THE UNIVERSITY of TENNESSEE KNOXVILLE

Deep Kernel Learning – II: Experiment, Explainability, and Human in the loop Intervention

Sergei V. Kalinin

What have we learned from lectures on GP/BO

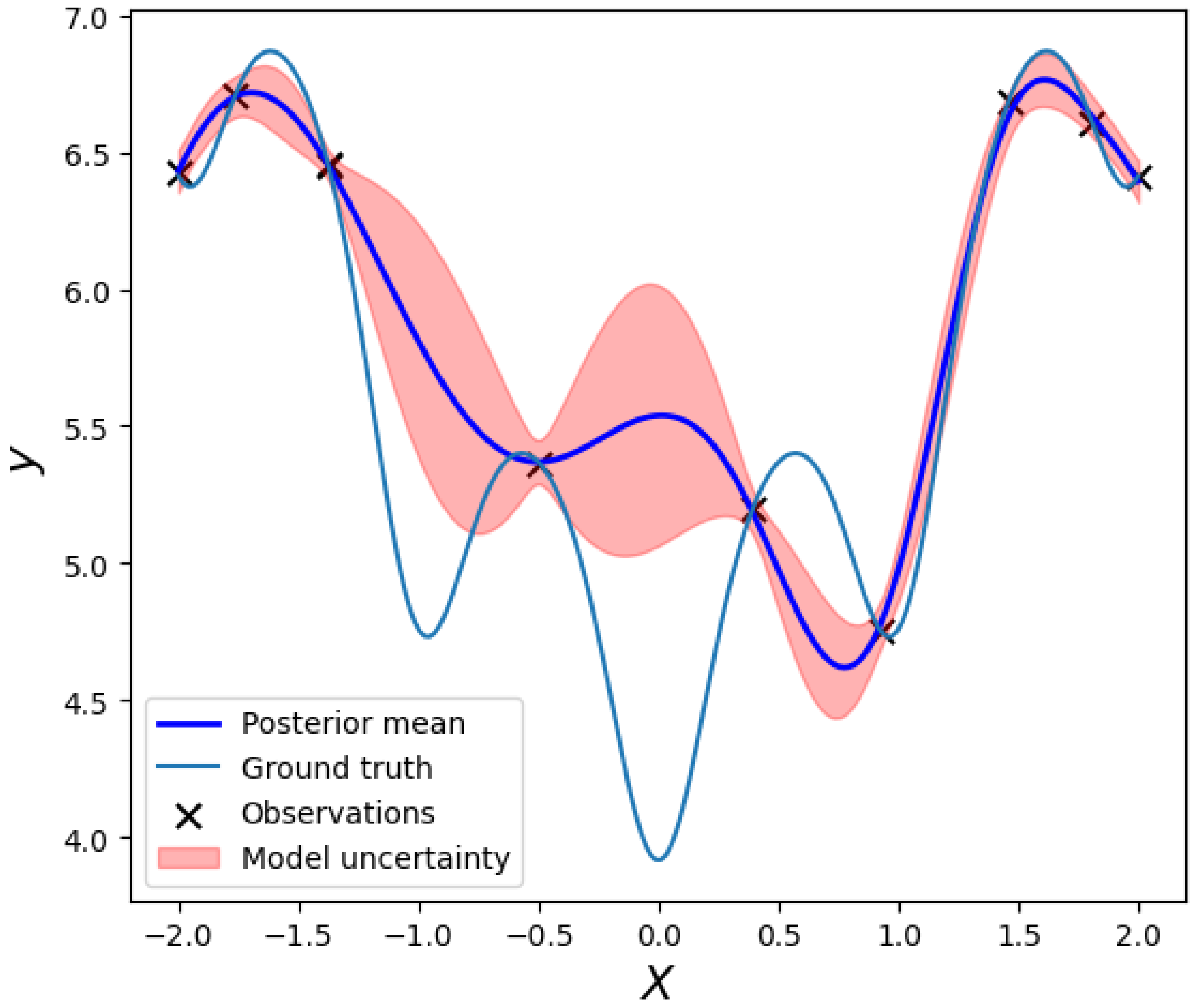
- Gaussian Process

- Kernel and kernel parameters
- Kernel Priors
- Noise Priors
- Mean function and priors

- Posteriors
- Bayesian Inference

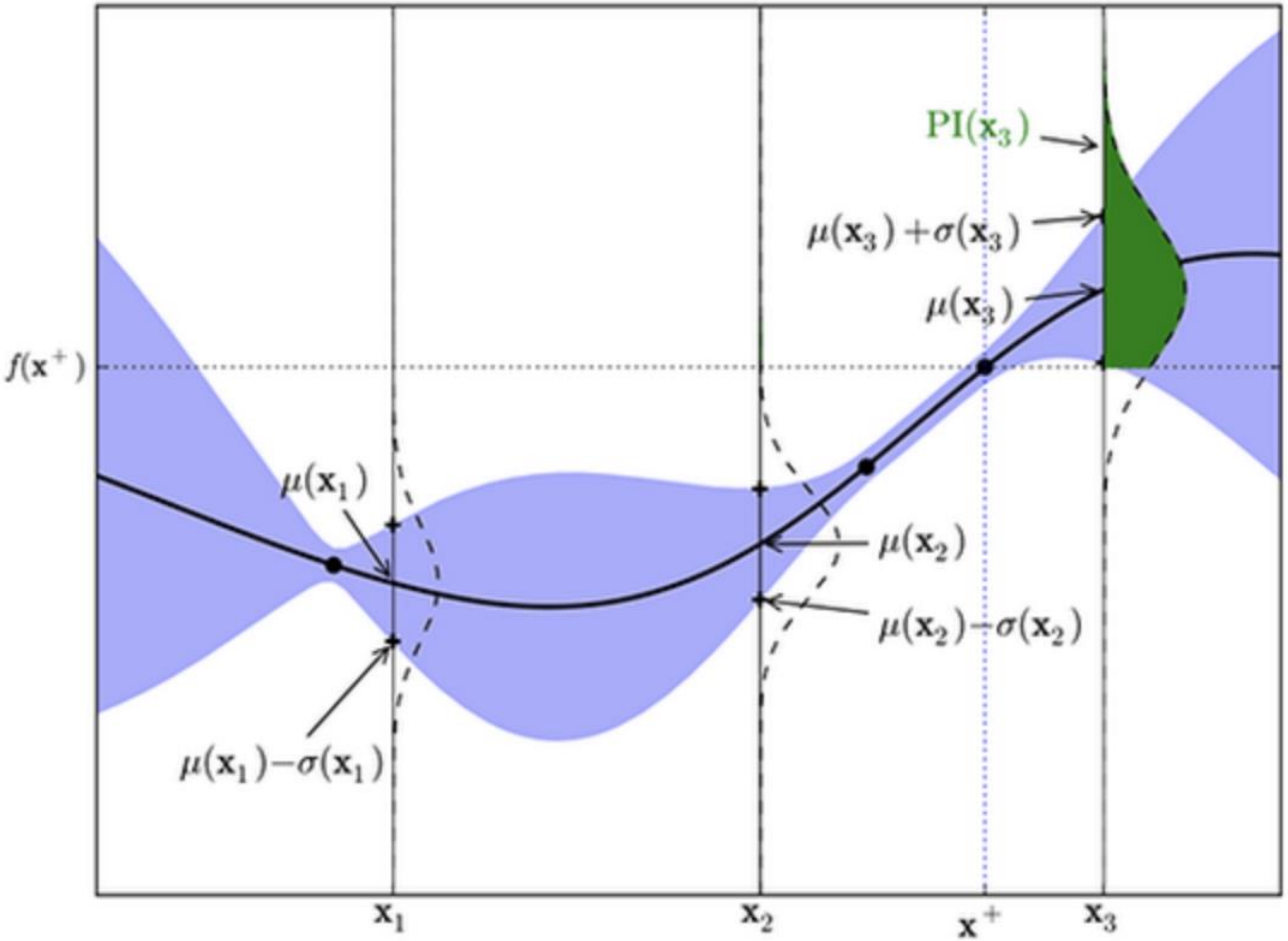
- Bayesian optimization
- Acquisition function

Gaussian Process and Bayesian Optimization



- We have some measurements in space X, and we want to maximize some property f(X).
- We create surrogate model: function and uncertainty based on measurements
- Gaussian Process: purely data driven
- Bayesian Inference: known model and some idea on parameters
- Structured Gaussian Process: physics-derived mean function

Acquisition Functions (Policies)



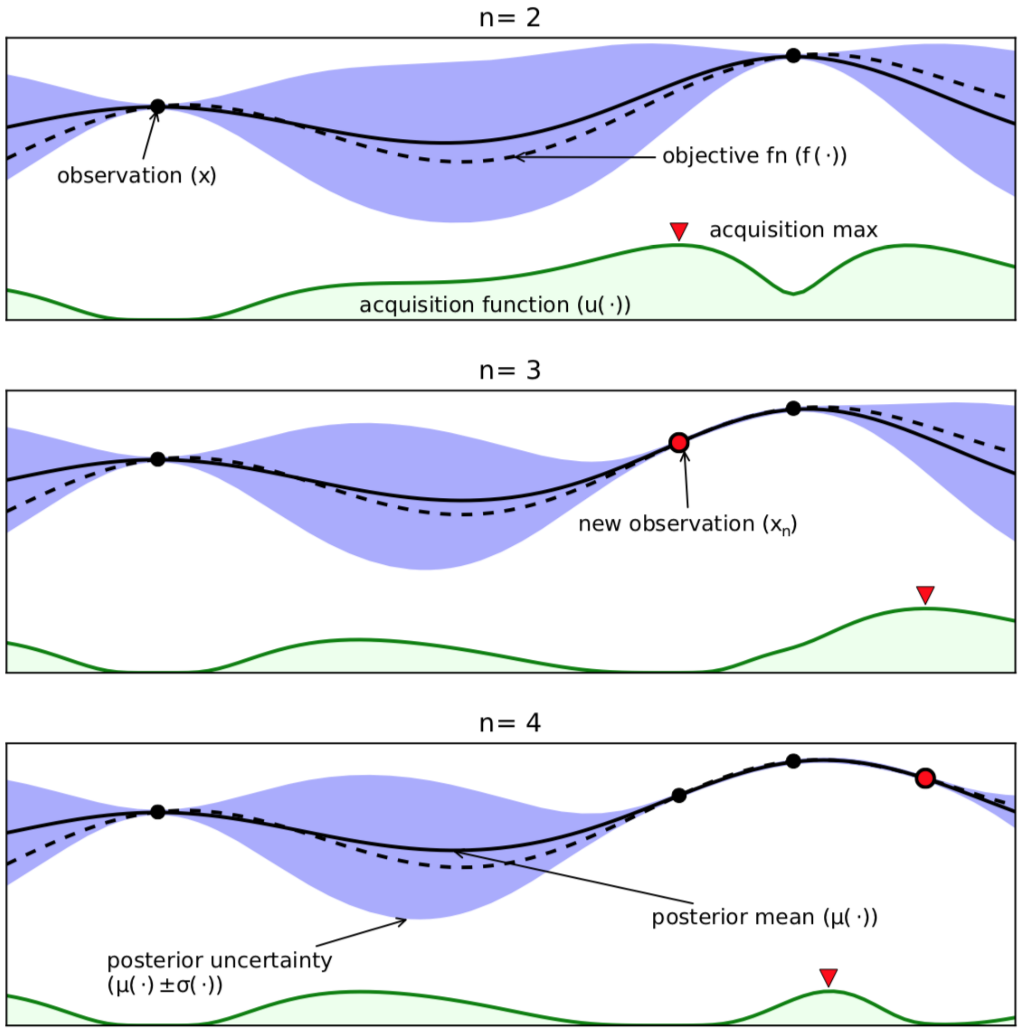
1. **Upper confidence bound:** simplest possible - just take the upper confidence bound from the prediction

2. **Probability of Improvement:** Integral from current functional maximum to upper limit of distribution as test point

