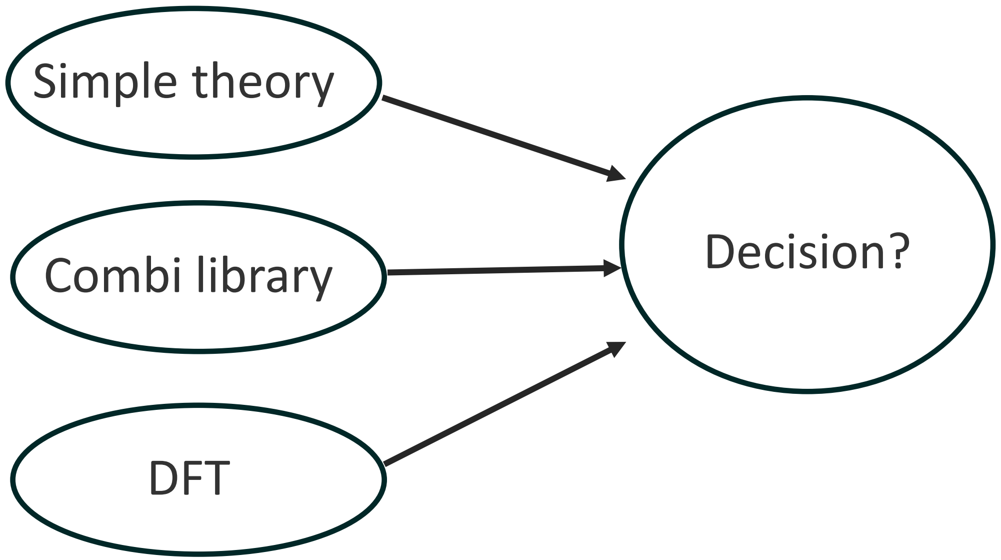




- **Experiment is a combinatorial space of opportunities:**
  → Investing only in scaling of throughput is only a linear improvement
- **Science is a cycle between theory-driven hypothesis generation and experiment**:
  → We need to accelerate all elements of the cycle
- **Experimental and computational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

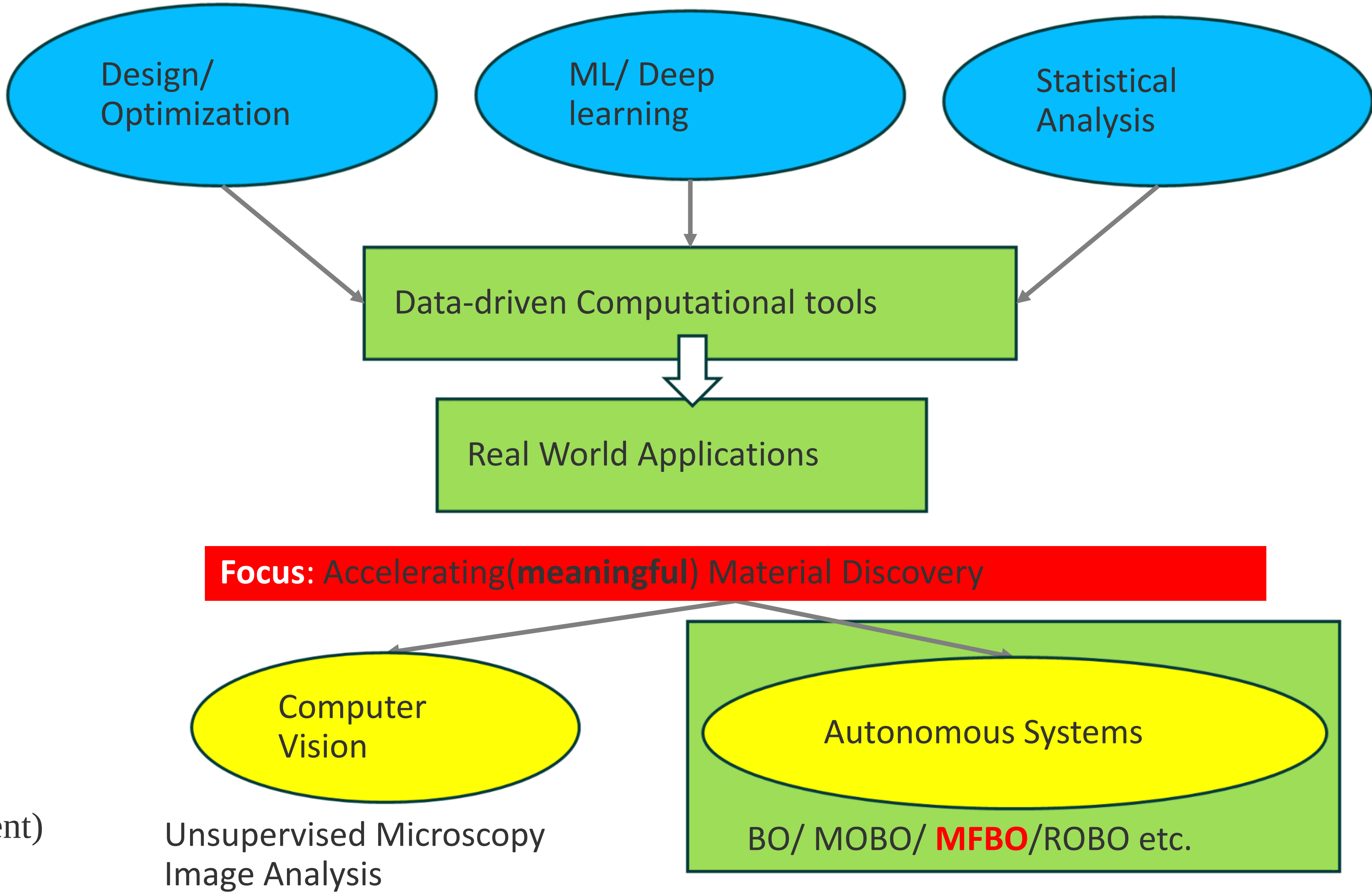
The real world is more complex!





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



Arpan Biswas

Postdoctoral Research Associate, DNA, CNMS, ORNL (April 2021 – Present)

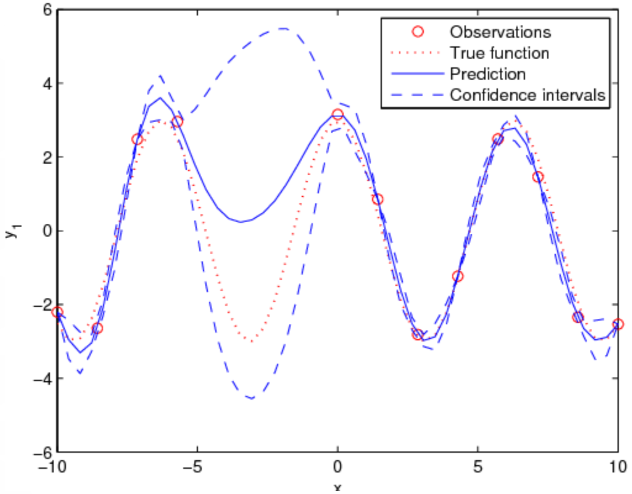
Multi-Objective Bayesian Optimization

Multi-objective Optimization:

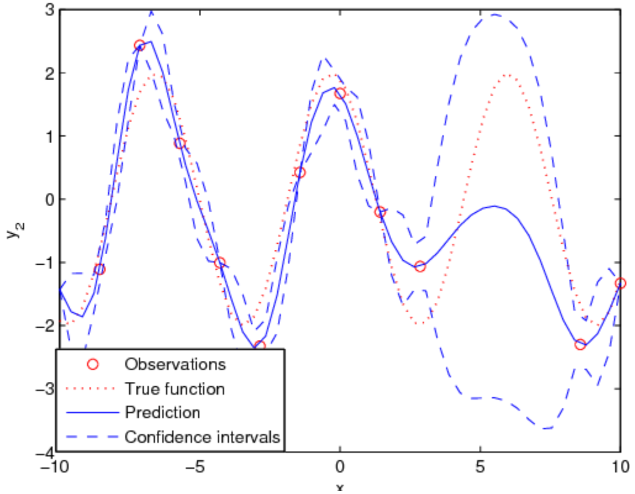
$$\min f(\boldsymbol{X}) = [\min f_1(\boldsymbol{X}), \min f_2(\boldsymbol{X}), \ldots . \min f_n(\boldsymbol{X})] \;\; s.t\; X \in \mathbb{R}$$

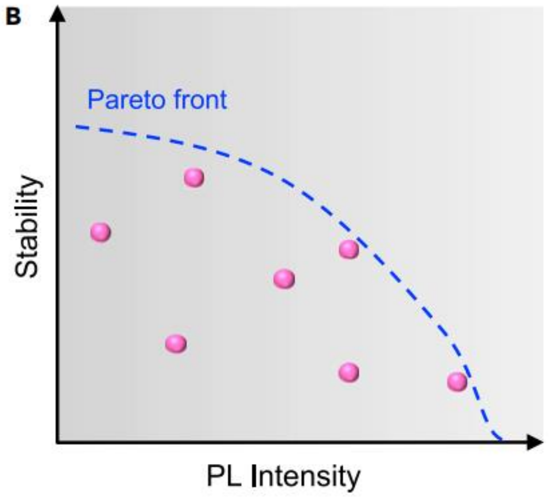
Multi-objective Bayesian Optimization: $\min f(\boldsymbol{X})$ where $f(\boldsymbol{X})$ is expensive to evaluate

(a) Output 1 (y_1)



(b) Output 2 (y_2)





