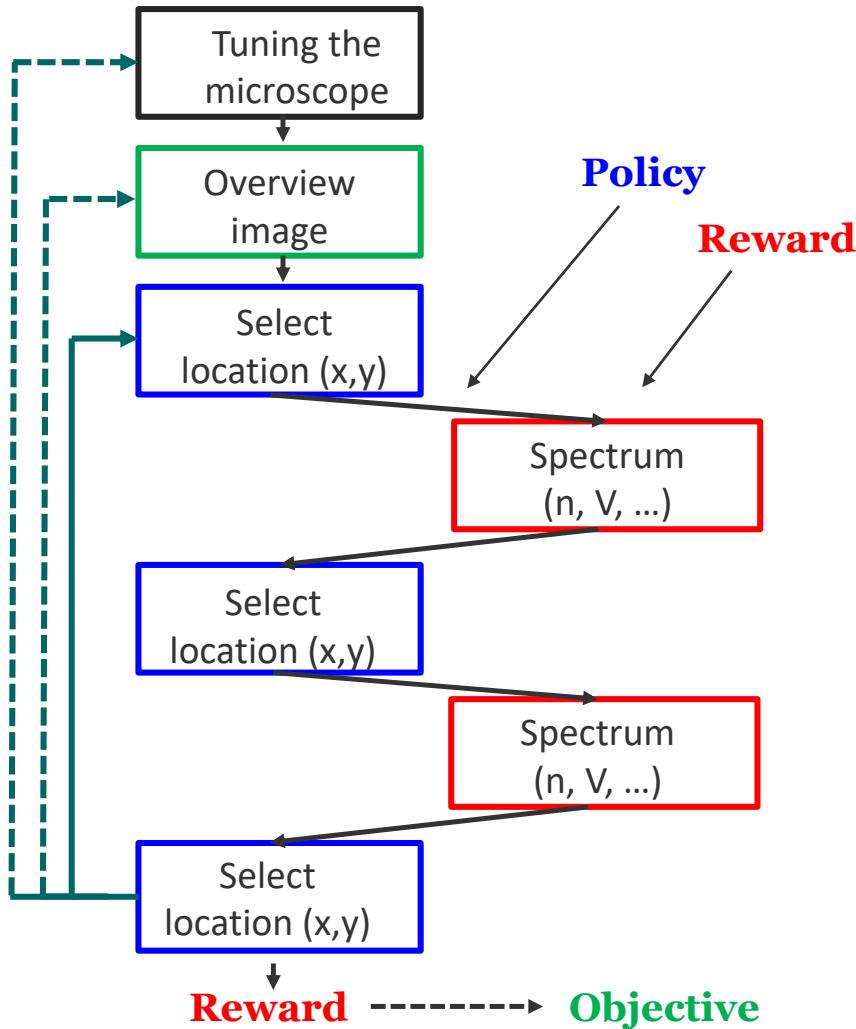


Day 5: Gaussian Processes and Bayesian Optimization

Sergei V. Kalinin

University of Tennessee, Knoxville, and
Pacific Northwest National Laboratory





To implement the ML workflows, we start from emulating the human operations:

- Well defined and explainable commands
- Extensive domain expertise
- Potentially available data from experiments