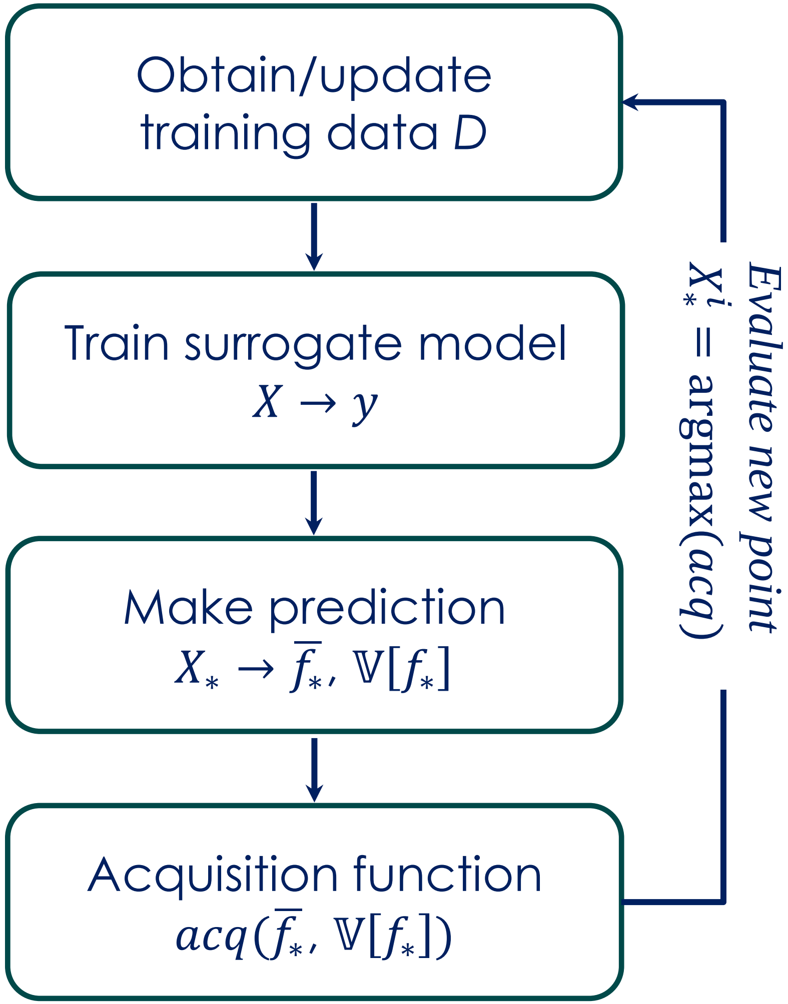
3. **Expected Improvement:** Instead of probability of improvement, we want to maximize the expected increase in the function value

4. **There are (always) more…**

Bayesian Optimization



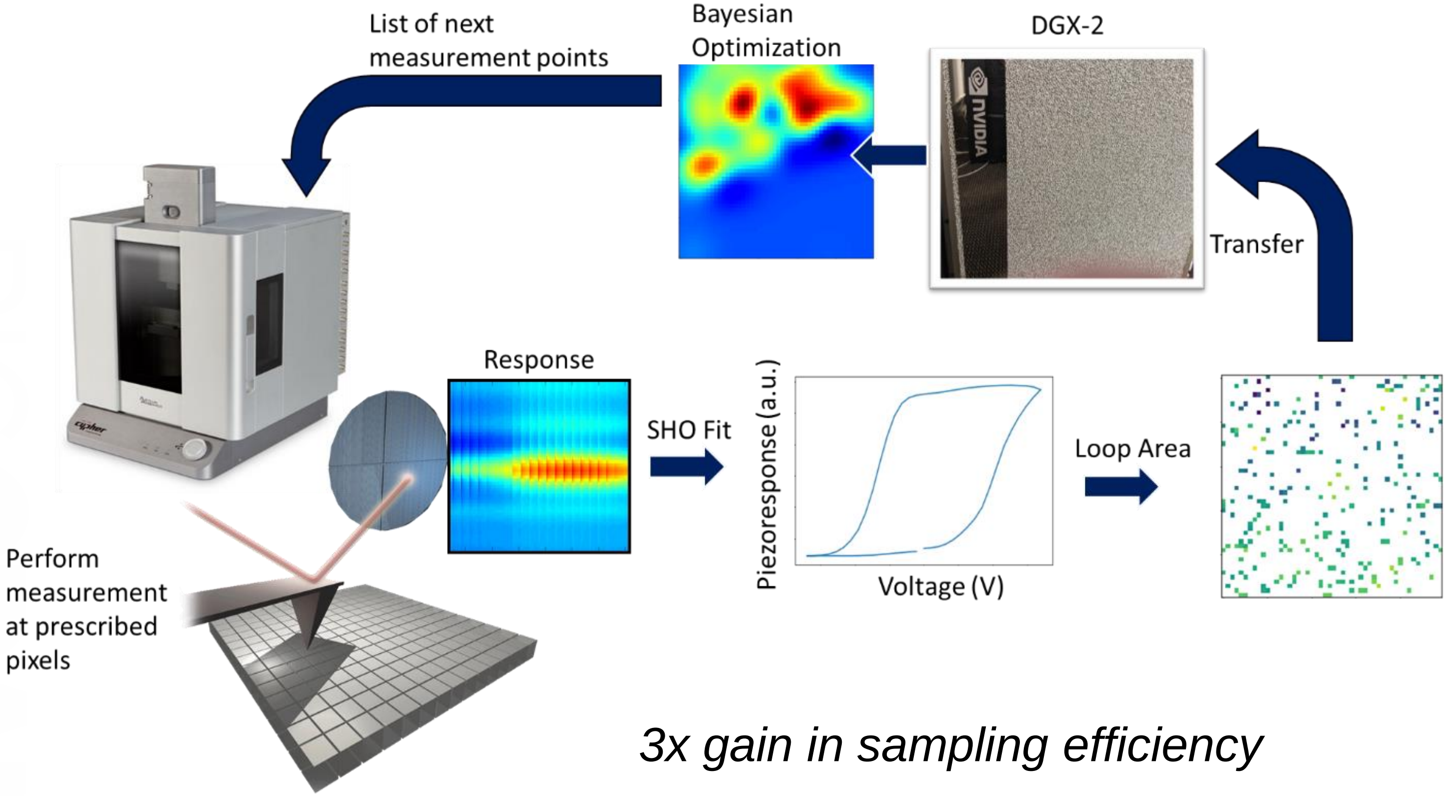
X, y: (sparse) Training data
X_*: New (not yet evaluated) points



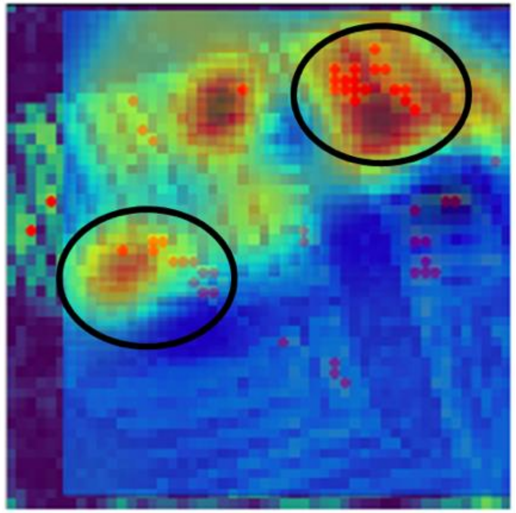
N. de Freitas et al., Taking the Human Out of the Loop: A Review of Bayesian Optimization , *Proceedings of the IEEE* **104**, 148 (2015)

BO for Self-Driving Microscope

First implementation of self-driving microscope: 2020

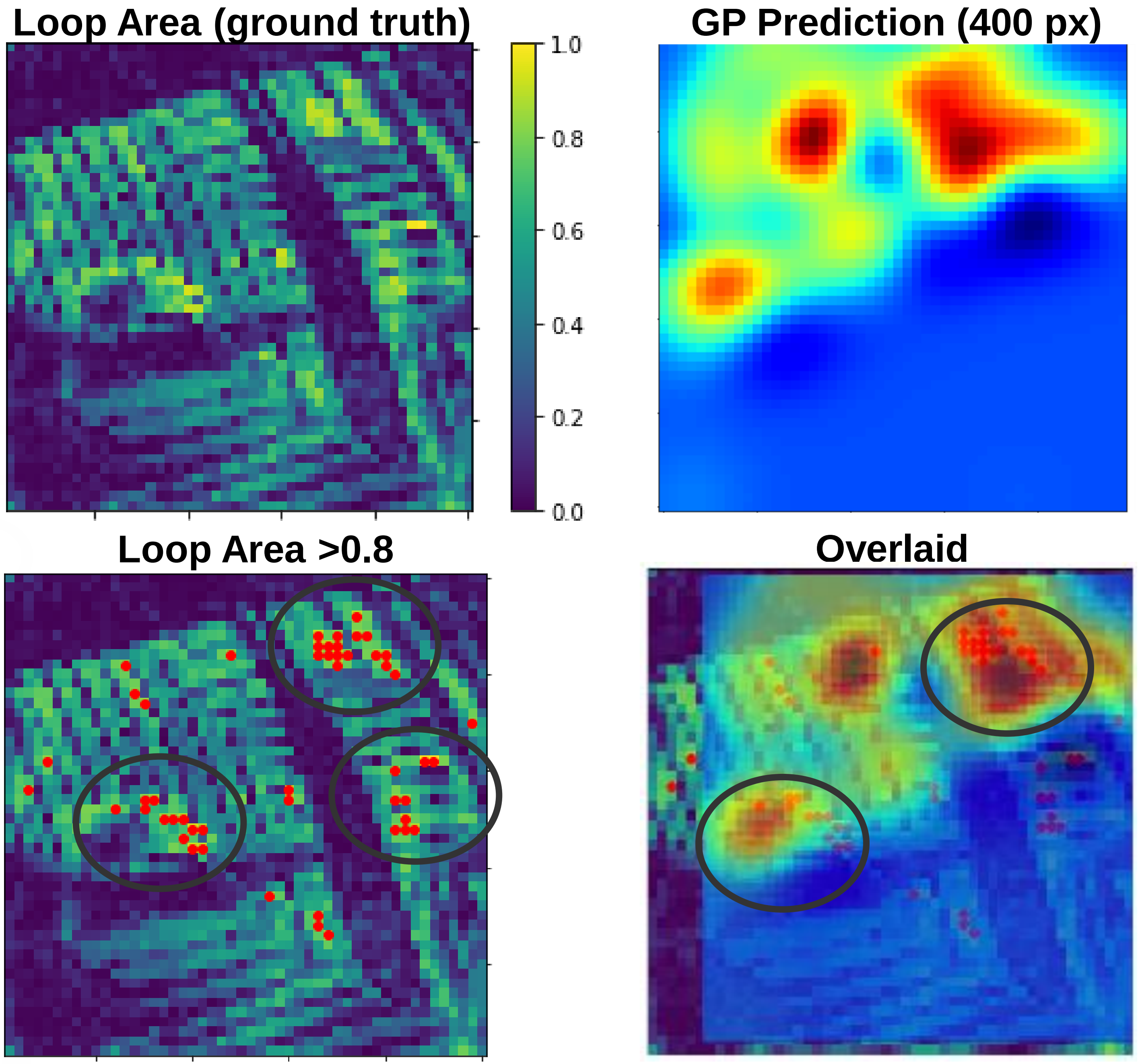


Comparison with "ground truth"



3x gain in sampling efficiency

R. K. Vasudevan, K. Kelley, H. Funakubo, S. Jesse, S. V. Kalinin, M. Ziatdinov, ***ACS Nano*** *(2021) https://doi.org/10.1021/acsnano.0c10239*

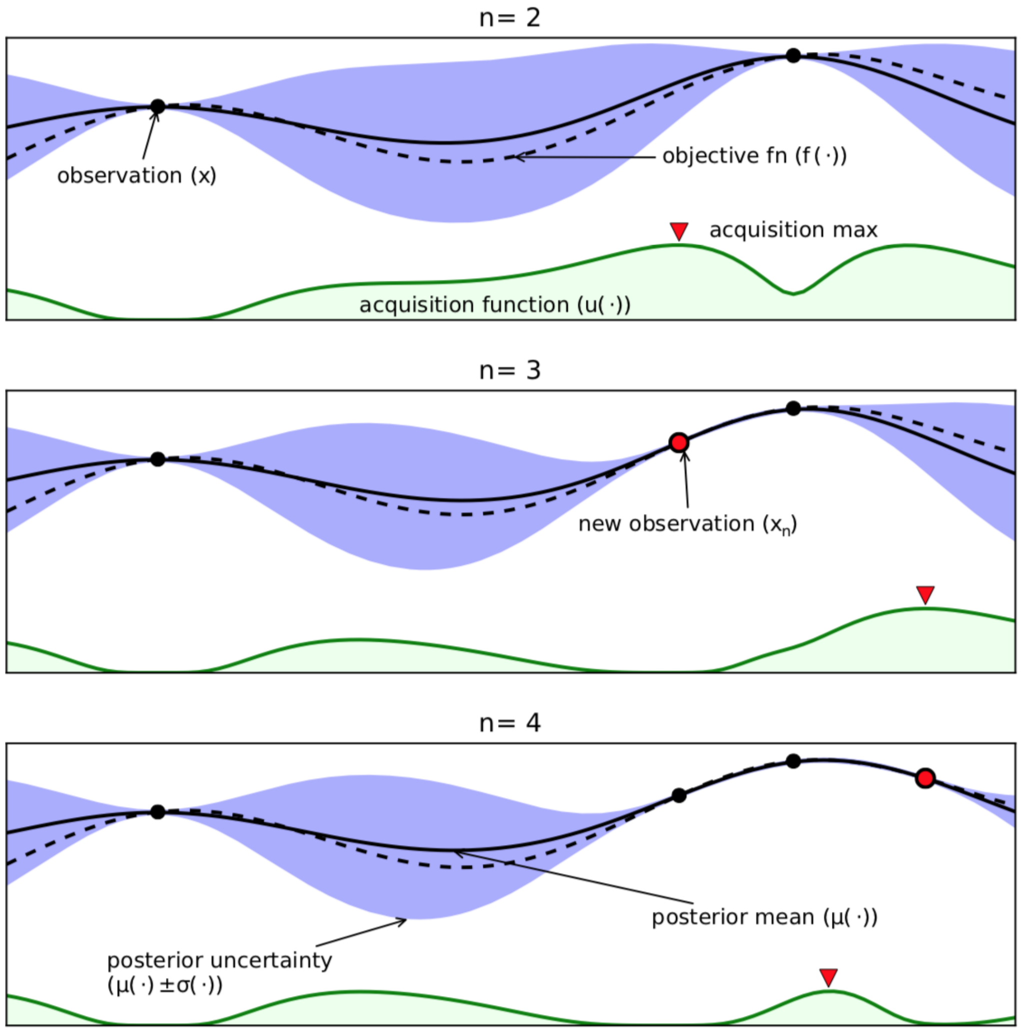


[arXiv:2103.12165]()

[arXiv:2011.13050]()

The application of simple data-driven GP for real world scenarios did not work particularly well.