Multi-output Gaussian Process

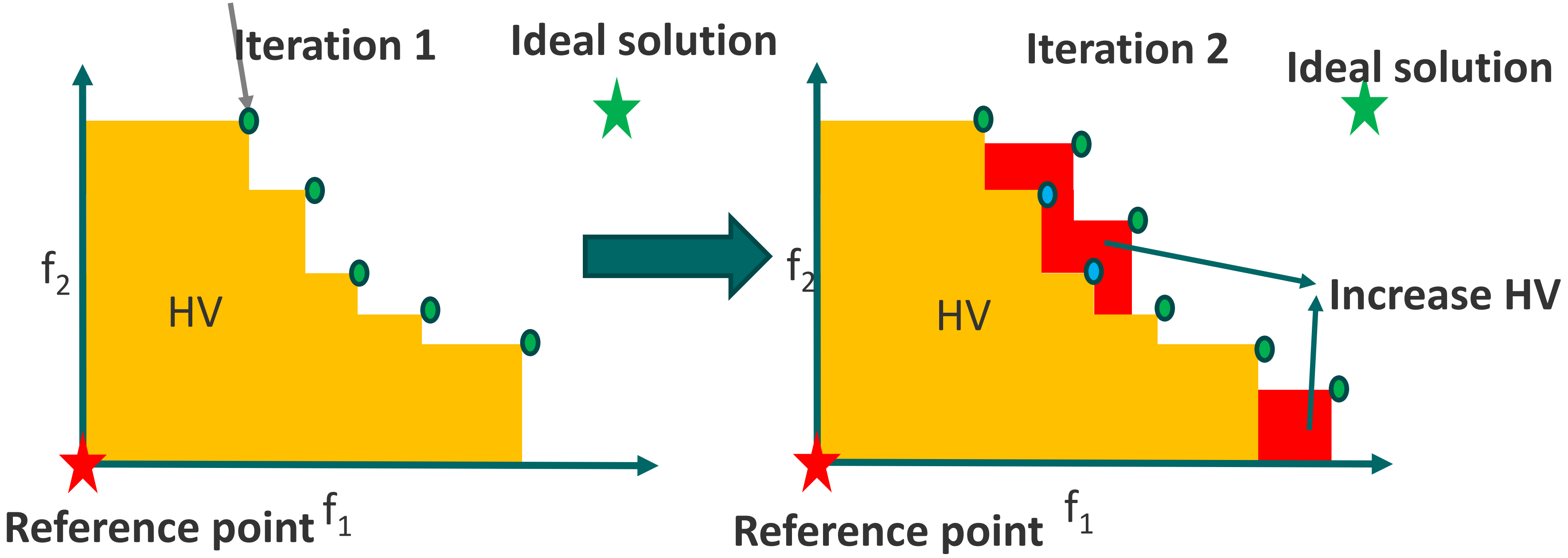
Multi-Objective Bayesian Optimization

Hyper-volume acquisition function

Posteriori Approach

Priori Approach

Weights NOT defined



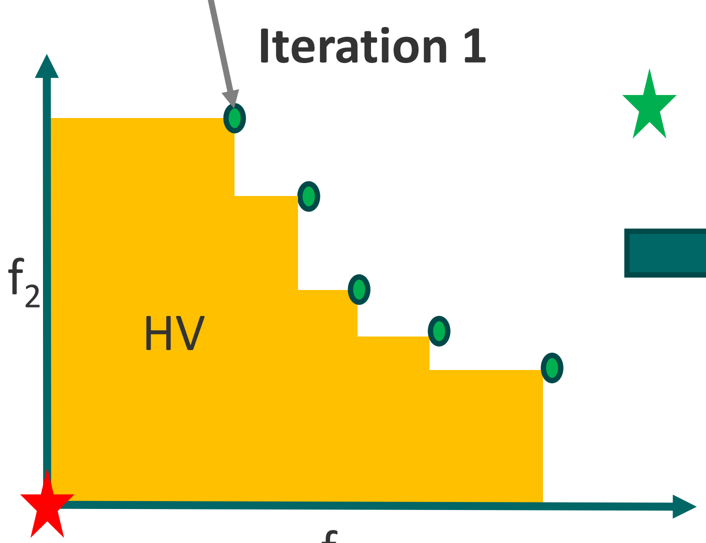
Multi-Objective Bayesian Optimization

Hyper-volume acquisition function

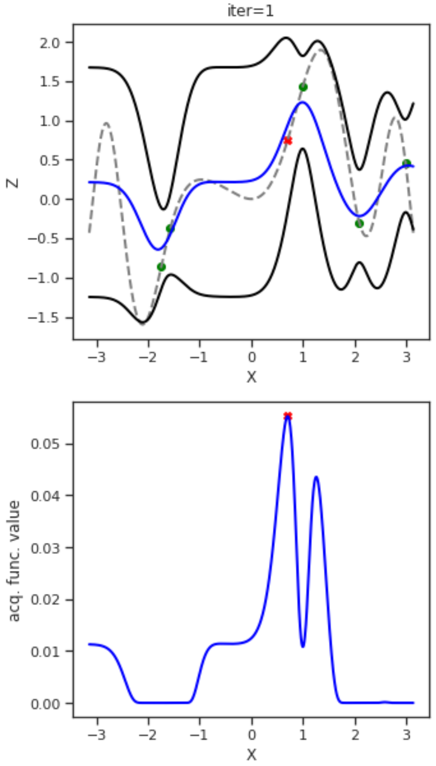
Posteriori Approach

Priori Approach

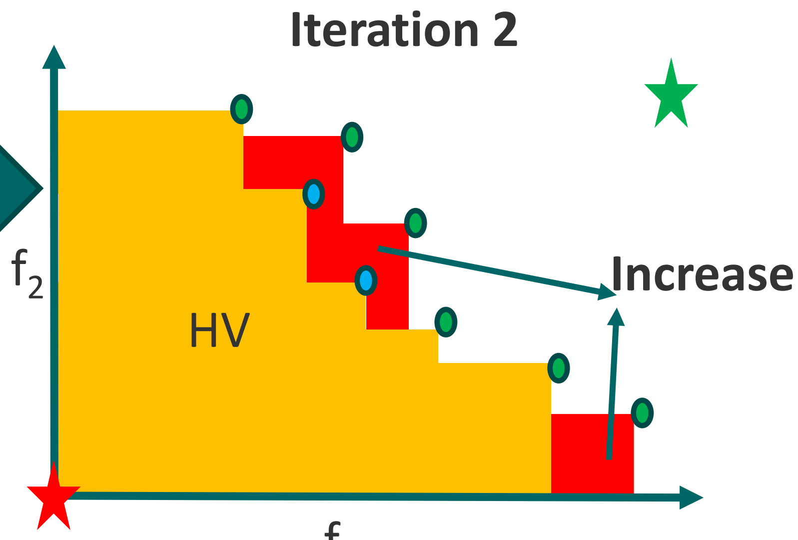
Weights PRE defined



2D objective space



1D multi-objective space



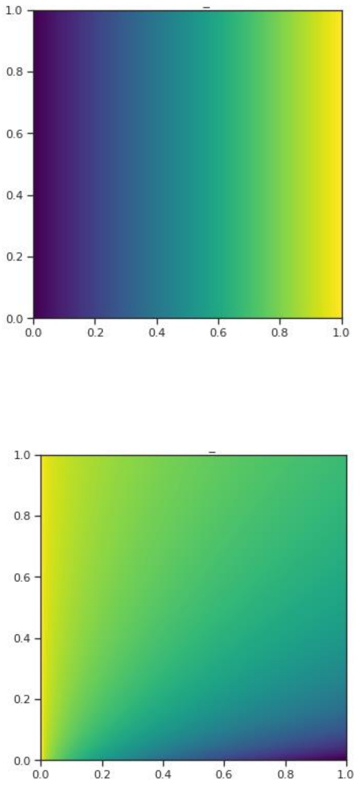
2D objective space

Case Study 1

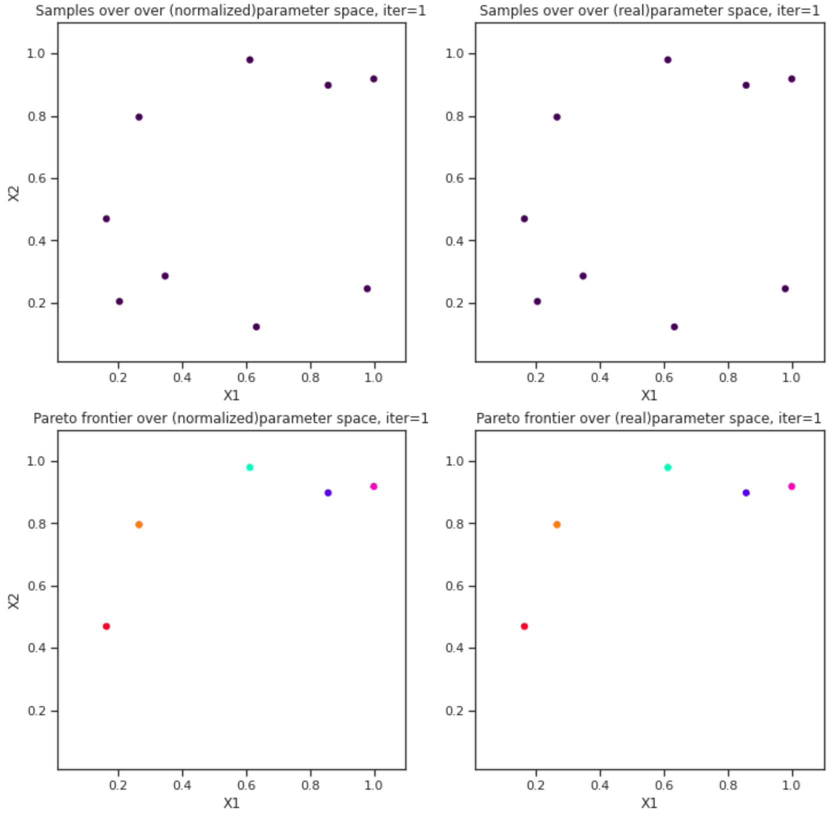
Ground truth

f$_1$

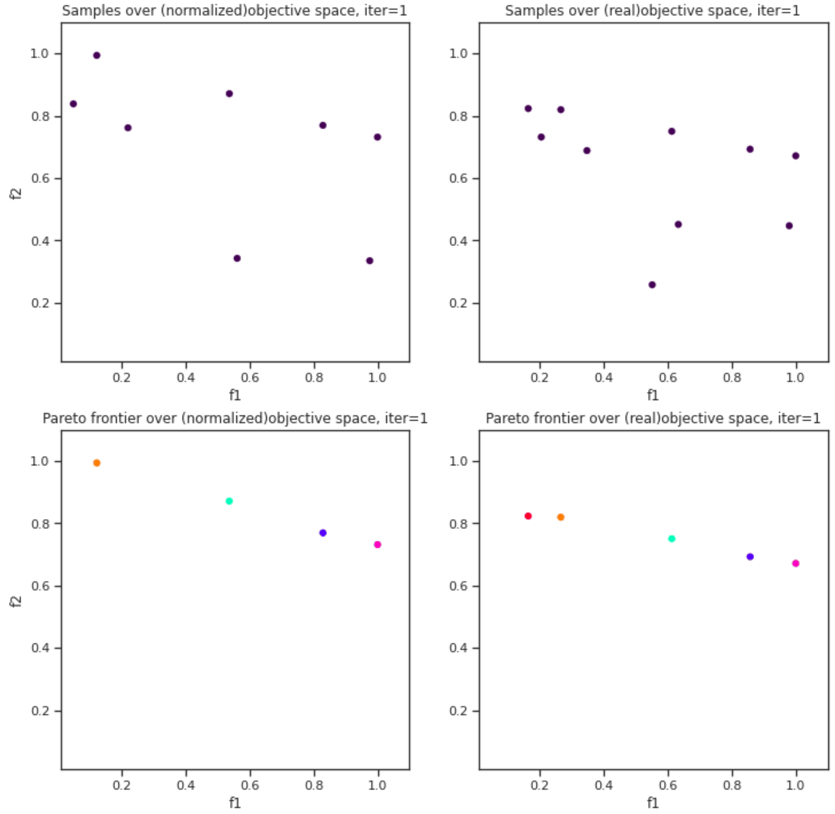
f$_2$



Building Pareto over design space



Building Pareto over objective space



Numerical Test Problems: ZDT1

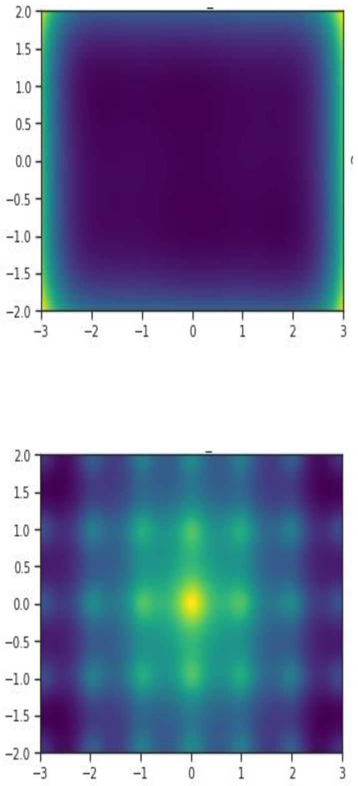
Acquisition Function: qEIHV (Posteriori); Batch_size: 4, Max BO sampling: 50 x Batch_size

Case Study 2

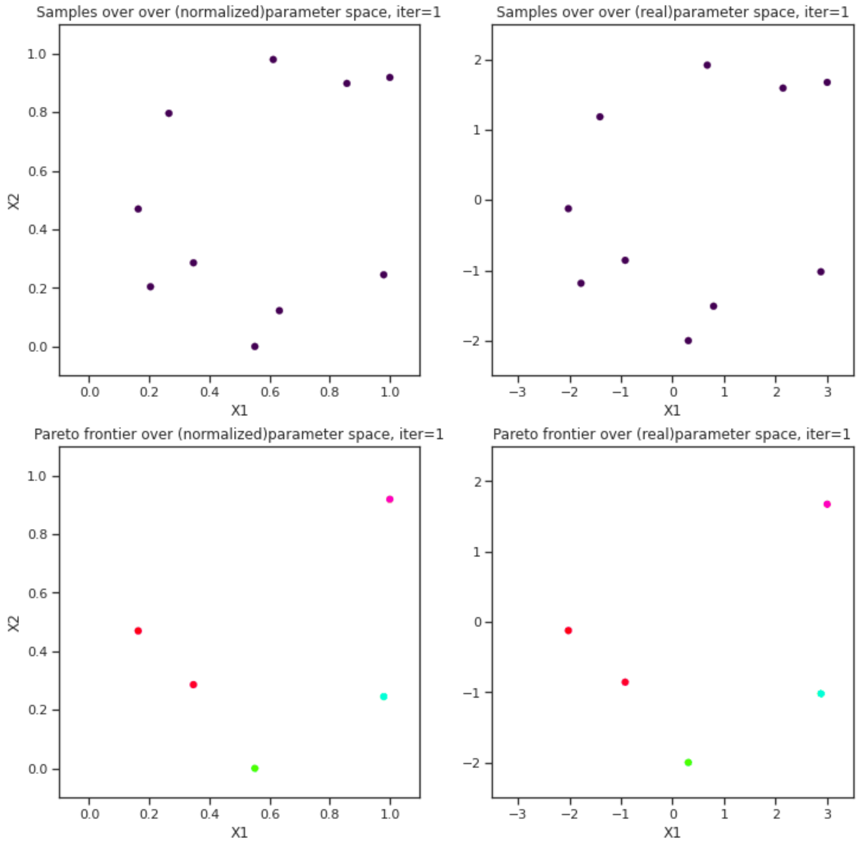
Ground truth

f$_1$

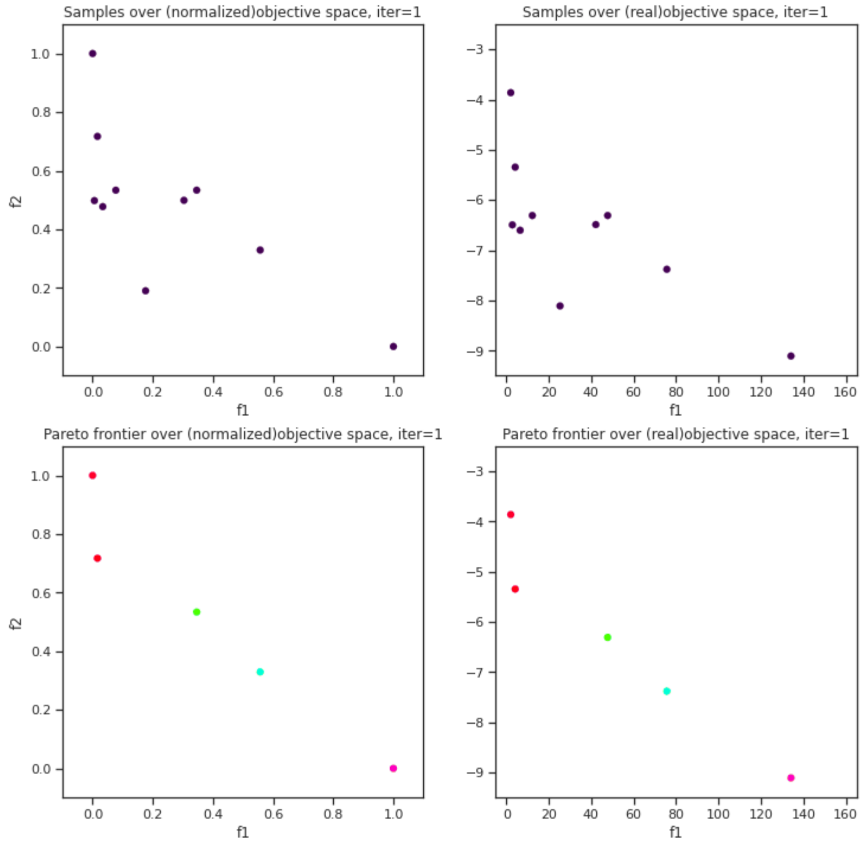
f$_2$



Building Pareto over design space



Building Pareto over objective space

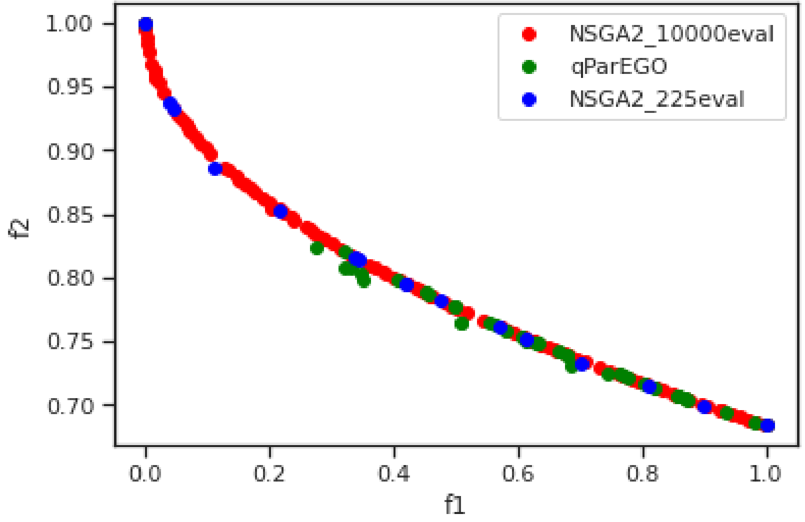


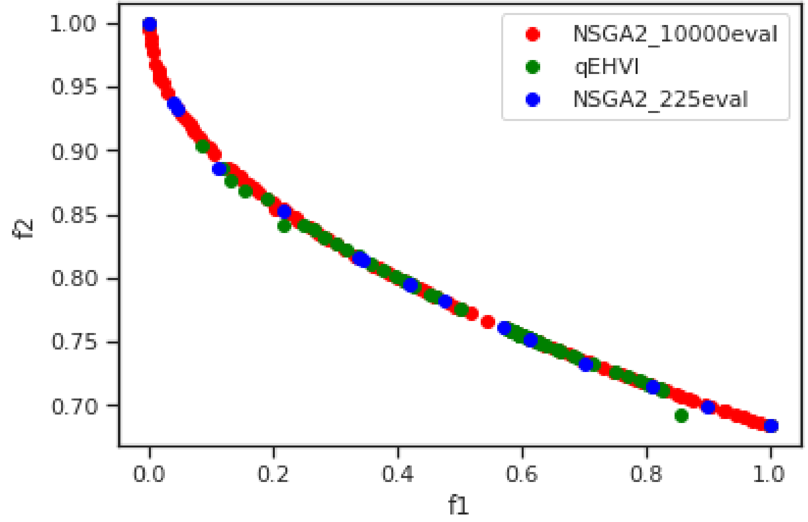
Numerical Test Problems (Non-Physics): 6-Hump Camel Back – Inversed Ackley's Path (6HC-IAP)