Bayesian Optimization



X, y: (sparse) Training data
X_*: New (not yet evaluated) points

Obtain/update training data D

↓

Train surrogate model
$X \rightarrow y$

↓

Make prediction
$X_* \rightarrow \bar{f}_*, \mathbb{V}[f_*]$

↓

Acquisition function
$acq(\bar{f}_*, \mathbb{V}[f_*])$

Evaluate new point
$X_*^i = \mathrm{argmax}(acq)$

N. de Freitas et al., Taking the Human Out of the Loop: A Review of Bayesian Optimization , *Proceedings of the IEEE* **104**, 148 (2015)

What is the limitation of the GP/BO?

1. Works only in low-dimensional spaces

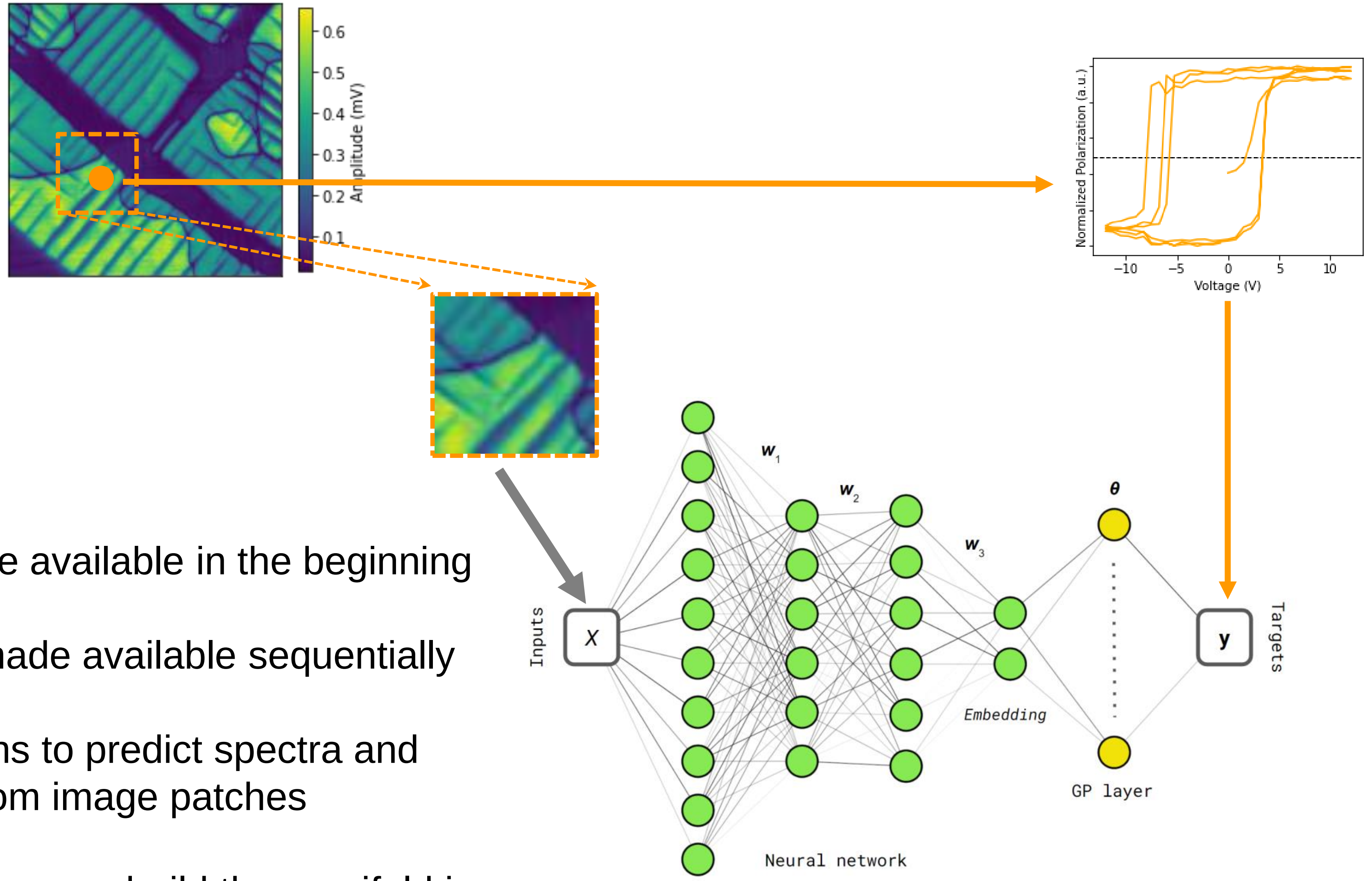
2. The correlations are defined by the kernel function (very limiting)

3. We do not use any knowledge about physics of the system

4. We do not use cheap information available during the experiment (proxies)

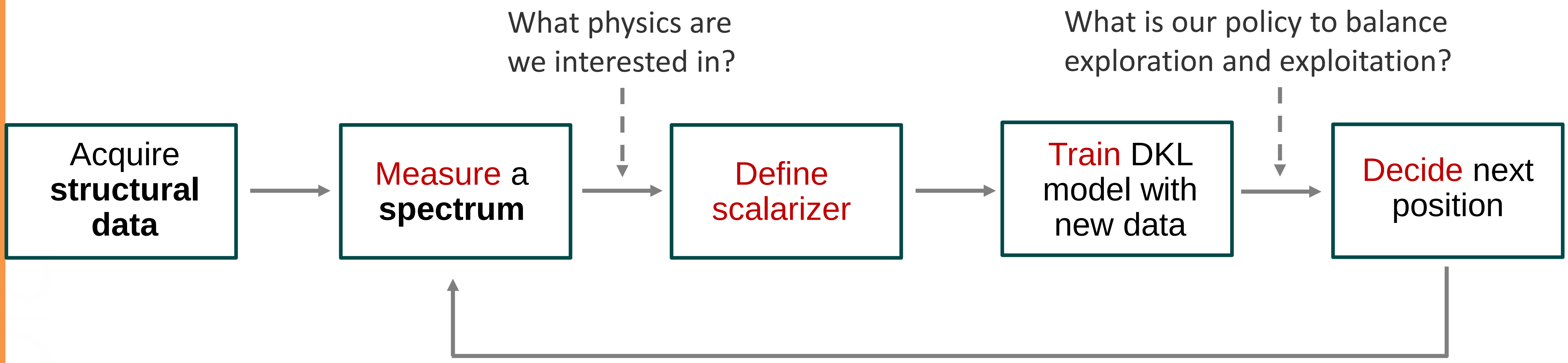
Can we somehow make high dimensional space low-D?

Deep Kernel Learning



- All patches are available in the beginning
- Spectra are made available sequentially
- The DKL learns to predict spectra and uncertainty from image patches
- Key aspect here: we build the manifold in latent space ***dynamically*** (unlike VAE)

Deep Kernel Learning based BO



Key concepts:

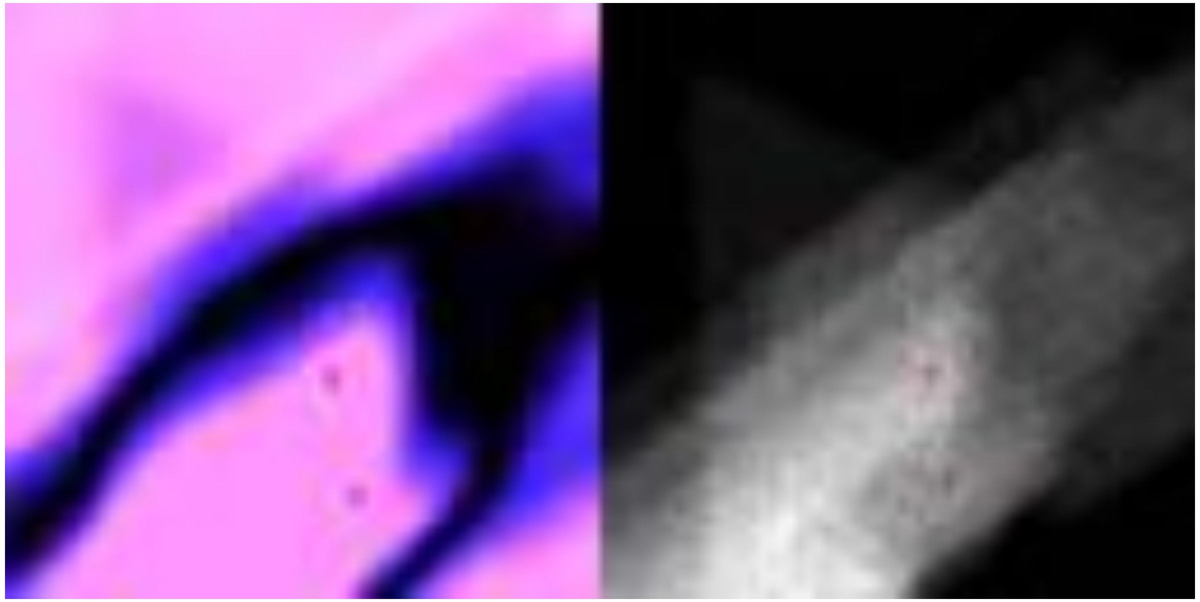
- **Scalarizer:** (any) function that transforms spectrum into measure of interest. Can be integration over interval, parameters of a peak fit, ration of peaks, or more complex analysis

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

Discovering Regions with Interesting Physics

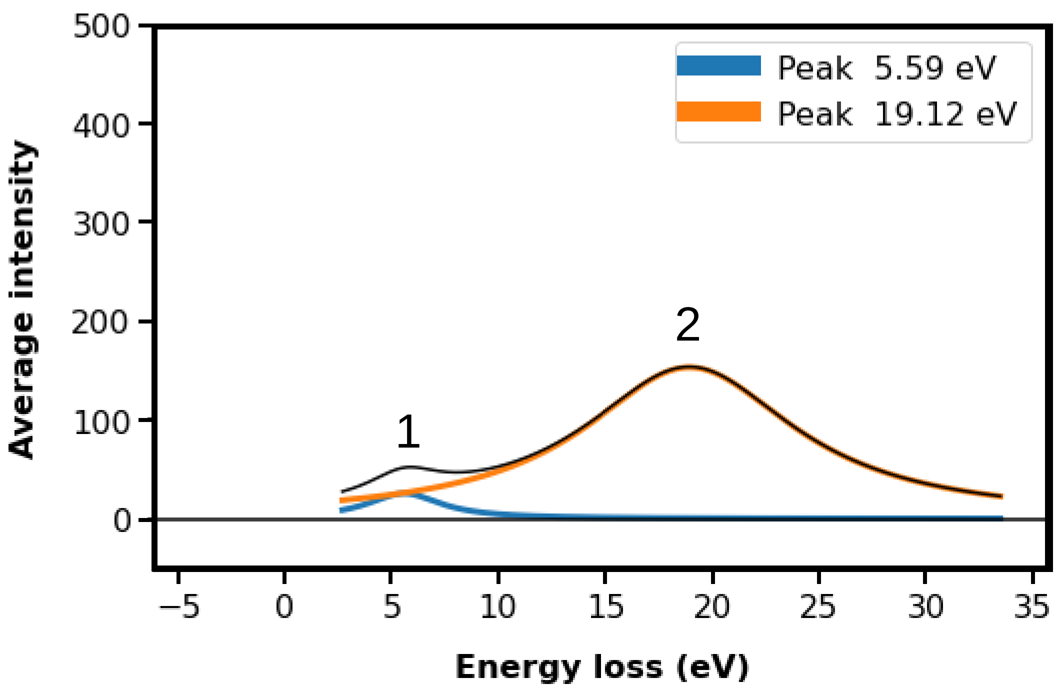
- **Discovering physics in a "new" material $MnPS_3$**
- **Curve fitting** to help enforce physical processes