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

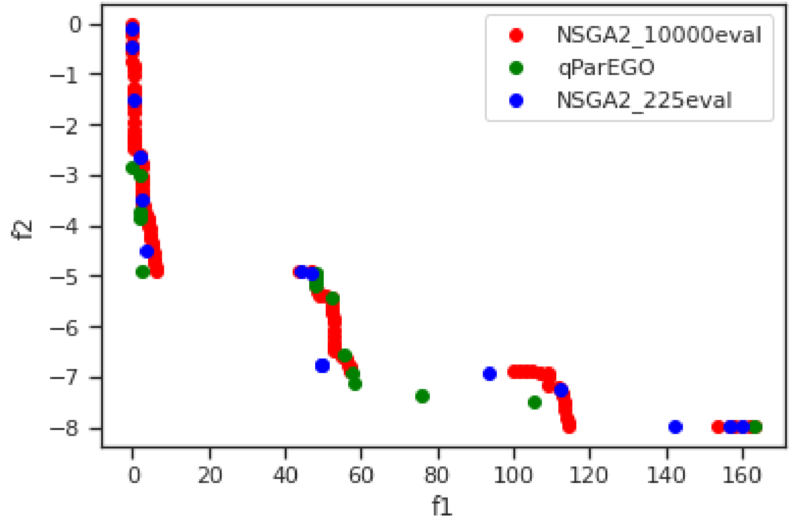
Case Study 1 and 2

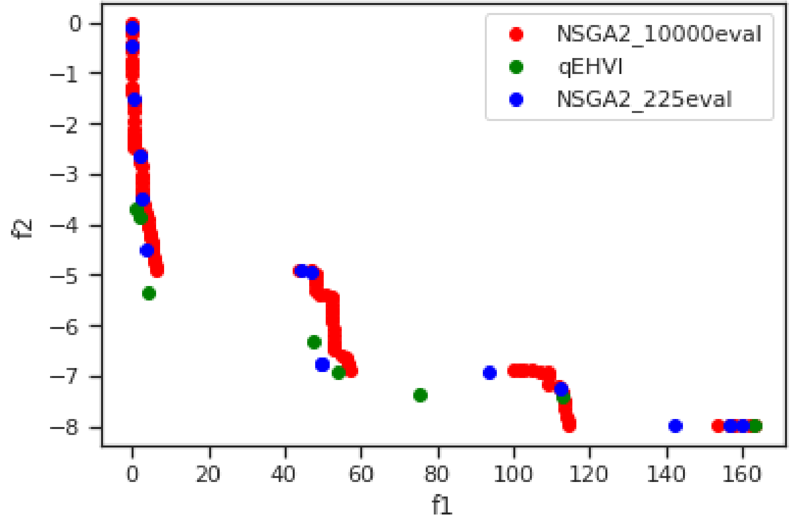
ZDT1



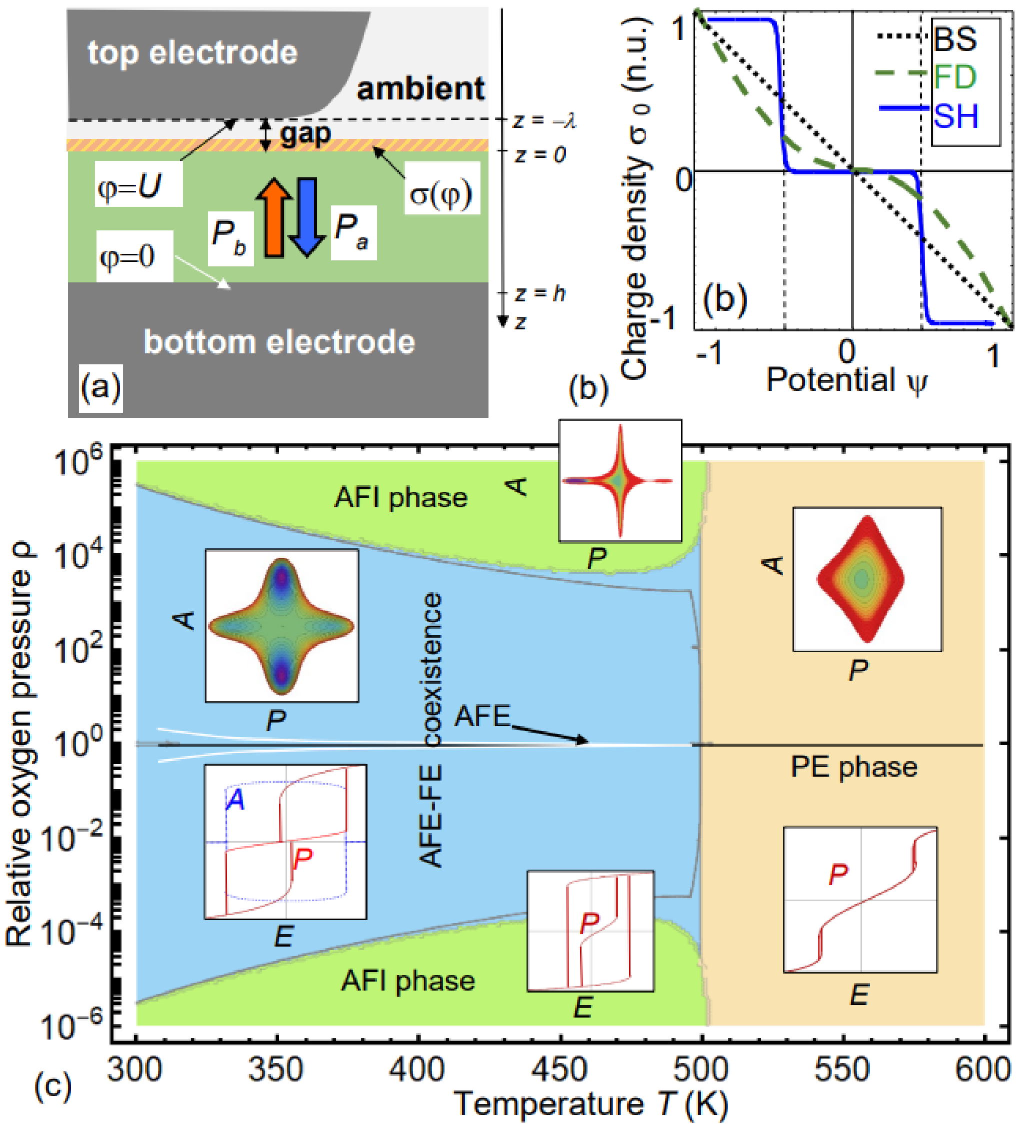


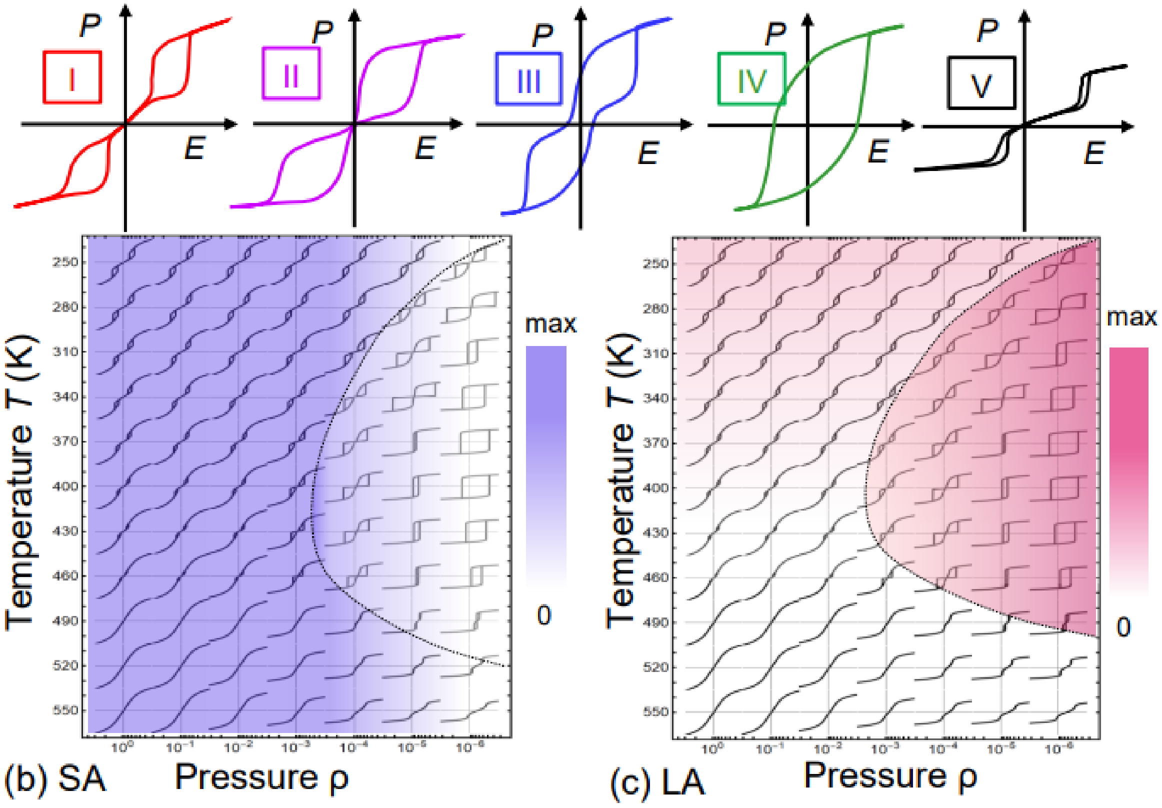
6-Hump Camel Back – Inversed Ackley's Path (6HC-IAP)





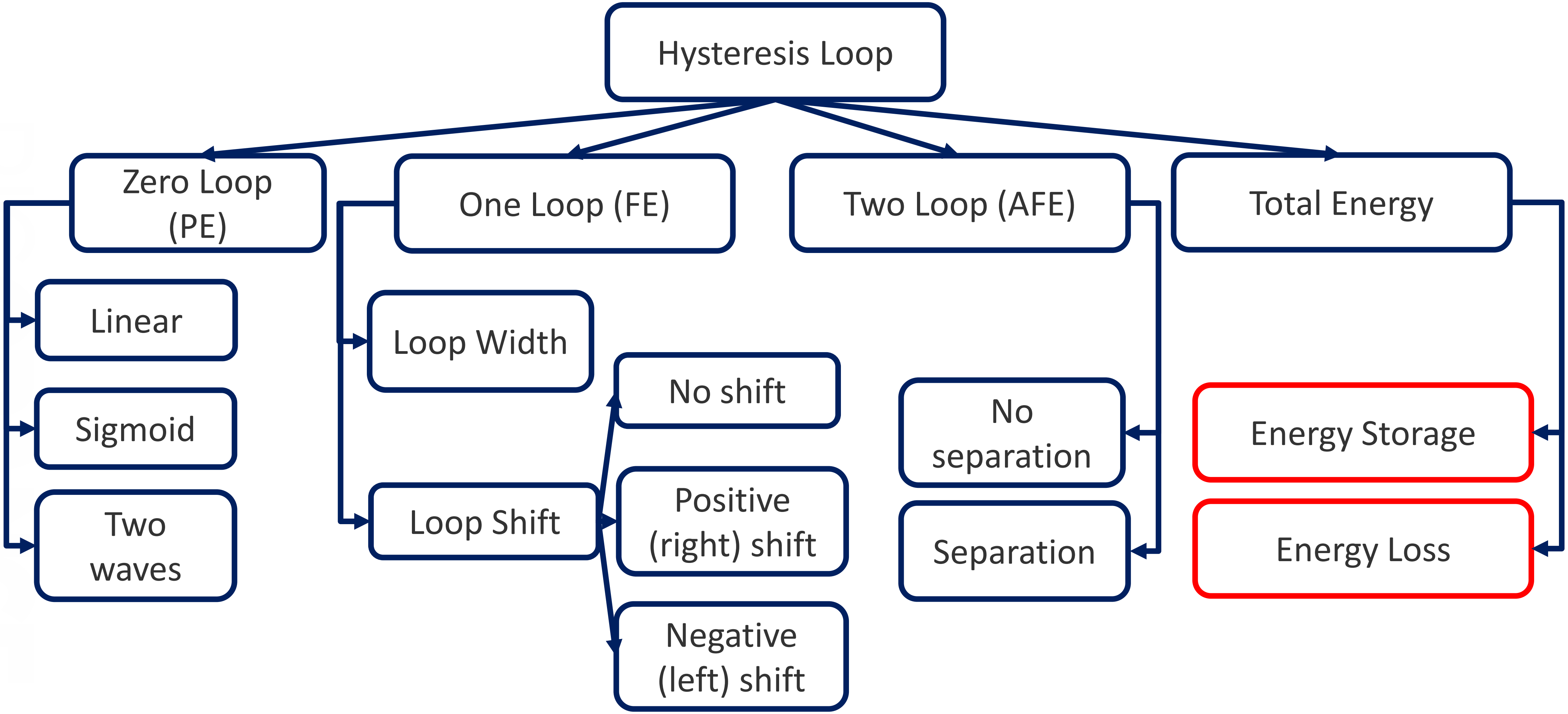
Applications for physical problems



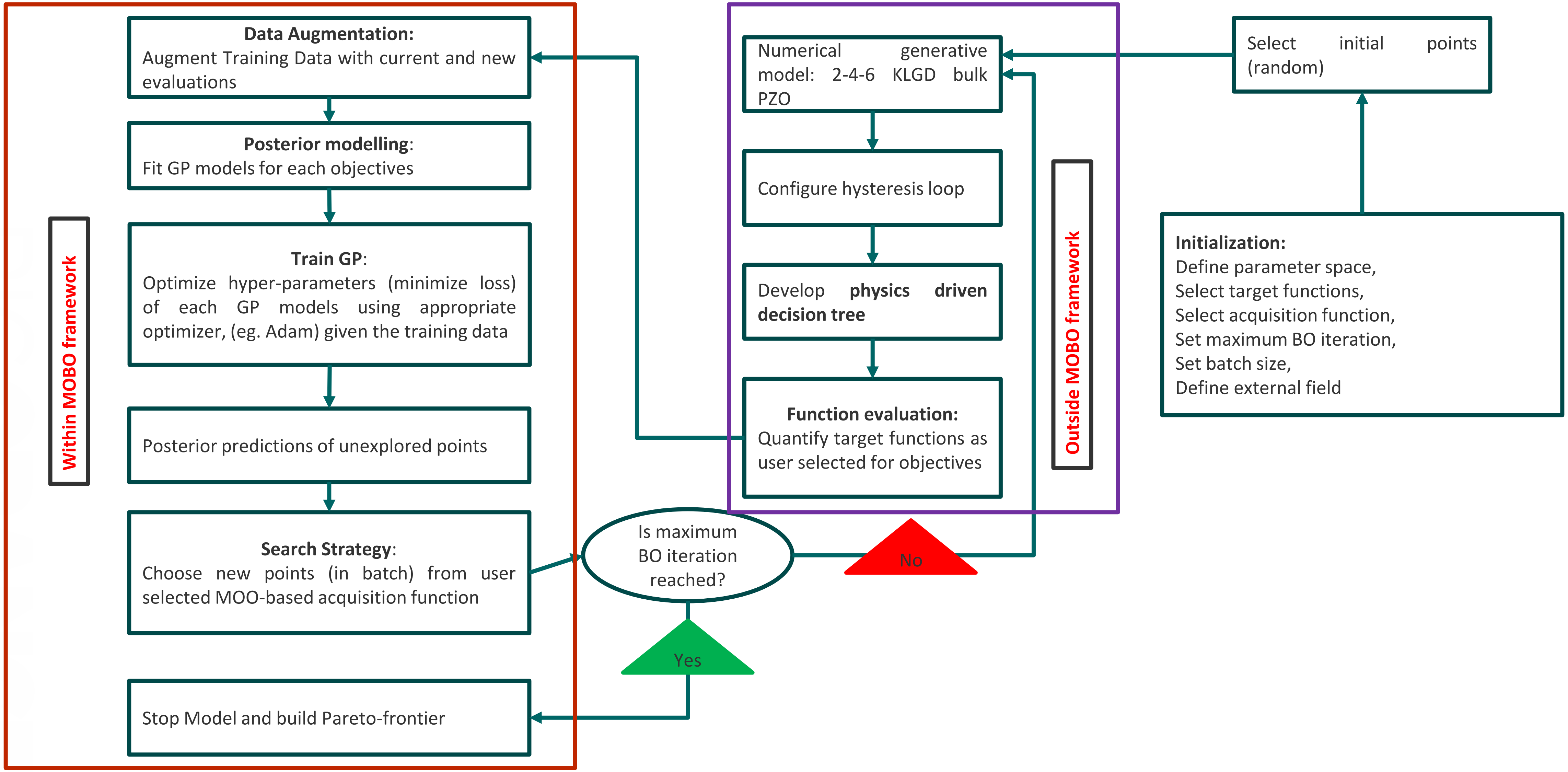


Applications for physical problems

Case Study Theoritical Model: 2-4-6 KLGD for bulk PZO
proposed by -- Anna N. Morozovska and Sergei V. Kalinin,
developed by -- Eugene A. Eliseev (in Mathematica) and Arpan Biswas (in Python)

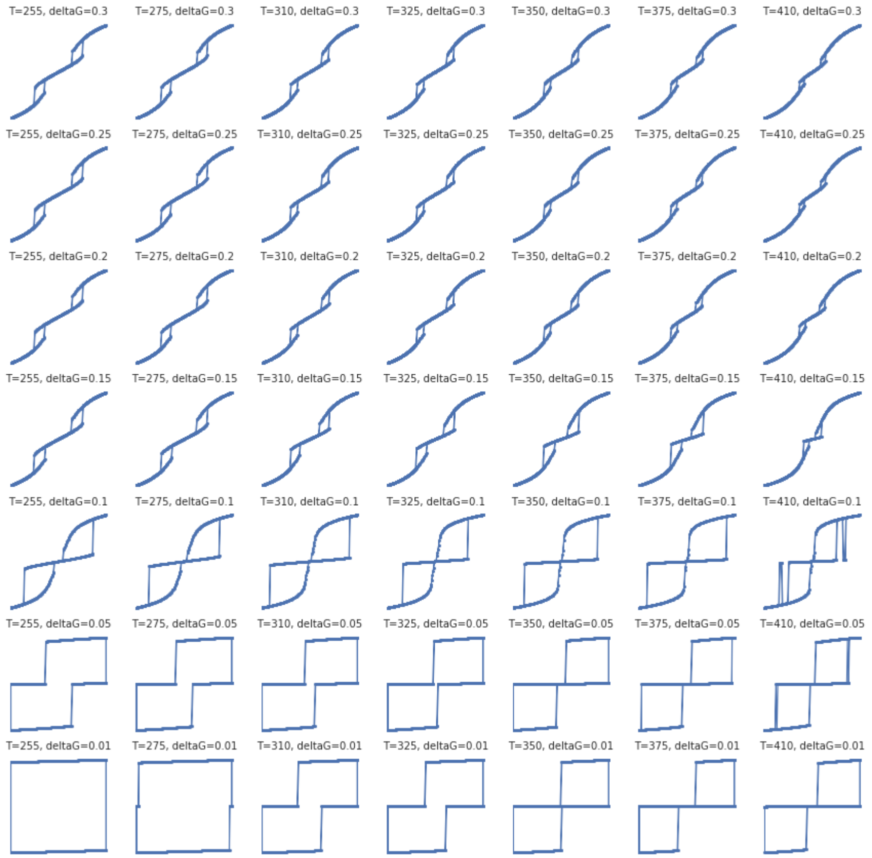


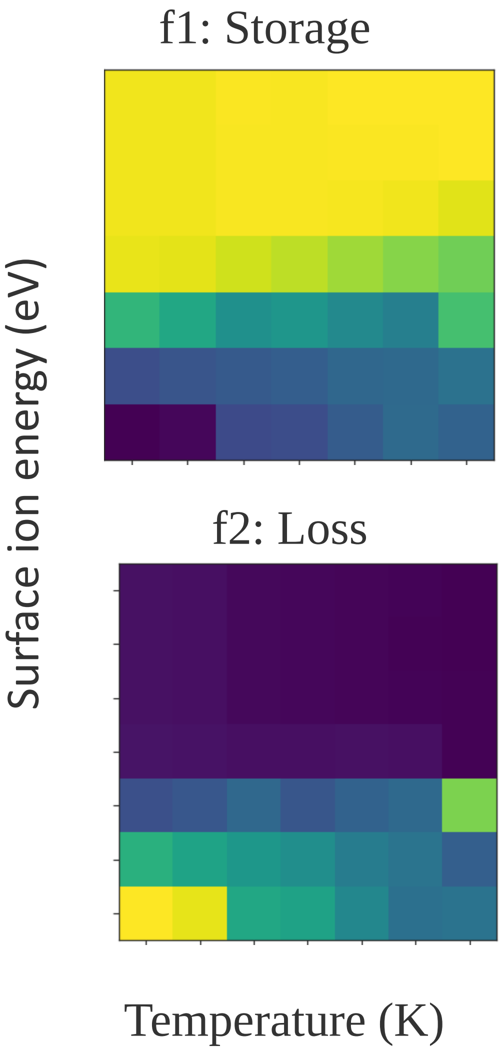
Applications for physical problems

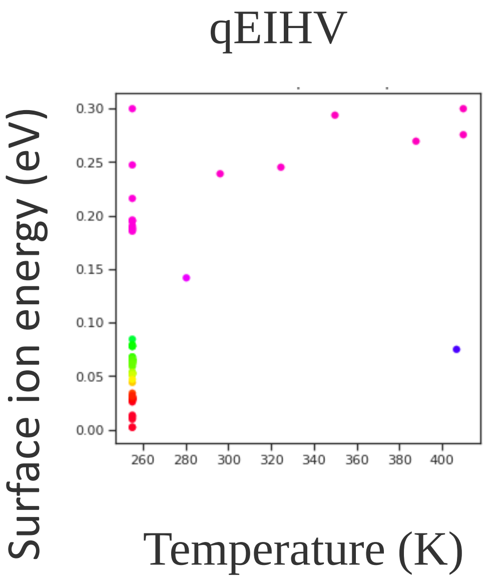


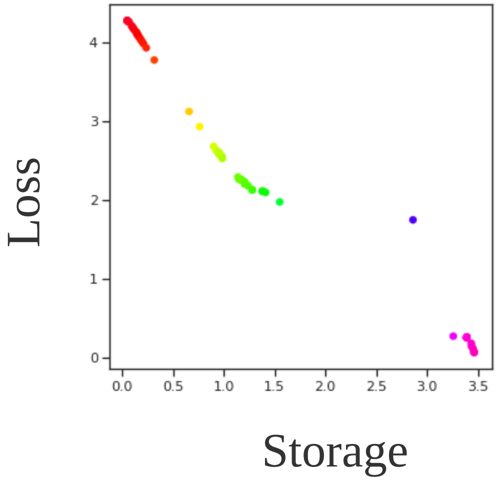
Applications for physical problems

Hysteresis Loops: FE and AFE









Maximize Energy Storage, *Maximize* Energy Loss

Parameter space $T = [255, 410]K, \rho = 10^2, h = 5nm, \Delta_G = [0.002, 0.3]eV$.

Arpan Biswas, Anna N. Morozovska, Maxim Ziatdinov, Eugene A. Eliseev, and Sergei V. Kalinin "Multi-objective Bayesian optimization of ferroelectric materials with interfacial control for memory and energy storage applications" Journal of Applied Physics 130, 204102 (2021);

https://doi.org/10.1063/5.0068903

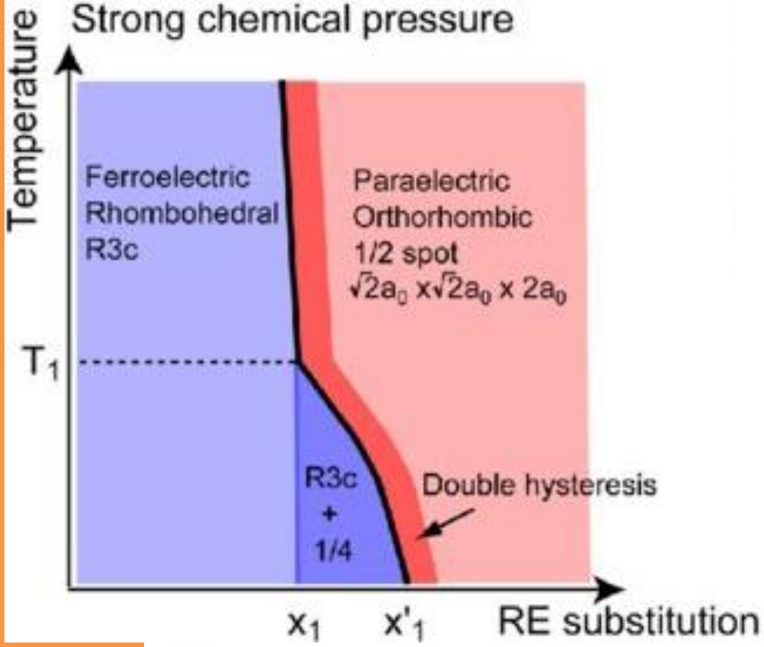
Full Notebook: https://github.com/arpanbiswas52/MOBO_PhysicsBasedModels

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.

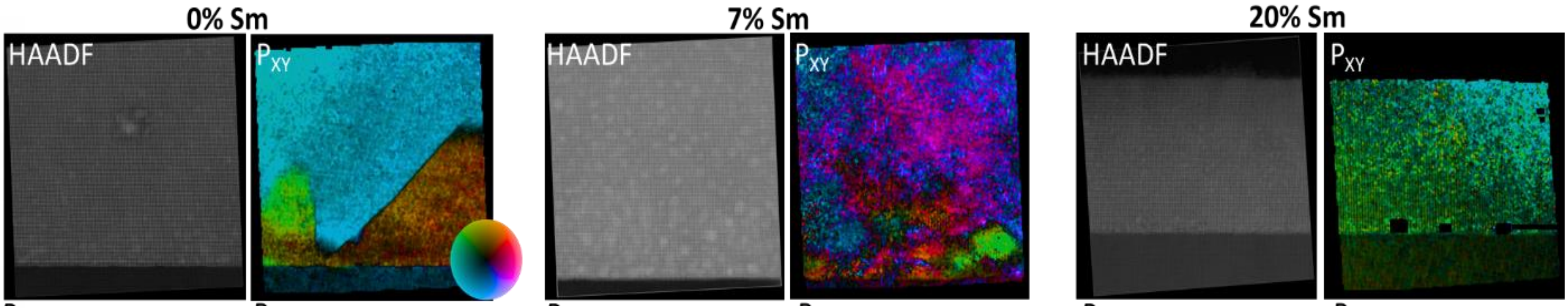
How can we utilize both ? Time for multi-fidelity !!

Combinatorial Synthesis: Expensive Measurements



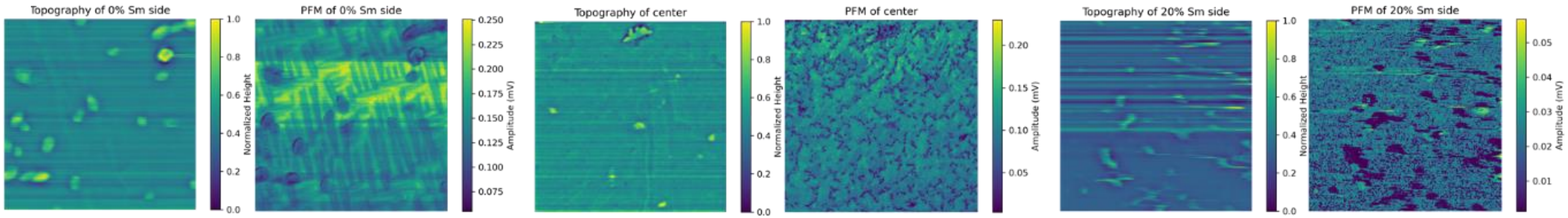
Sample by I. Takeuchi, UMD
Phase diagram by N. Valanoor et al.

$BiFeO_3$ Linear est. 7%Sm $BiFeO_3$ 20%Sm $BiFeO_3$



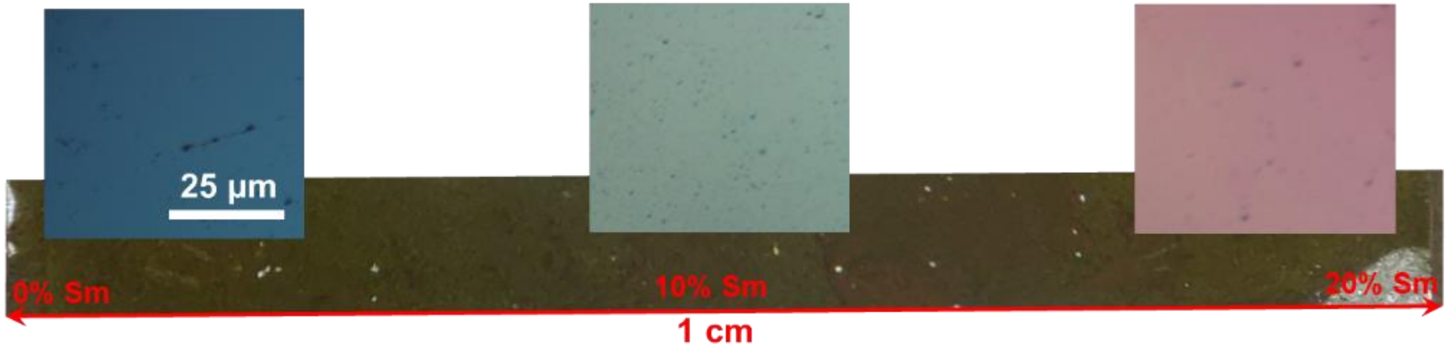
arXiv:2004.11817

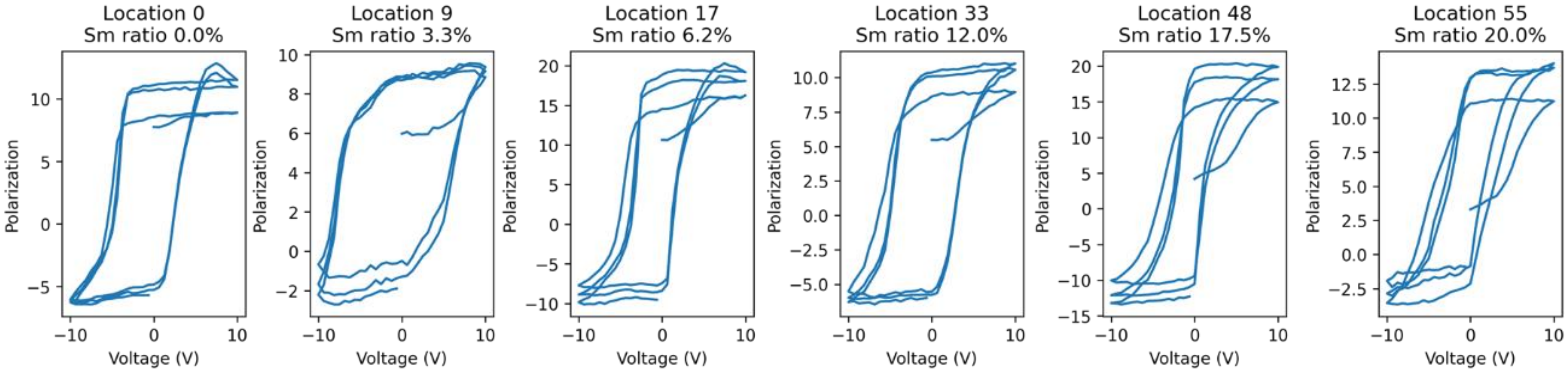
Combinatorial Synthesis: Cheap(er) Measurements