"Acquisition function" HAADF-STEM

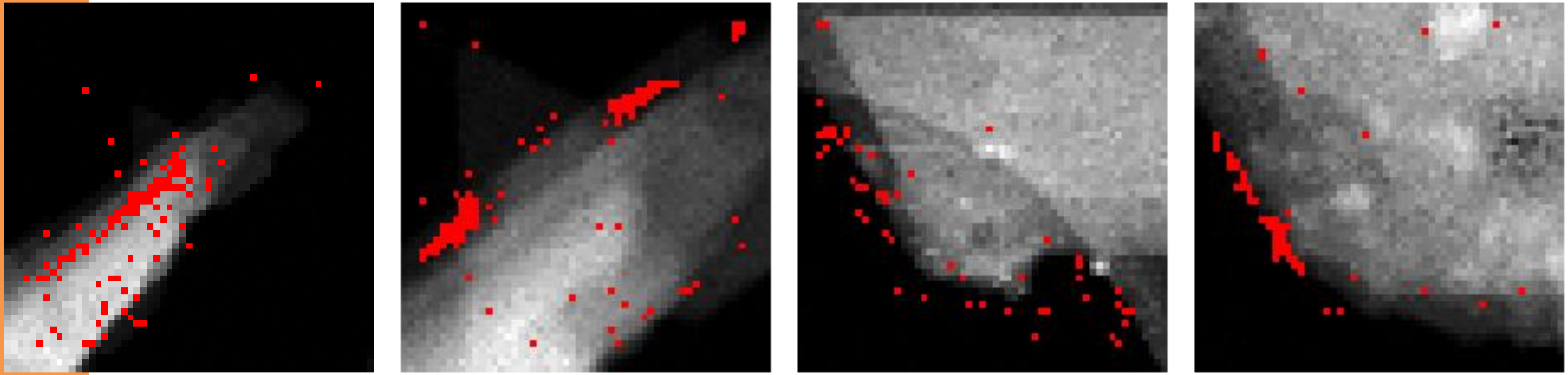


Physics search criteria:

$$Ratio = Peak\,1\ /\ peak\,2$$



More Examples of Physics Discovery



Discovery pathway depends on the reward structure (scalarizer that defines signature of physics we want to discover)!

Changing the Criterion

- (Same region) **Simple physics search:** peak max in selected region

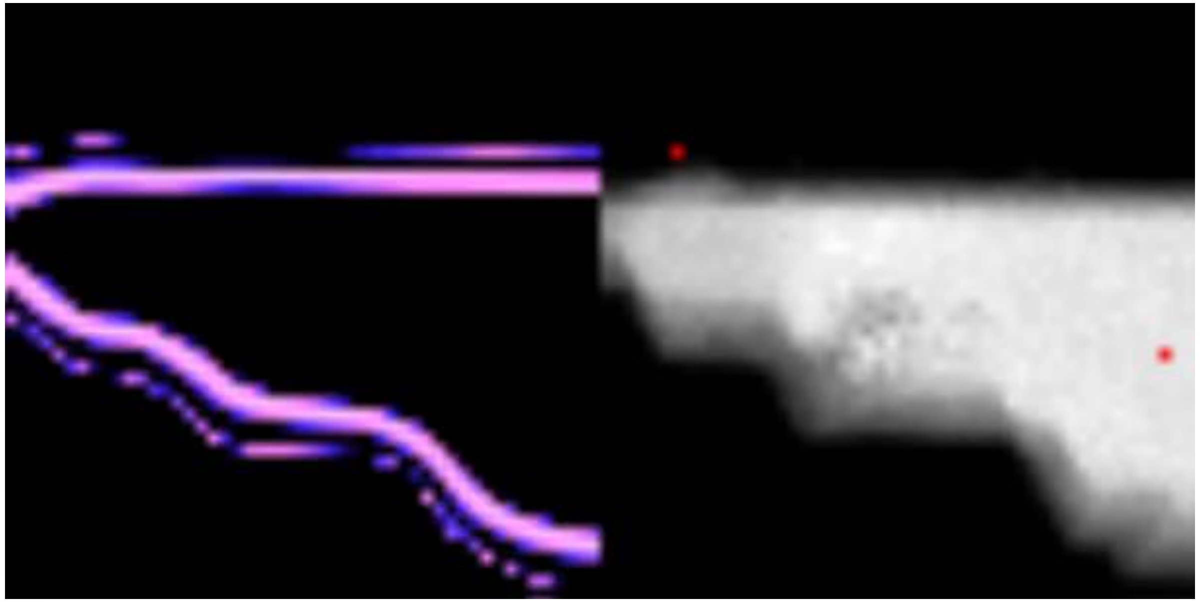
Physics search criteria:

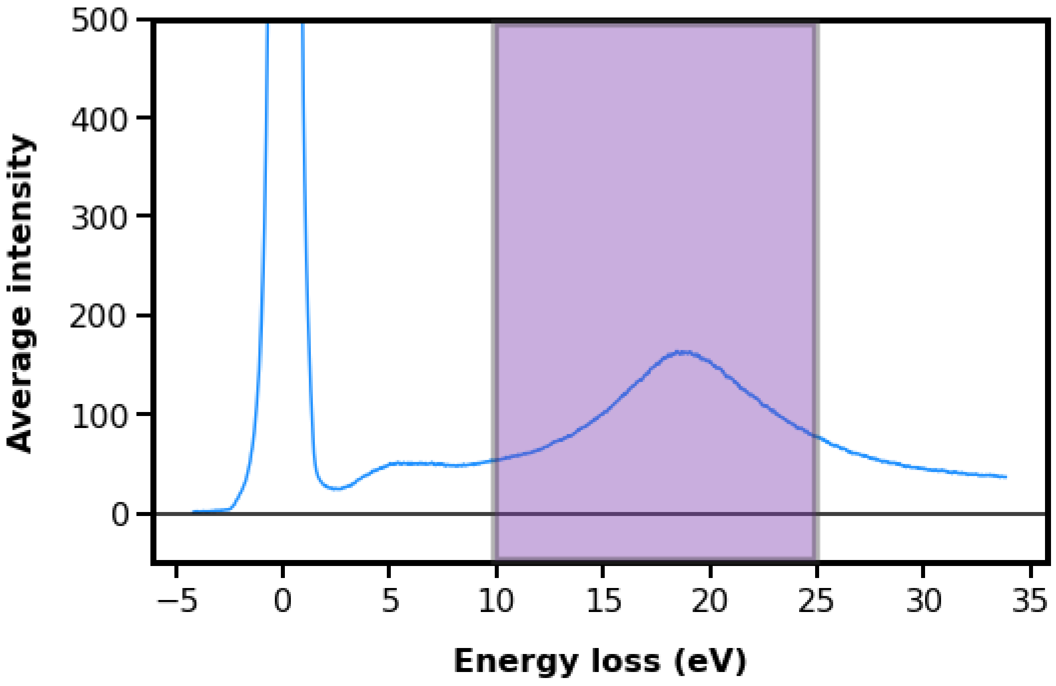
$$\boldsymbol{Maximize(f)}$$

(Specific peak intensity)

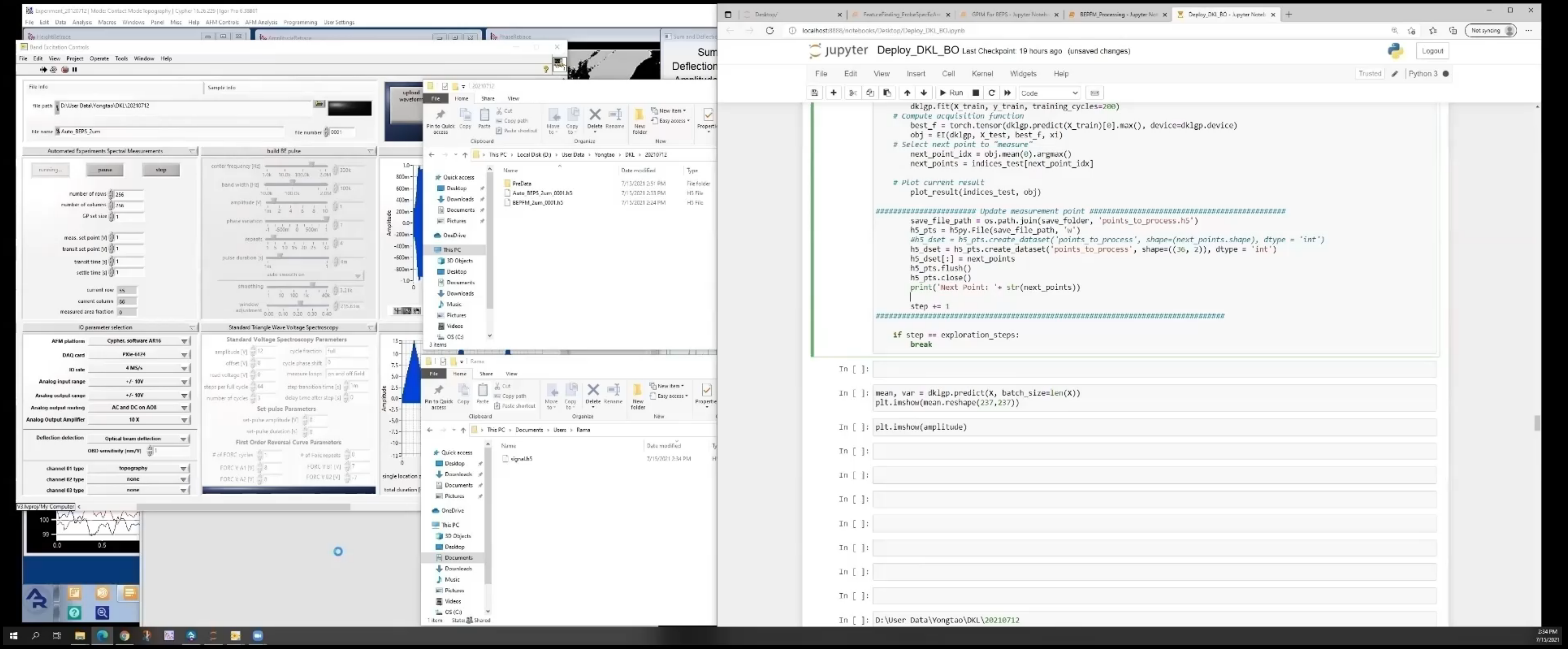
“Acquisition function”