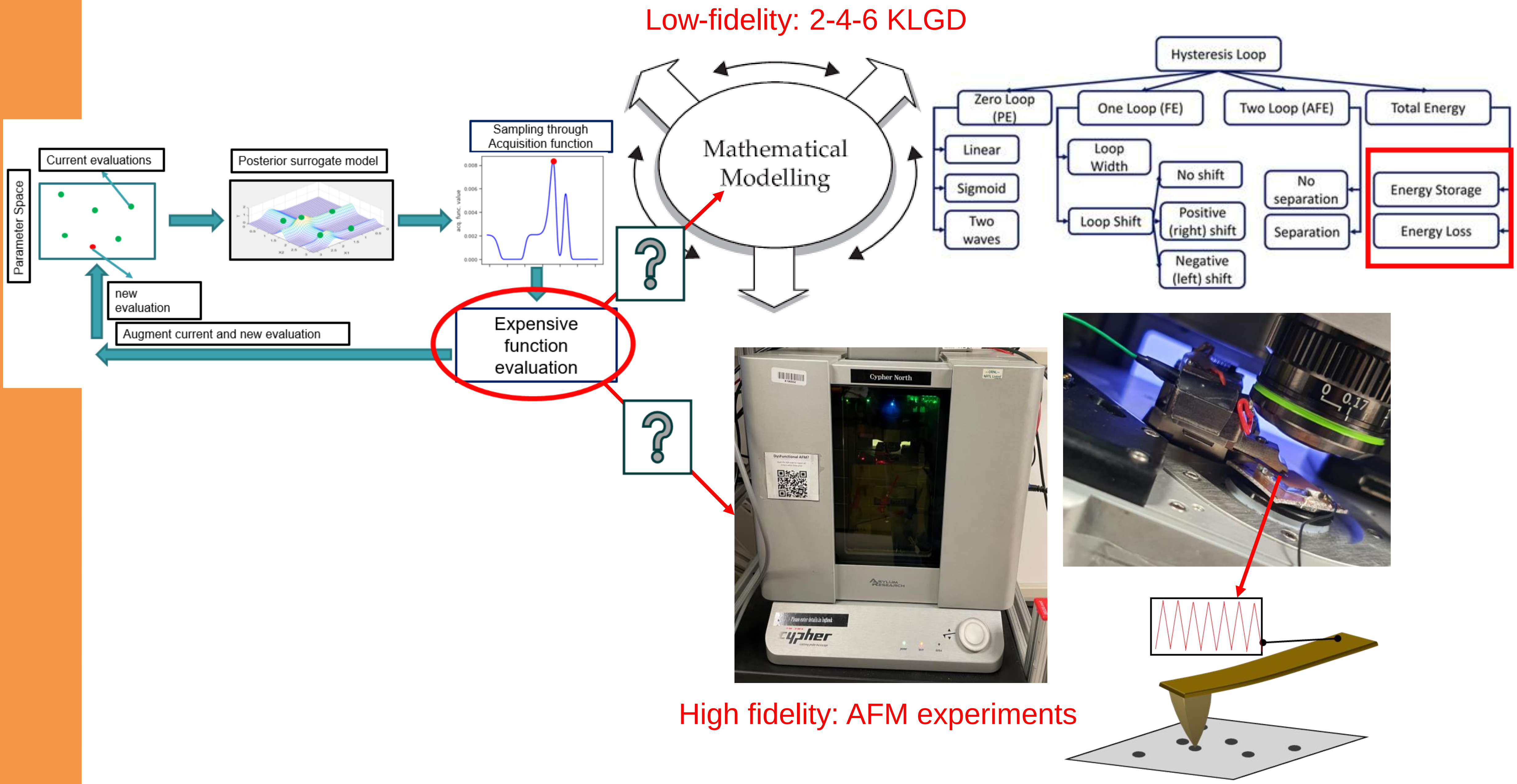
- optical images

sample picture

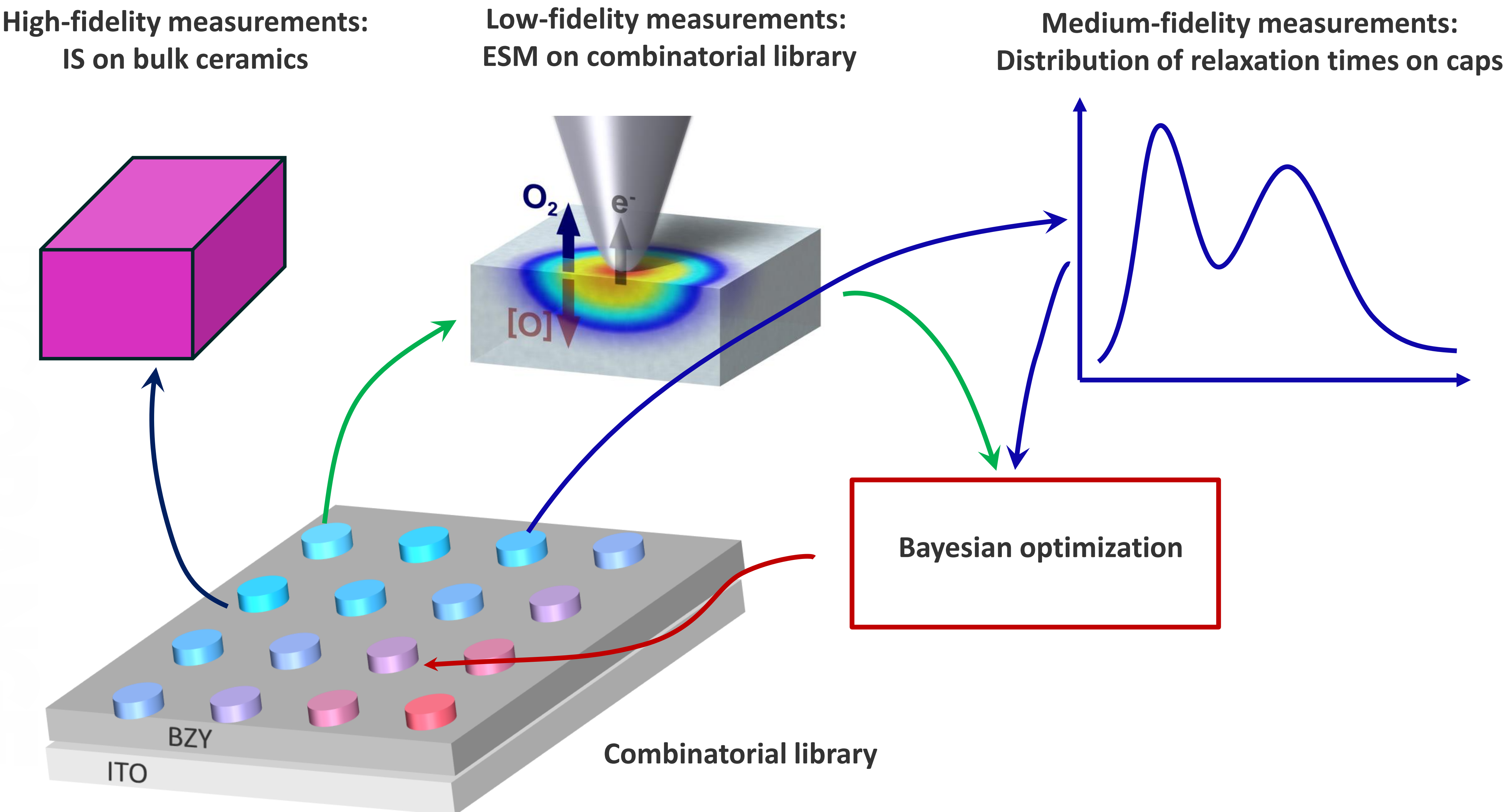




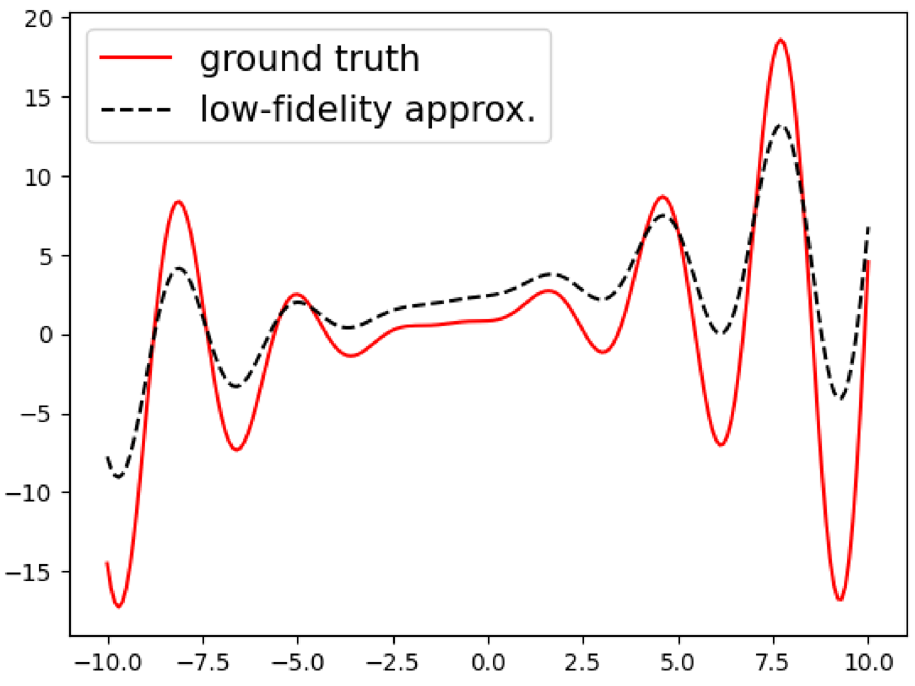
Multifidelity Optimization

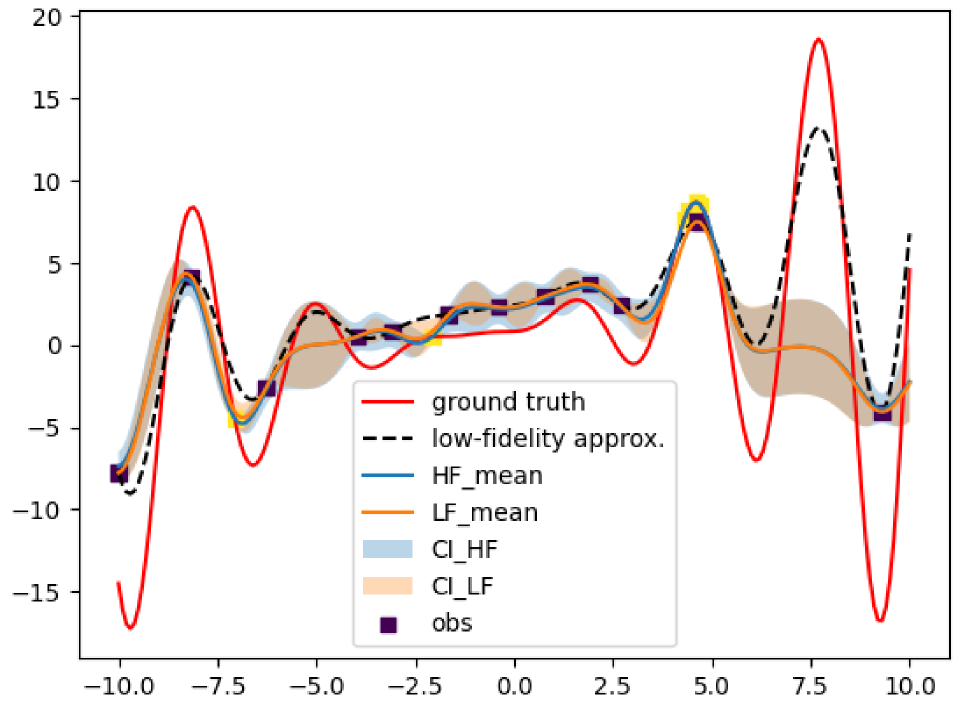


Multiple fidelity cycles are possible



Multifidelity Gaussian Processes





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

Polynomial Regression

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

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

Fidelity kernel

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

$$\delta > 0$$

hyperparameter

Multifidelity acquisition functions

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

Acquisition value of x, given HF

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

Same as EI for standard BO

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

Acquisition value of x, given LF

$$\Delta EI_l(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*)) = EI(\boldsymbol{x}^*) - EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))$$