HAADF-STEM
+ points visited



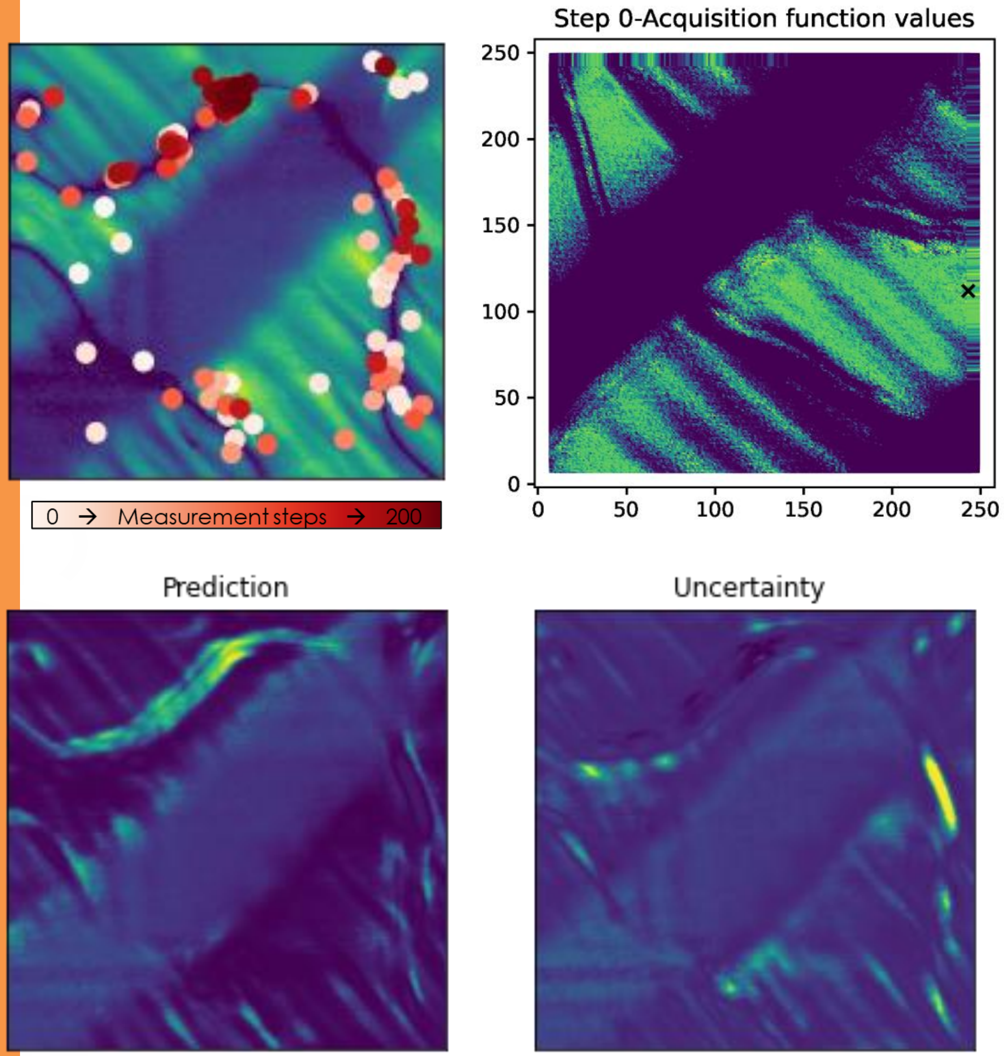


Deep Kernel Learning AE SPM

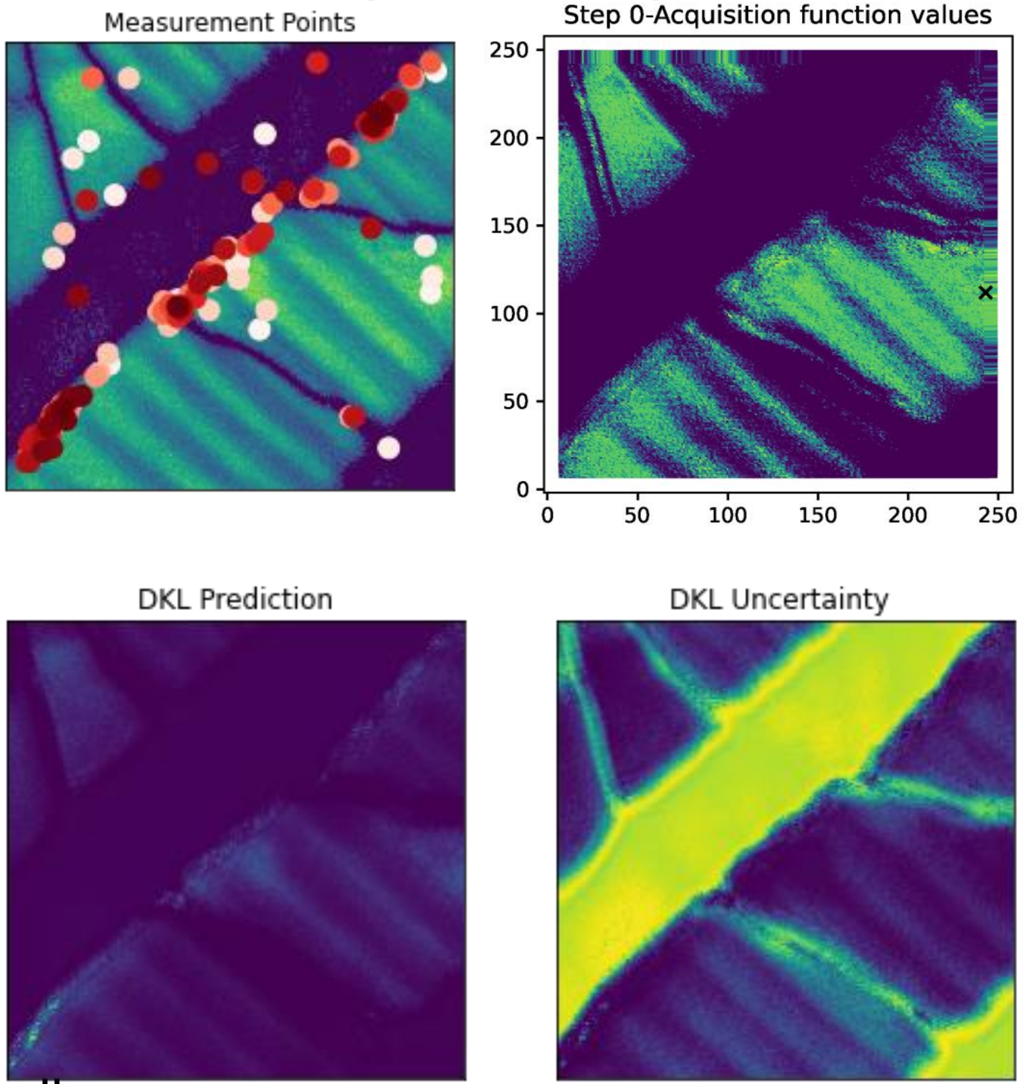


DKL SPM

Guided by: On field loop area



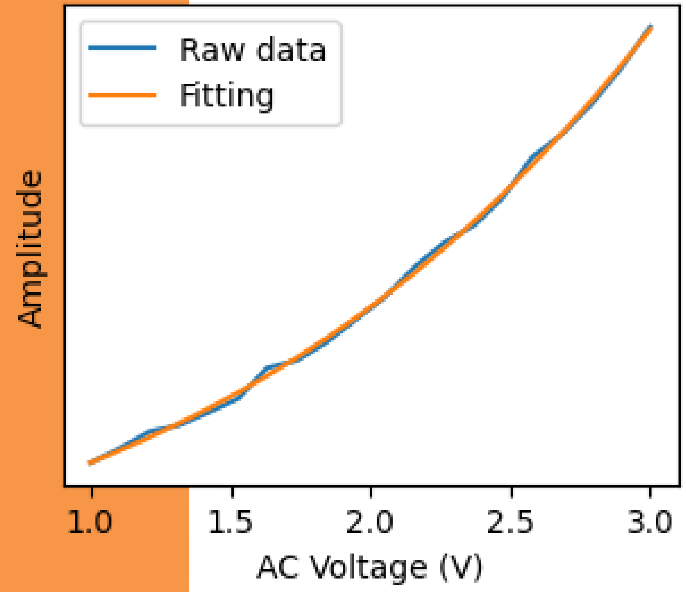
Guided by: Off field loop area



- ➢ Large loop opening corresponding 180º domain walls
- ➢ This behavior can be attributed to the large polarization mobility of 180º walls

Liu, Yongtao, et al, Nature Machine Intelligence 4, 4 (2022): 341-350.

Exploring Non-Linearity



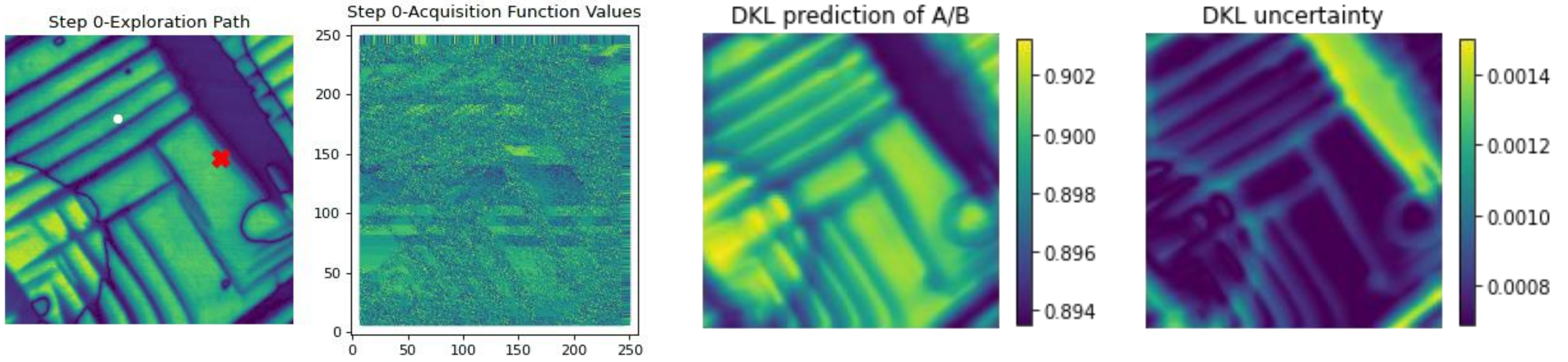
V_{AC} sweep curve at each location was fitted as $y = Ax^3 + Bx^2 + Cx$

A, B, C, and A/B were us as the target function to guide DKL-V_{AC} measurement.

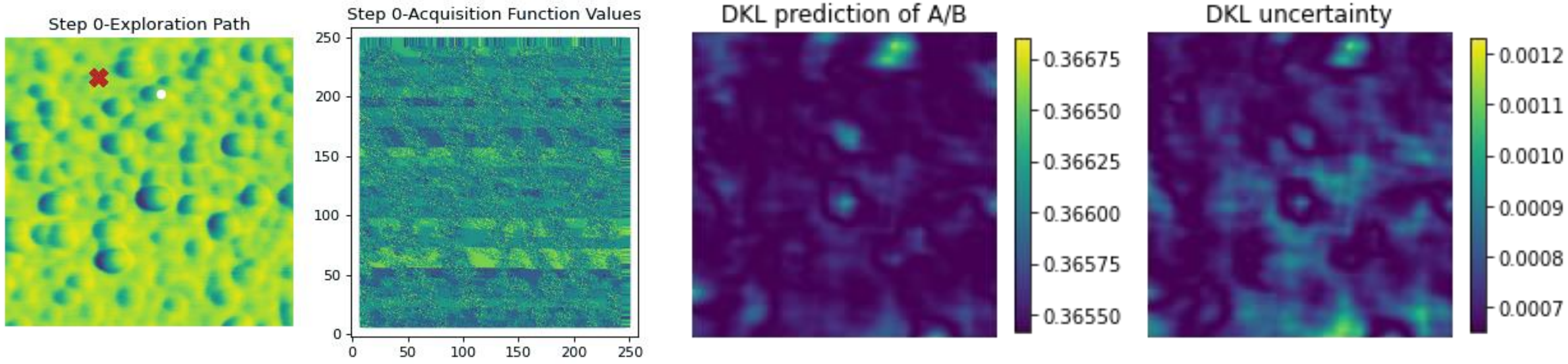
PTO and HZO thin films were studied.

- Shown are 200-step measurements of PTO and HZO thin films
- PFM amplitude was used as structure image; A/B was used to guide the measurement.

PTO experiment process and results



HZO experiment process and results



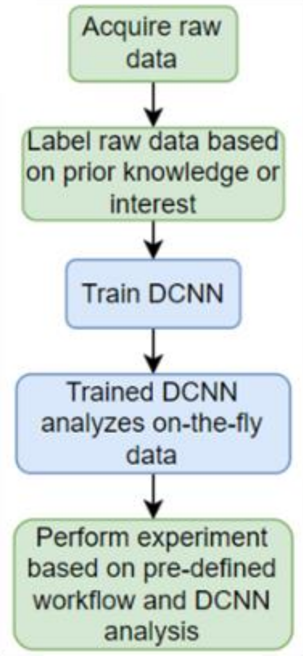
- In conventional microscopy experiment, human runs everything directly – defines scan, positions the probe, defines measurement parameters.

- In AE SPM, the policies are defined before the experiment and do not change. Sometimes it works – but not always.

- How would we:
  – (a) explain the AE progression after the experiment and
  – (b) control it during the experiment ?