Further improvement with LF samples

Similar trend

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

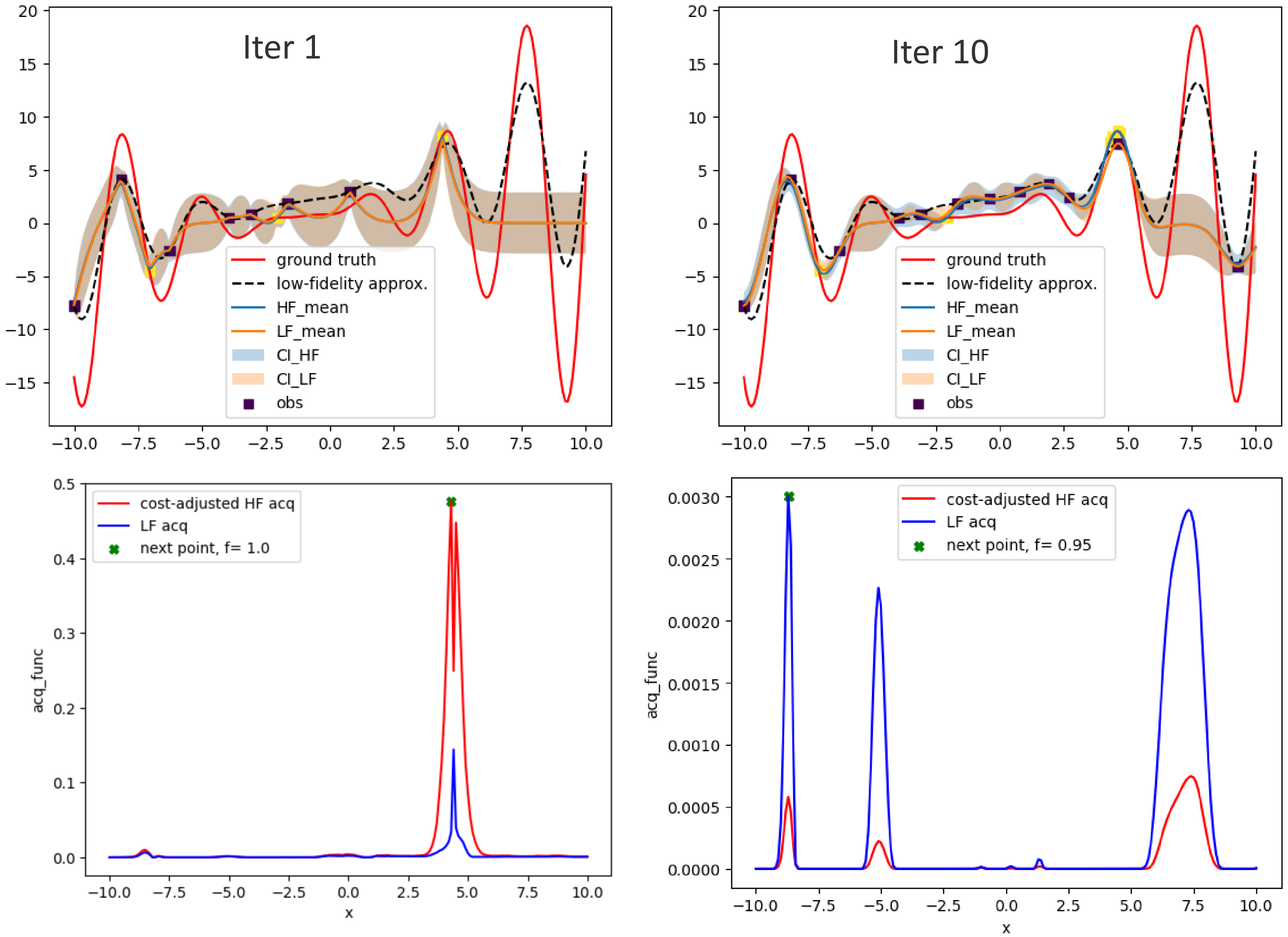
Multi-fidelity acquisition function

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

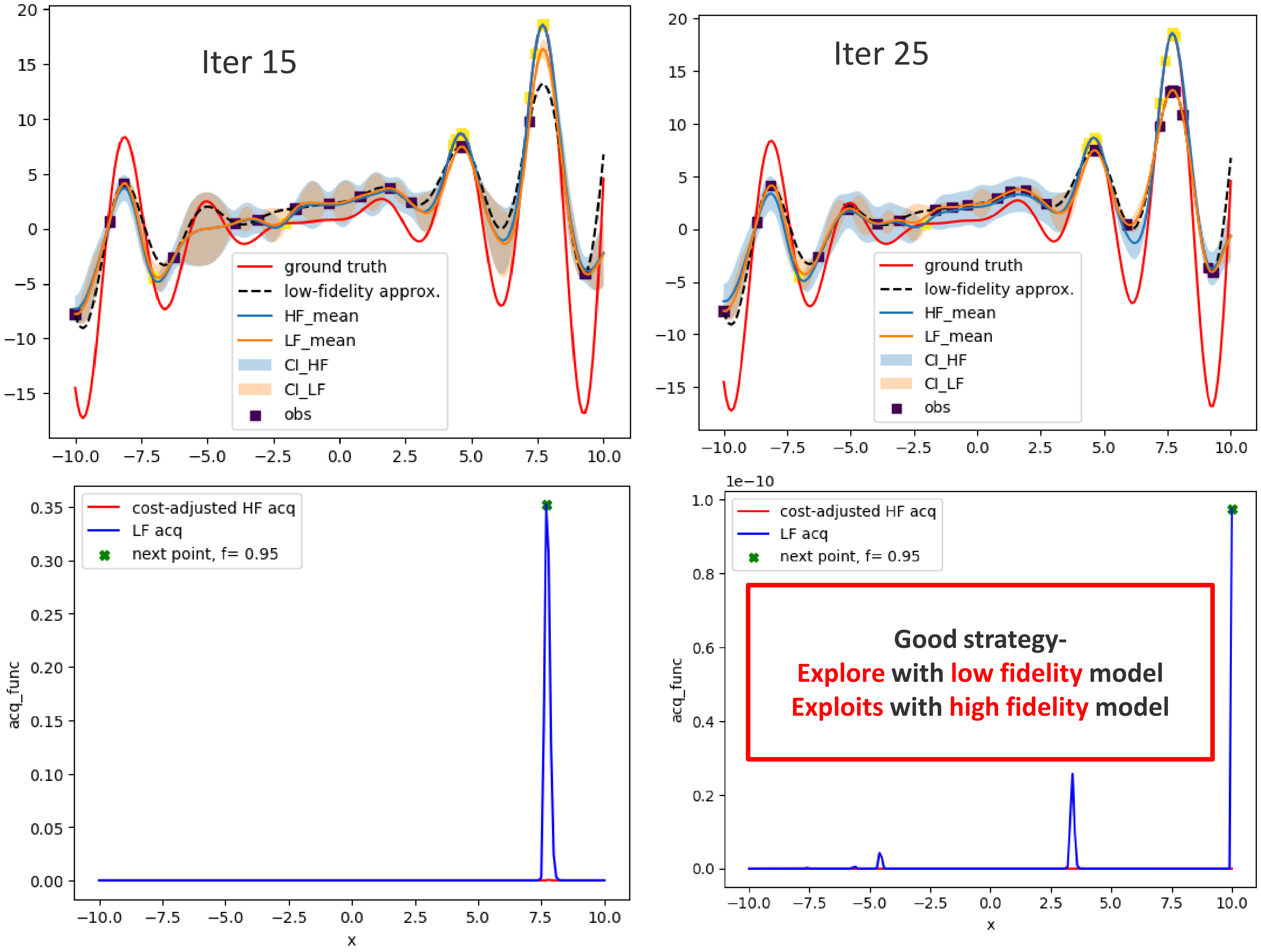
Cost-ratio: Can be derived from model complexity and domain knowledge

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:

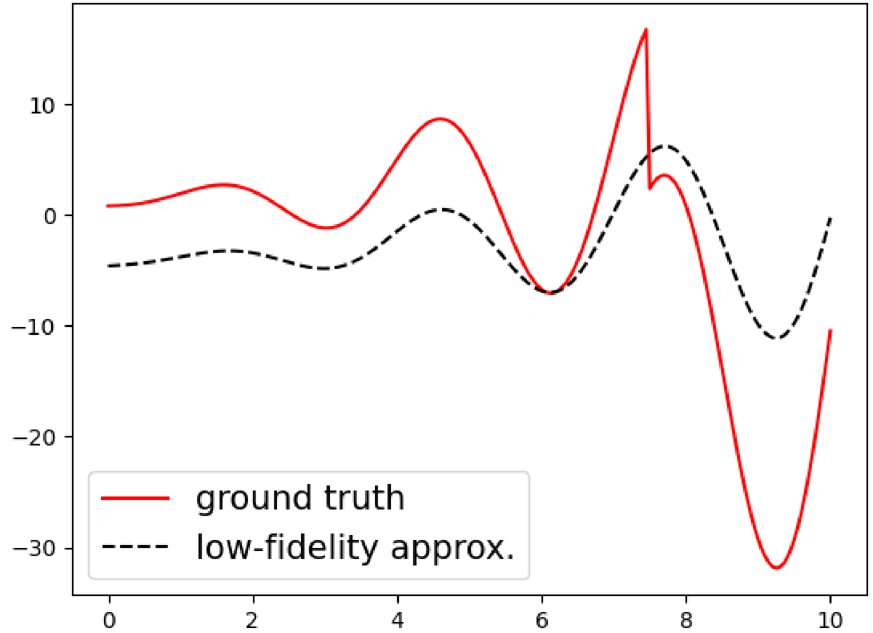
1. **2D Ising Model:** Low-fidelity (20x20) → High-fidelity (60x60)
2. **Hybrid perovskites:** Low-fidelity (cheap measurements) → High-fidelity (expensive measurements)
3. **Material Synthesis:** Low-fidelity (combi library) → High-fidelity (PLD synthesis)
4. and many more to come ….

Hence this tutorial!

Arpan Biswas

Postdoctoral Research Associate, DNA, CNMS, ORNL (April 2021 – Present)

Structured Multifidelity GP



Mean function

$$M_f(x,a,b,c) = \begin{cases} f_1(x,a), & x < c \\ f_2(x,b), & x \geq c \end{cases}$$

$$a = Unif[5,10], b = N(0,1), c = N(15,2)$$

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

Polynomial Regression

$$z(x) \sim GP\left(\boldsymbol{M_f}, \sigma^2 R(x^i, x^j) * K_I(f^i, f^j)\right)$$