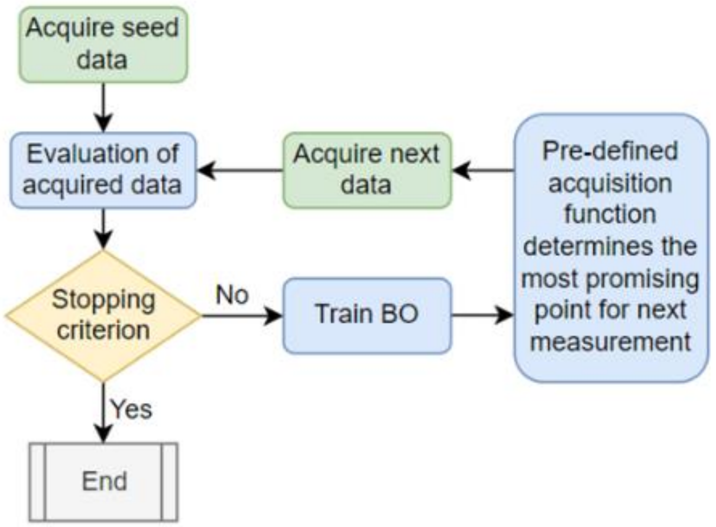
Types of automated experiment

Direct



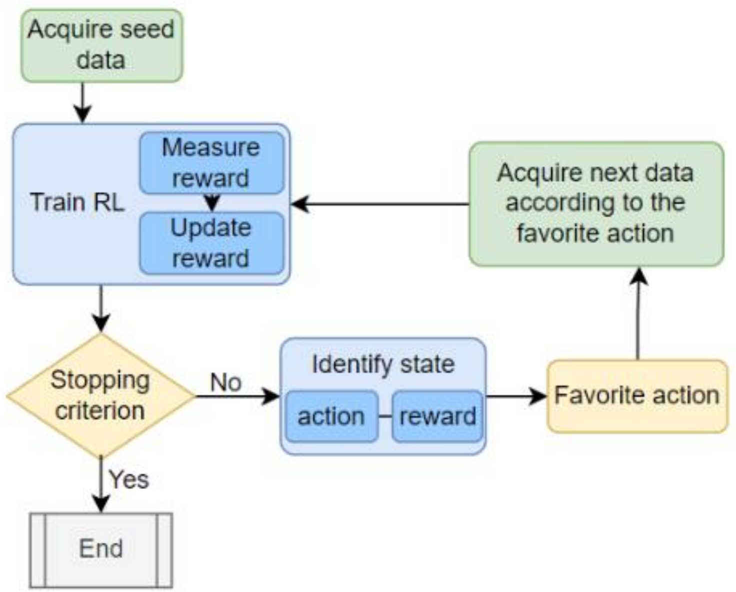
- Fixed policies
- Need DCNNs stable wrt. out of distribution shift

Myopic discovery



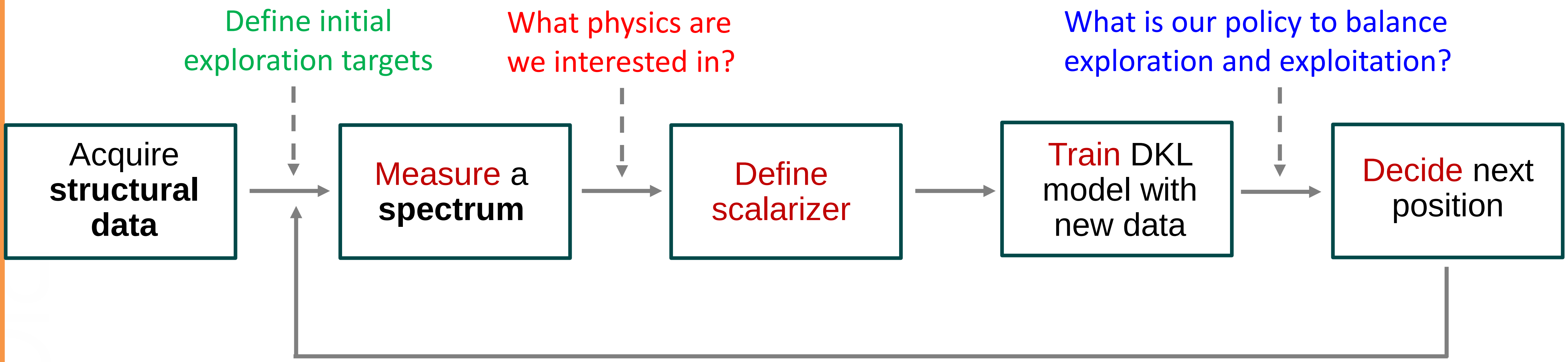
- Adjustable policies
- One step planning
- Can be implemented via Bayesian workflows
- Can be human in the loop

Multistage discovery



- Adjustable policies
- Multi-step planning
- Requires heuristic to start
- Requires **reward function**

Bringing Human into the Loop

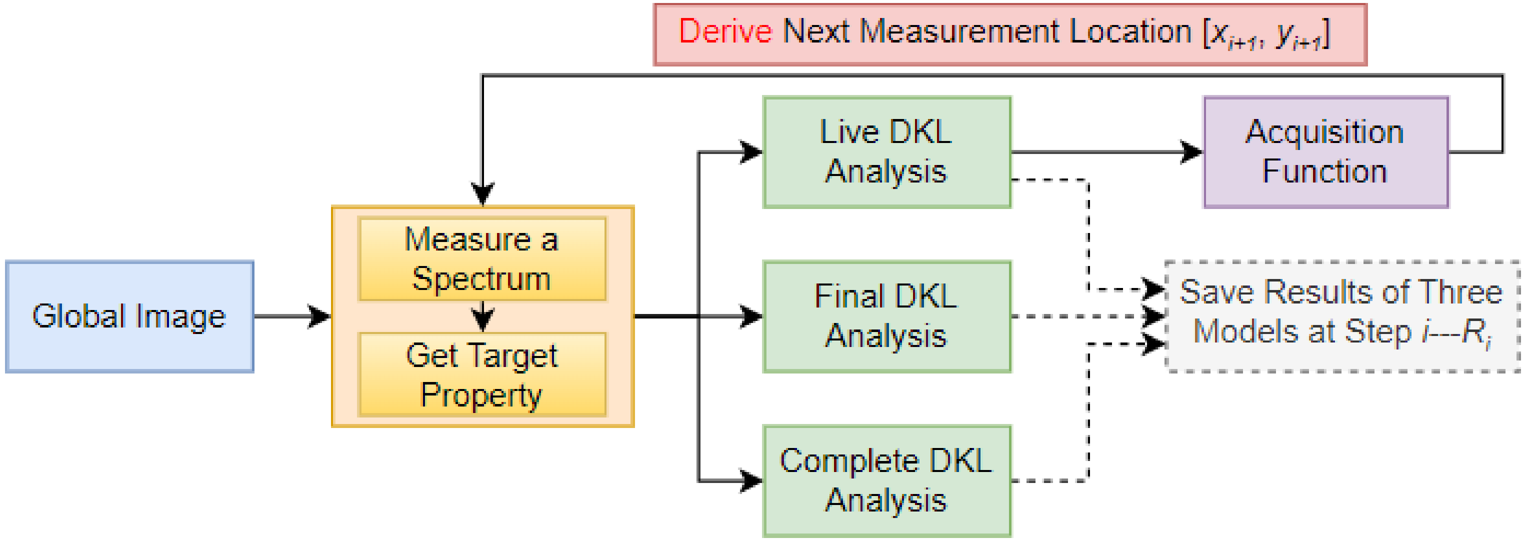


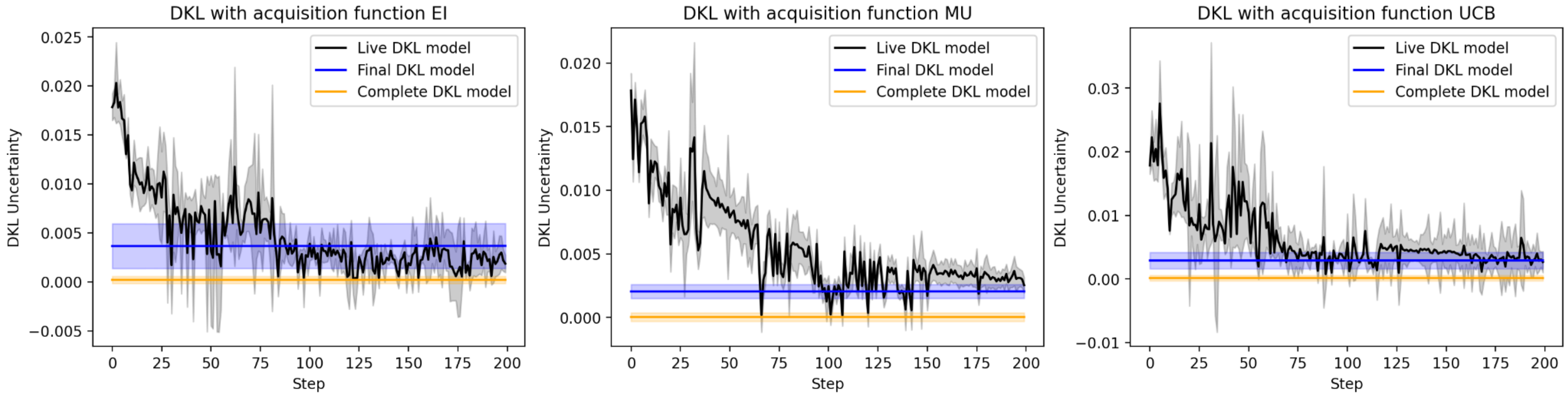
Key concepts:

- **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

AE Forensics

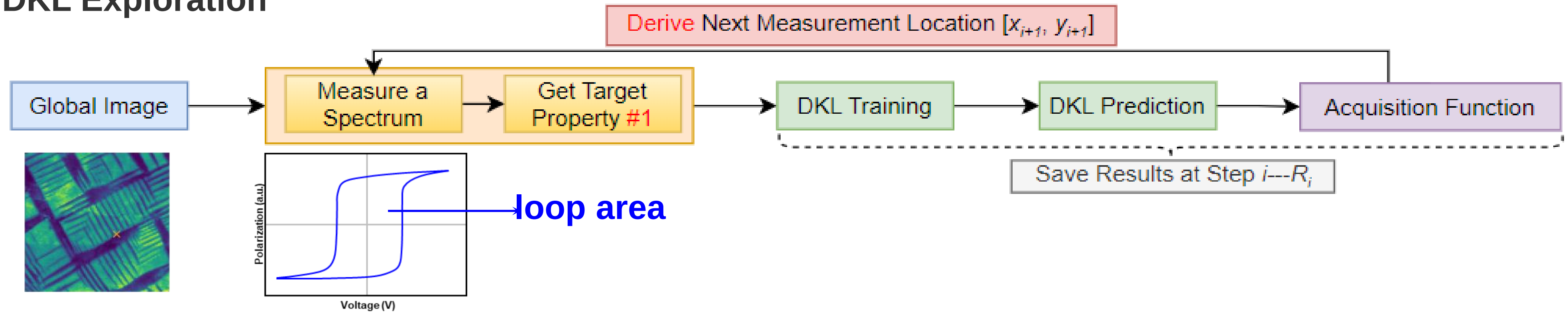
- During the AE, model learns structure-property relationships.
- What if we retrace the experimental steps – using the fully trained model?



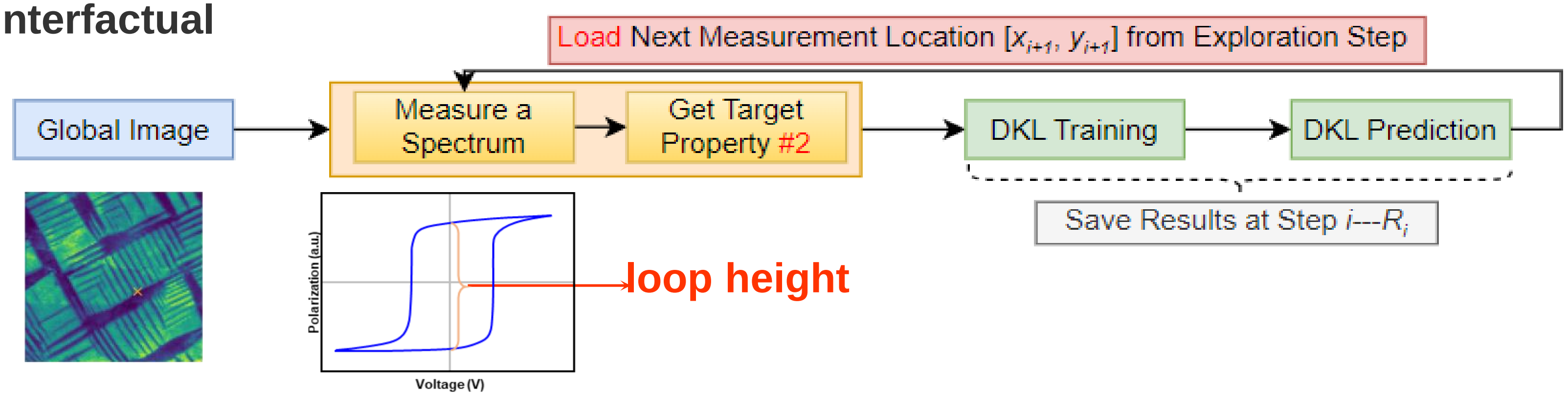


Counterfactual scalarizers

DKL Exploration



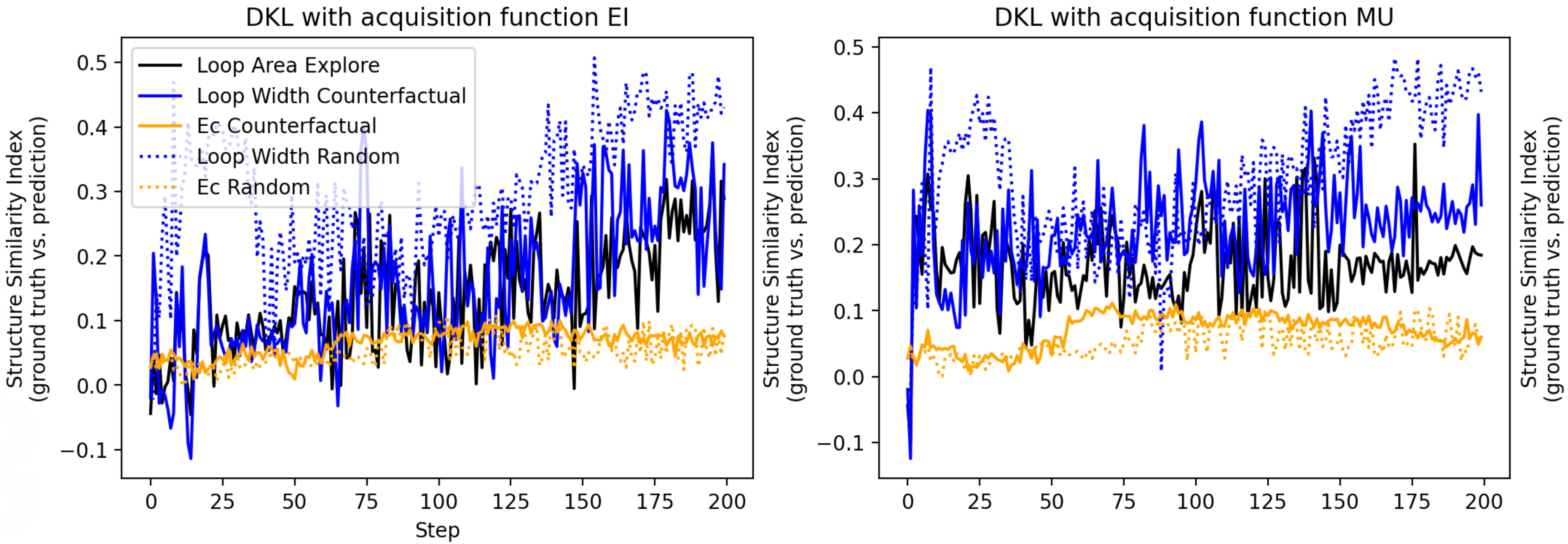
DKL Counterfactual



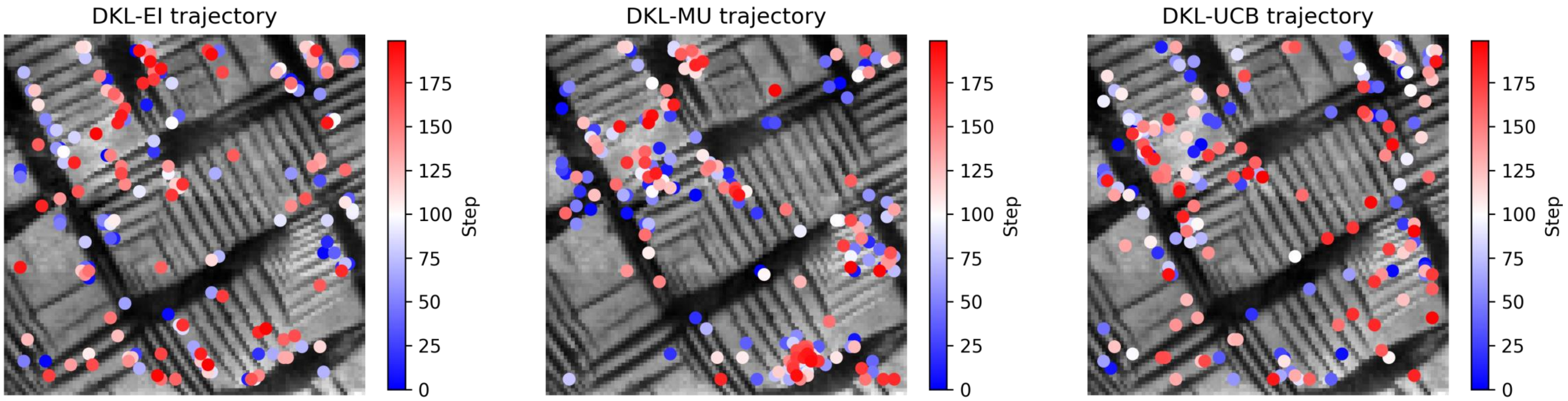
Target properties:

1. Loop Area
2. Loop Height
3. Coercive Field
4. …

- We save the full experimental trace
- What if we follow the actual experimental path – but calculate alternative (counterfactual) scalarizers?

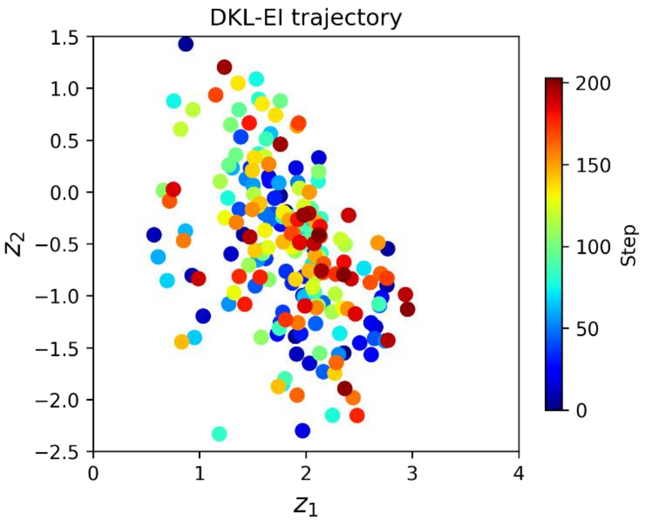

DKL with acquisition function EI
Loop Area Explore
Loop Width Counterfactual
Ec Counterfactual
Loop Width Random
Ec Random
Structure Similarity Index (ground truth vs. prediction)
Step
DKL with acquisition function MU
Structure Similarity Index (ground truth vs. prediction)

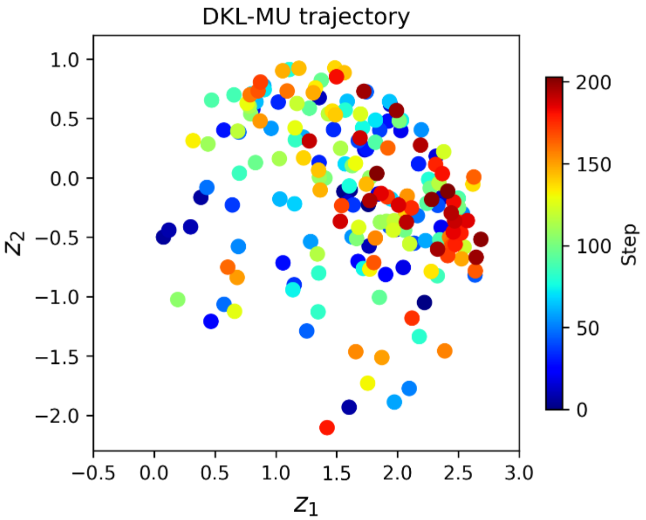
Monitoring the AE

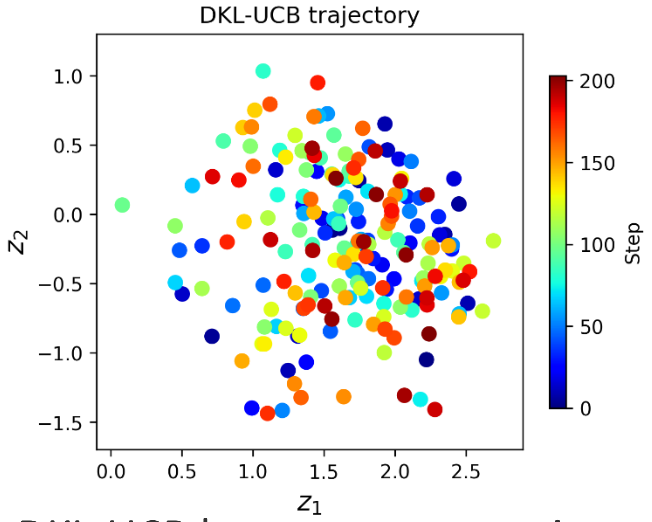


- Different acquisition functions (policies) give different experimental paths for AE
- Can we analyze what is special about points visited?

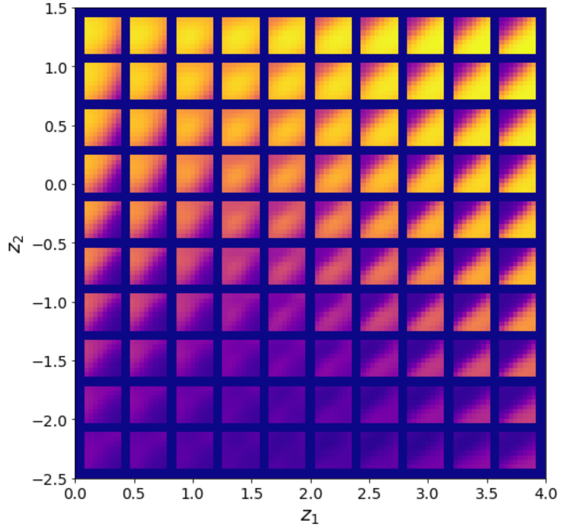
VAE approach: feature space of visited points