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

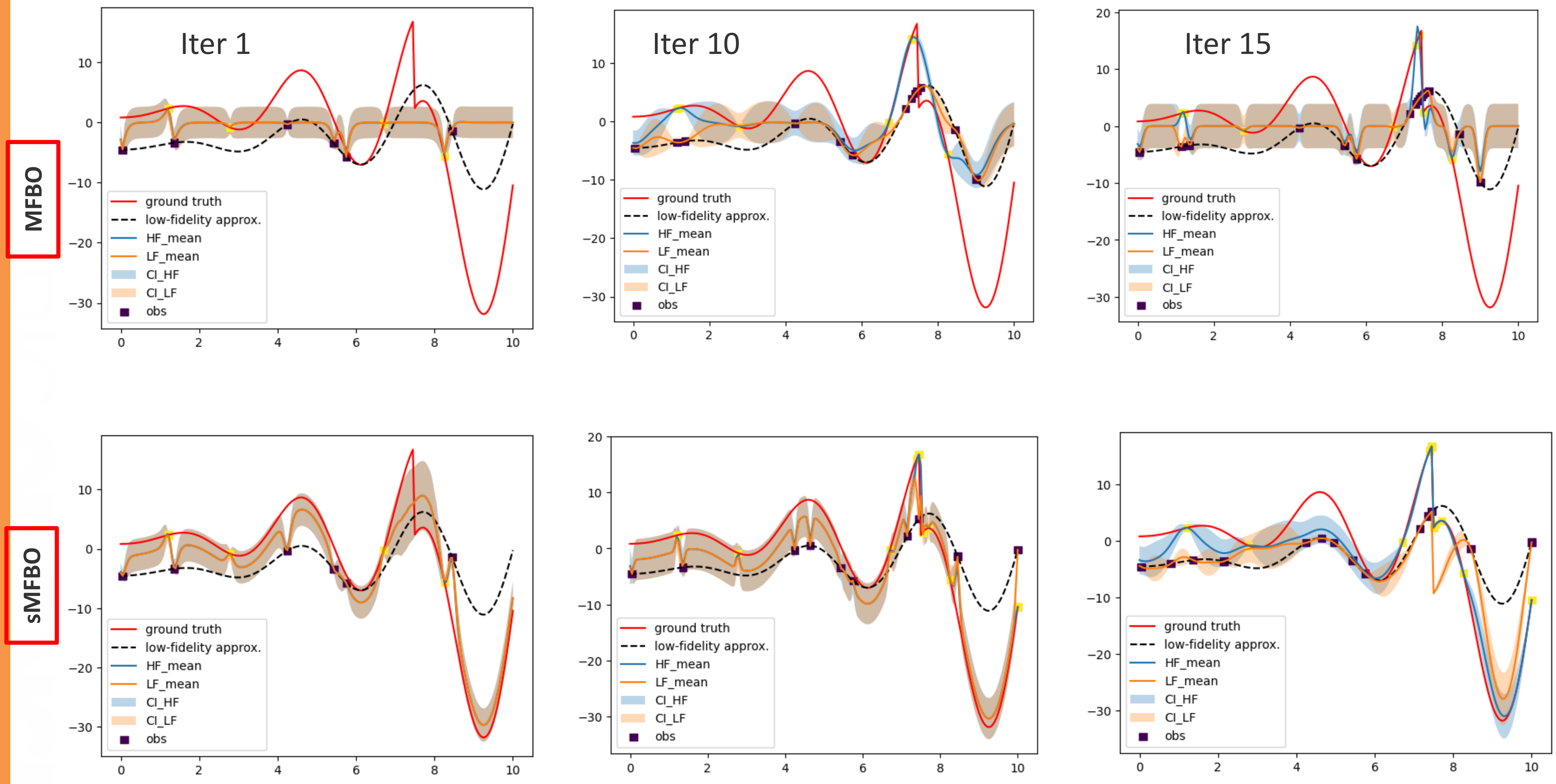
Fidelity kernel

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

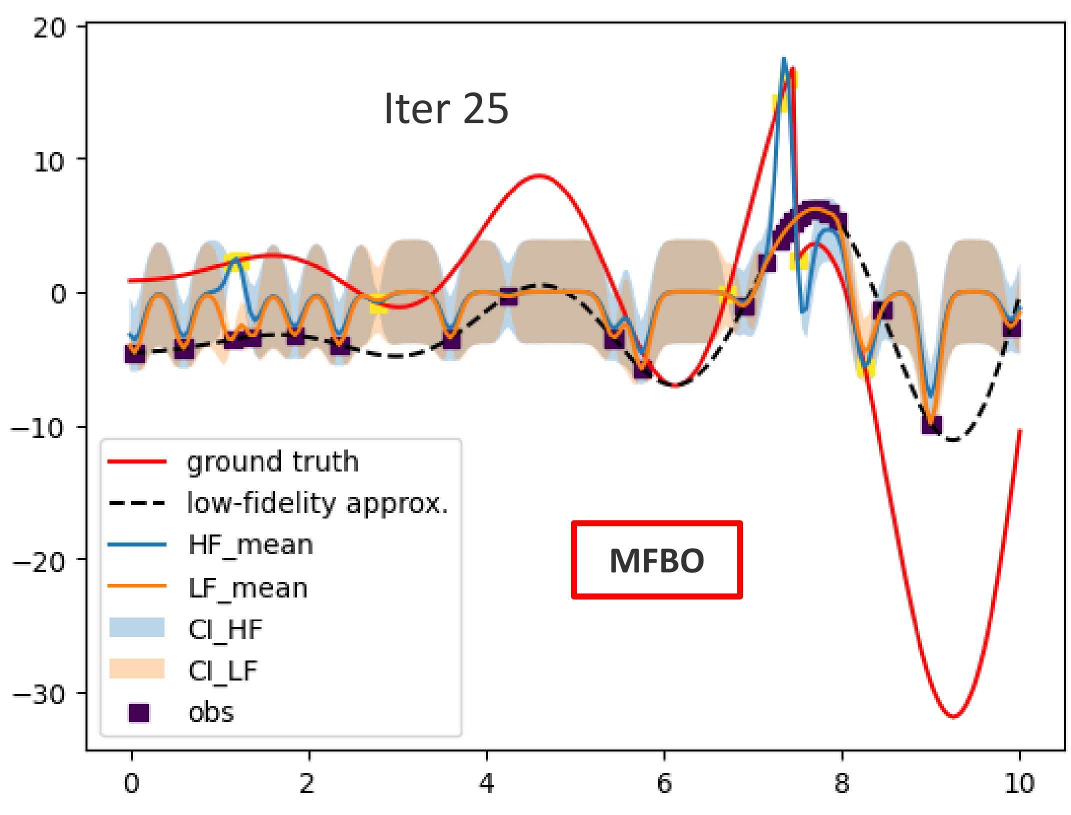
$$\delta > 0$$

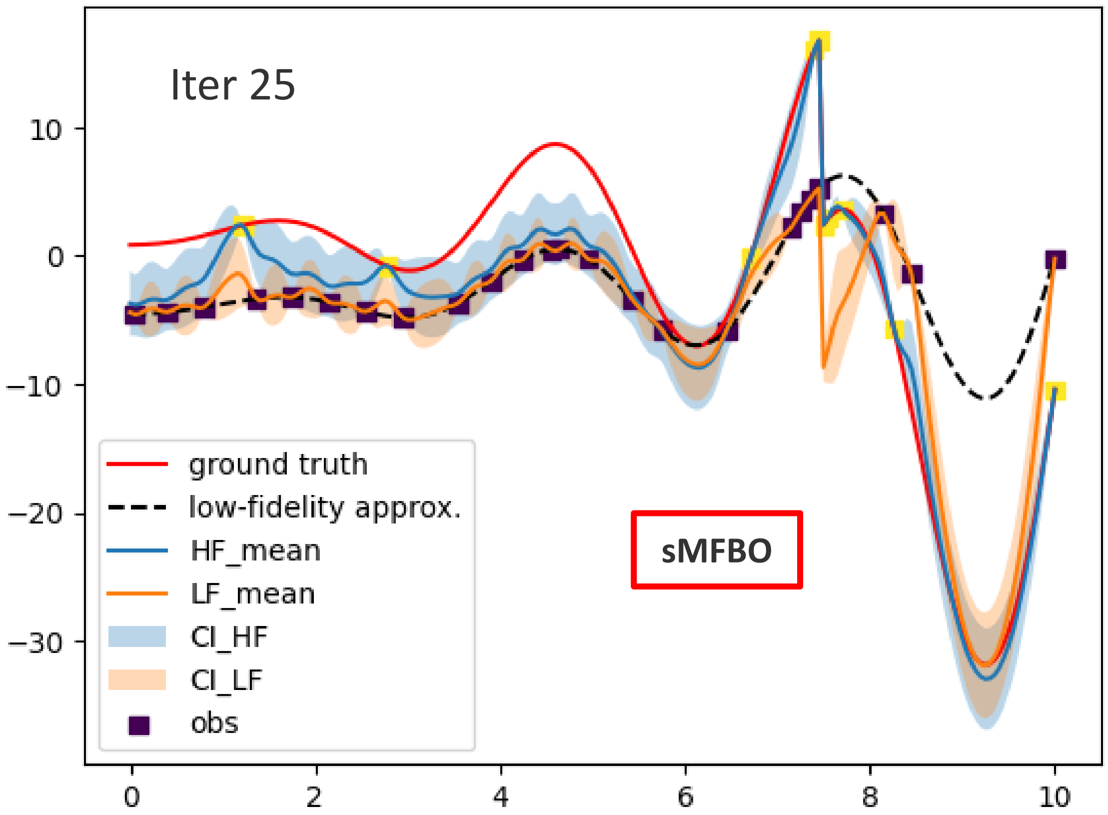
hyperparameter

Structured Multifidelity GP



Structured Multifidelity GP

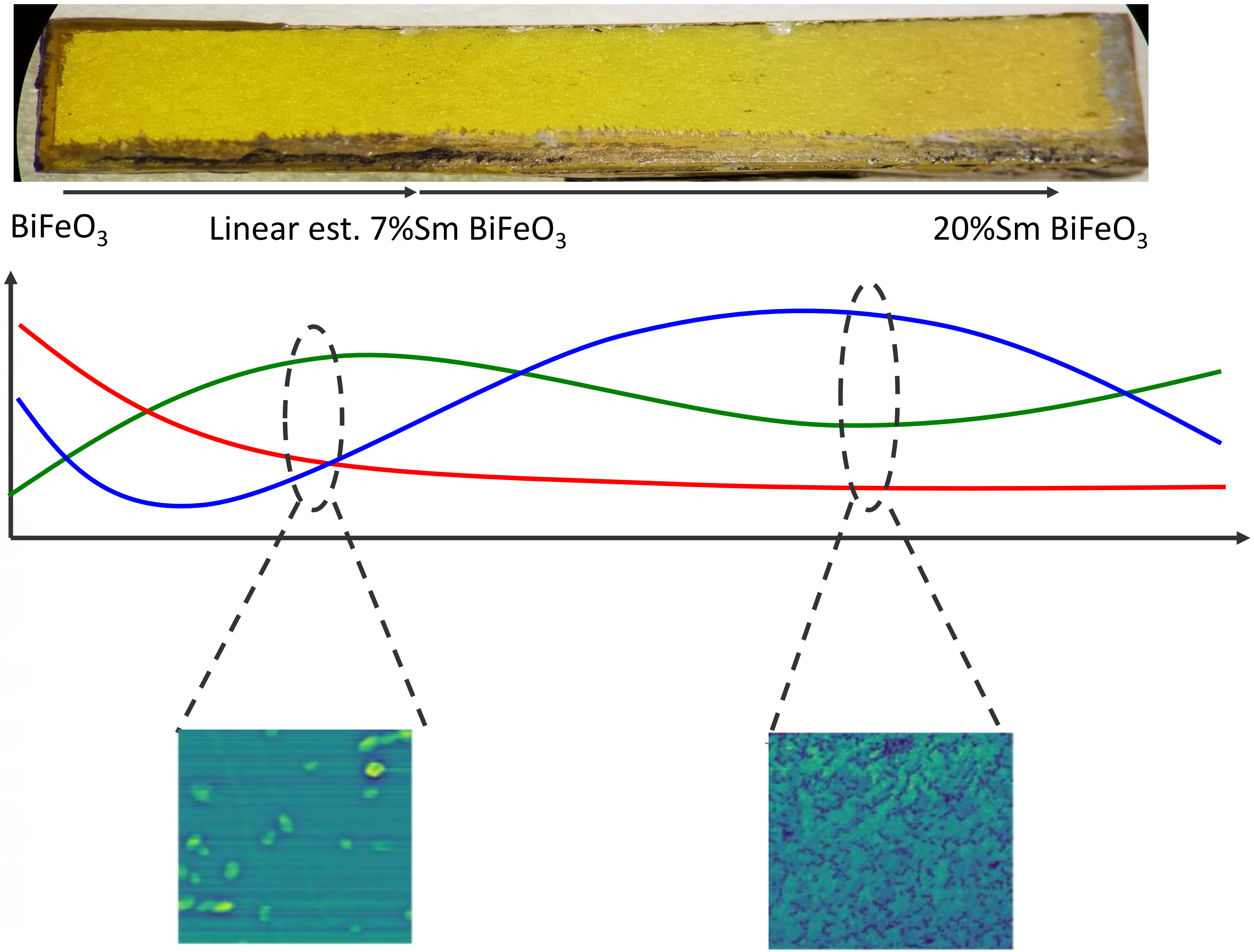




1. Multifidelity **structured** GP:
   - We have the easy to evaluate function with probabilistic model and expensive to evaluate function
   - The easy function is a proxy for expensive one and has some correlative relationship to it
   - We create policy that balances evaluation costs

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))

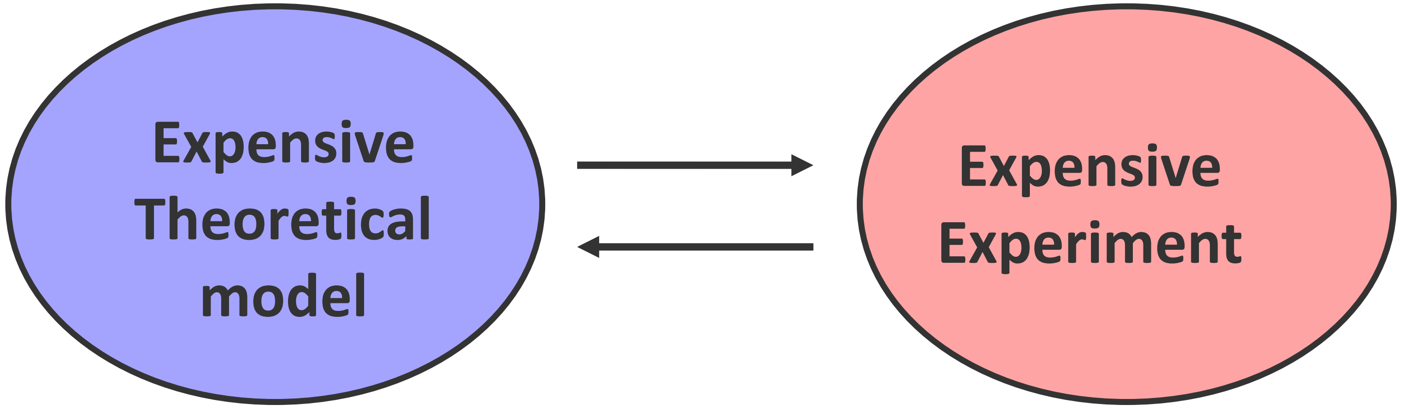
2. **Scenario II:** Active learning with one high dimensional imaging/spectroscopy method.
   a. Normal GP/sGP/HL if measured property is scalar (if we have good scalarizer for image/spectra)
   b. If it is active learning on images/spectra we do not have way to do it.

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).

Next step: co-navigation theory and experiment space via human supervised policies



1. Human-in-the-loop automated experiment
2. Policy tuning:
   - Exploration-exploitation balance
   - Fidelity of theory
   - Local physical model