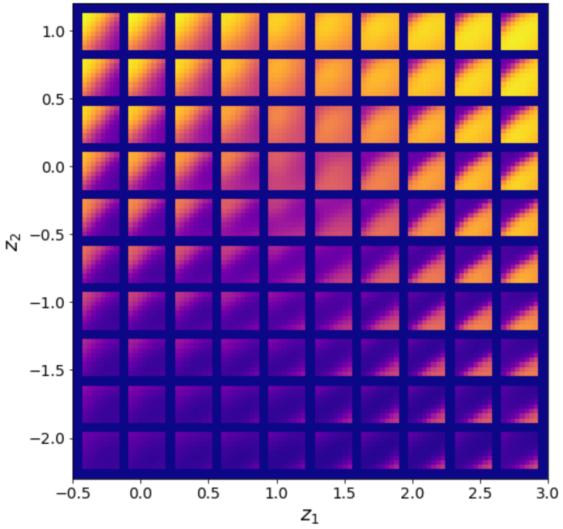




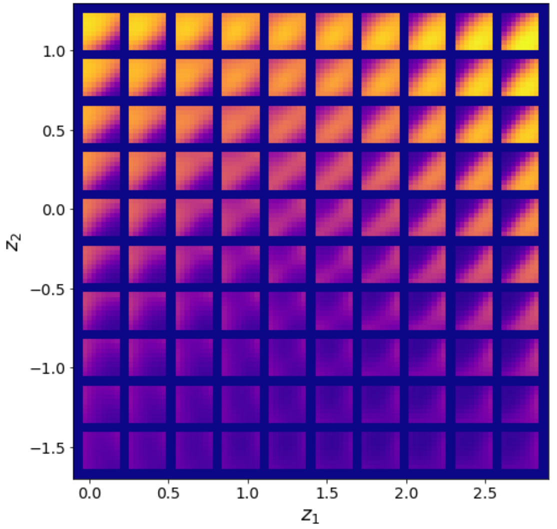
DKL-EI latent representation



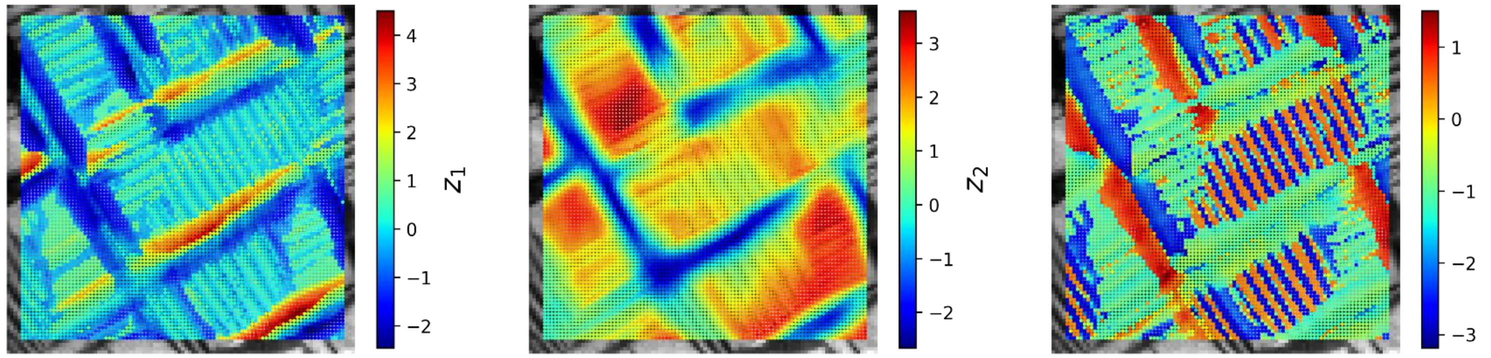
DKL-MU latent representation

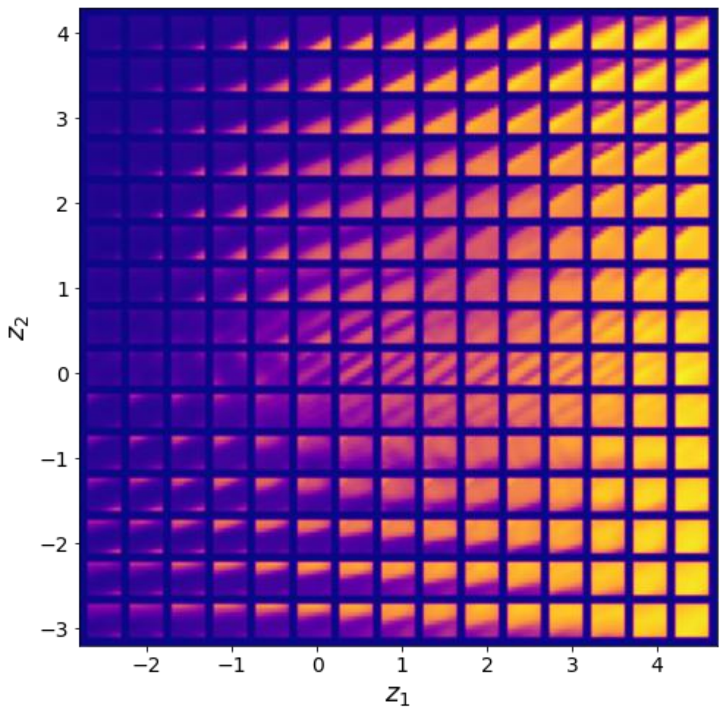


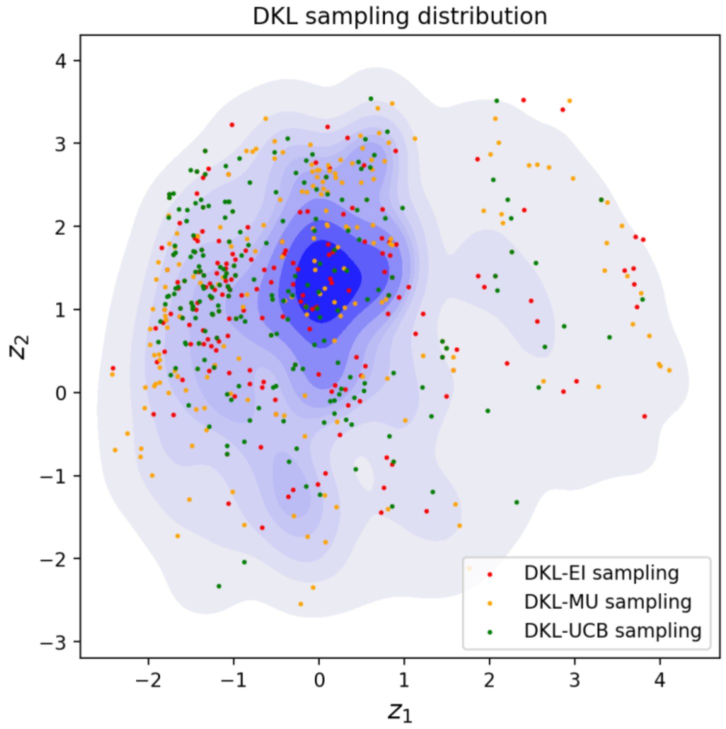
DKL-UCB latent representation



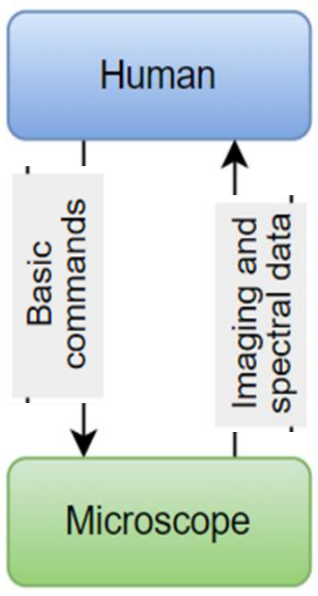
VAE approach: full feature space

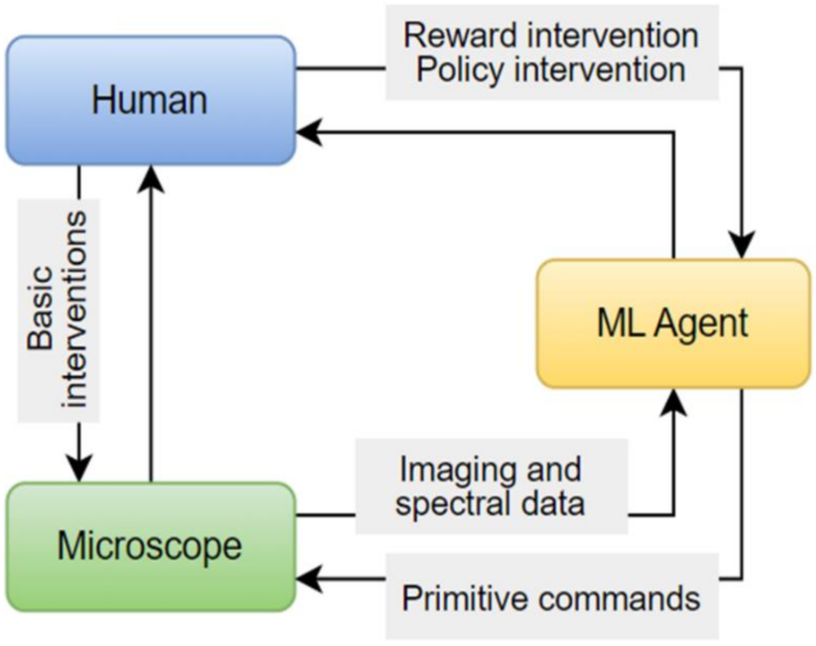






Human in the loop AE

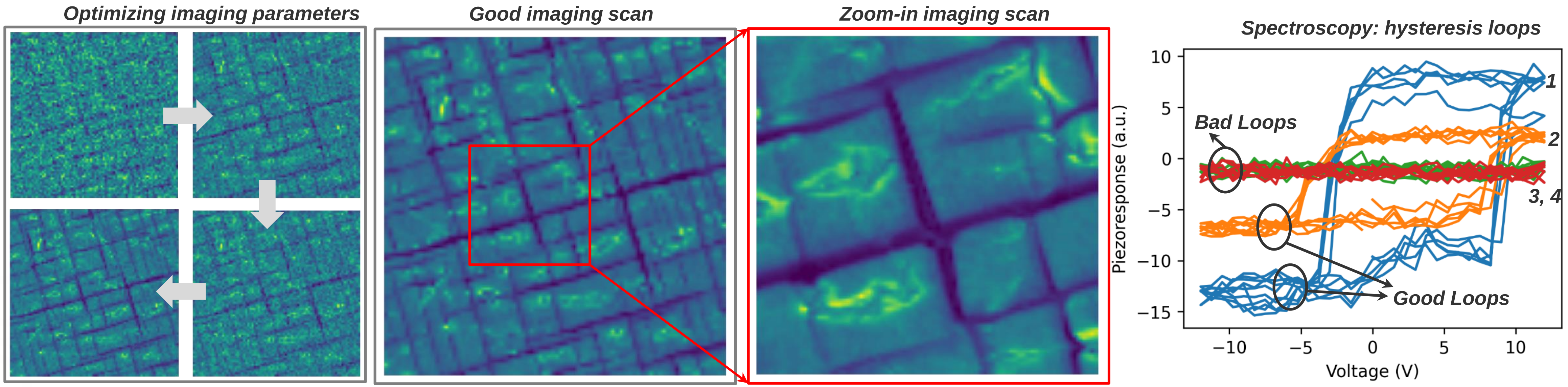


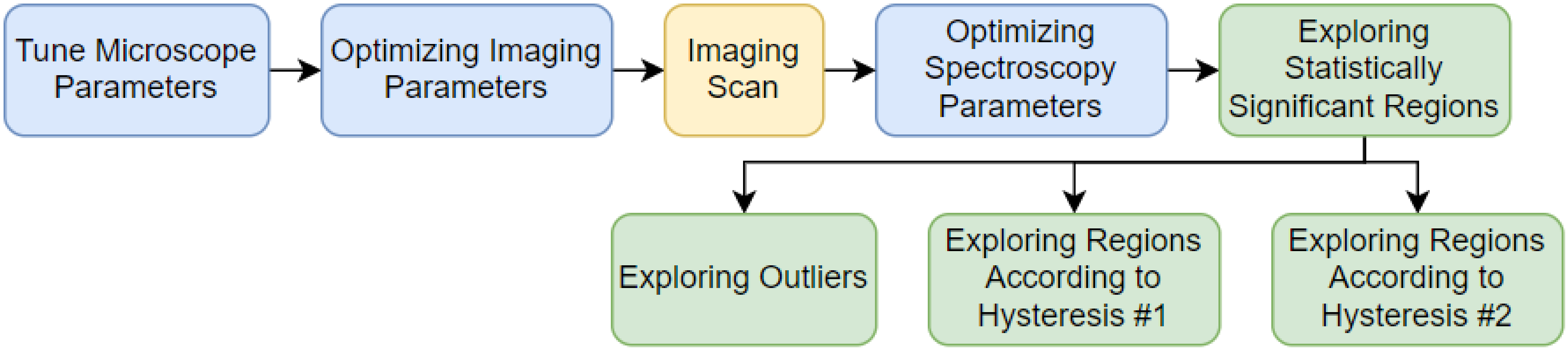


We can intervene on:

- Policies (acquisition functions): type and parameters
- Scalarizers: type and parameters
- Knowledge injection
- Direct operation

Future: full workflow optimization





Characteristic	**Definition**	**Availability**
Global image	Initial structural data set available before DKL experiment. Used to create patches for DKL training	Before
DKL latents	The latent variables encoding the structural information in the patches	During**
Scalarizer function	Function defining what characteristic of spectrum guides Bayesian Optimization	Before*
Acquisition function	Function combining DKL prediction and uncertainty of the scalarizer function	Before*
Policy	Principle for selection of next path. Simplest policy is maximization of acquisition function, but can be more complex including epsilon-greedy or switch between multiple scalarizers or acquisition functions. Human in the loop intervention tunes some aspect of the policy	Before*
Experimental trace	Collection of patches (and their coordinates) and spectra derived during experiment. Trace and global image are the results of AE SPM.	During

Characteristic	**Definition**	**Availability**
Live DKL model	DKL model in the state corresponding to the *n*-th experimental step	During
Final DKL model	DKL model in the state corresponding to the end of the experiment	After
Complete DKL model	DKL model trained on the full data set (if available from grid measurements, etc).	
Regret analysis	The difference between predictions of live DKL model and final DKL model after the whole experiment (i.e., after 200 steps in this work)	During** and After
Learning curve	Change of the DKL uncertainty (mean and deviation), indicative of the predictability of the patch-scalarizer relationship	During
Counterfactual scalarizer	The availability of full spectral data as a part of experimental trace allows to estimate what the BO step would be if scalarizer were chosen to be different	During
Trajectory analysis	Real-time trajectory of the probe that can be represented in the global image plane	During
Feature discovery	Analysis of the latent variables and latent representations of image patches and spectra in the trace. Here, we realize only patch analysis but extension to spectra is straightforward.	After
Latent trajectory analysis	Analysis of the experimental trajectory in the latent space of the full collection of the image patches derived from the global image	During**