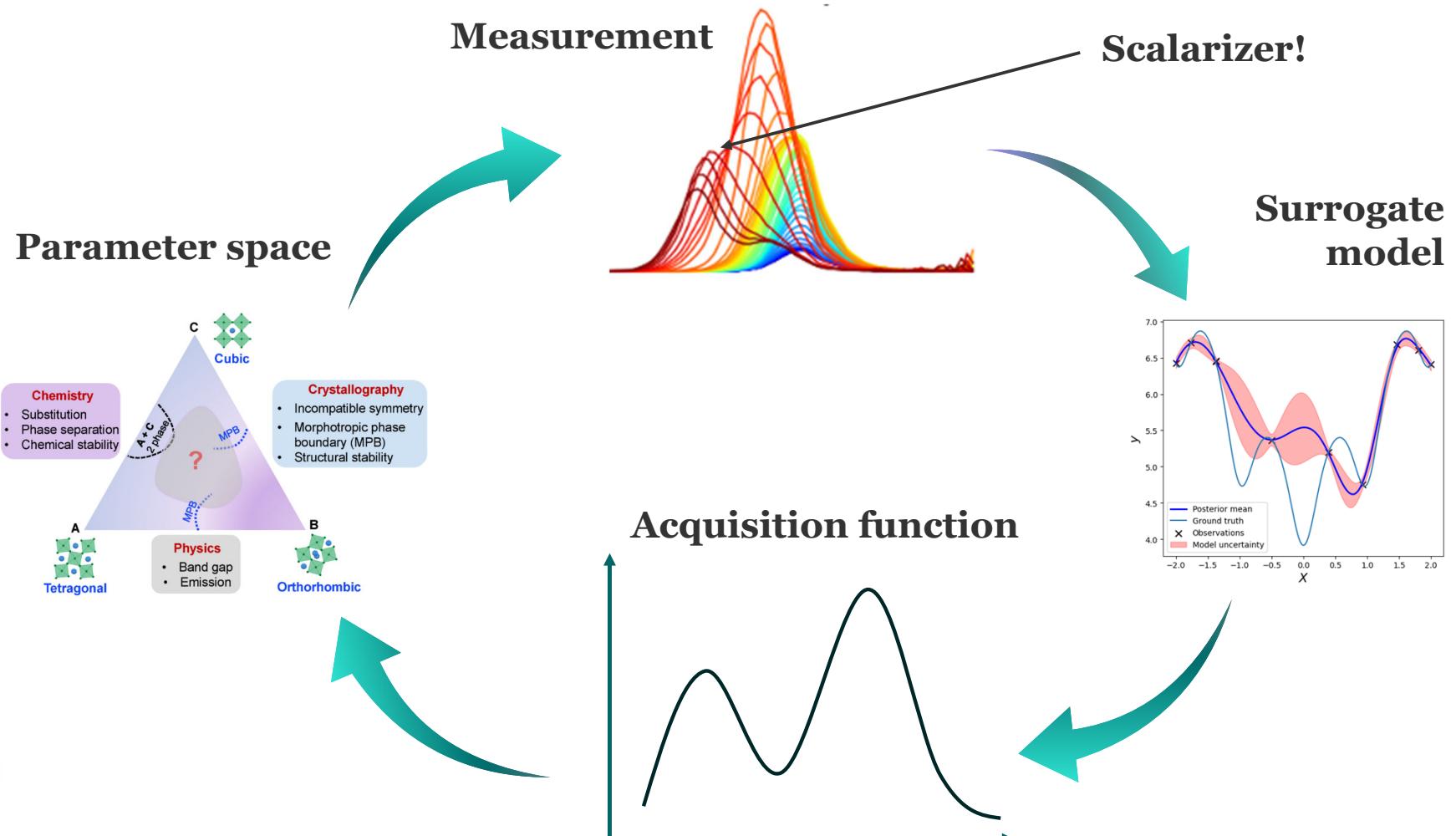
Development of ML workflows can give rise to more complex imaging modalities

- Data volumes and dimensionalities above human level
- More complex modes of sampling
- “Guardian angel” modules

However, we always have to think about

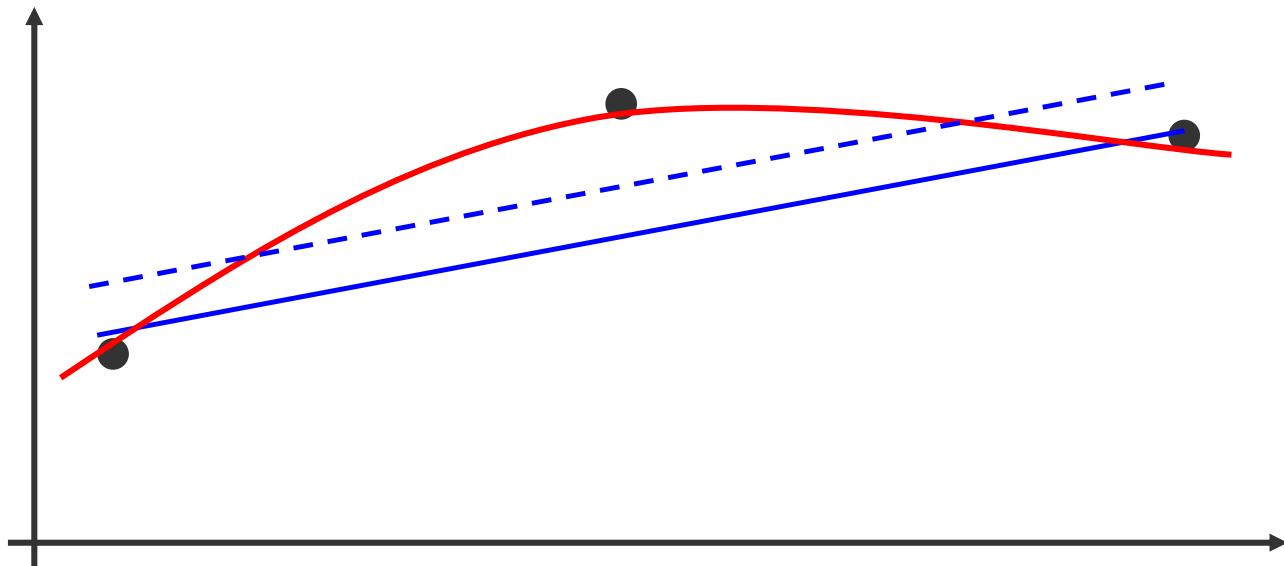
- Reward function(s) for imaging problem
- Reward functions for materials problem
- Overall objective

Classical Bayesian Optimization

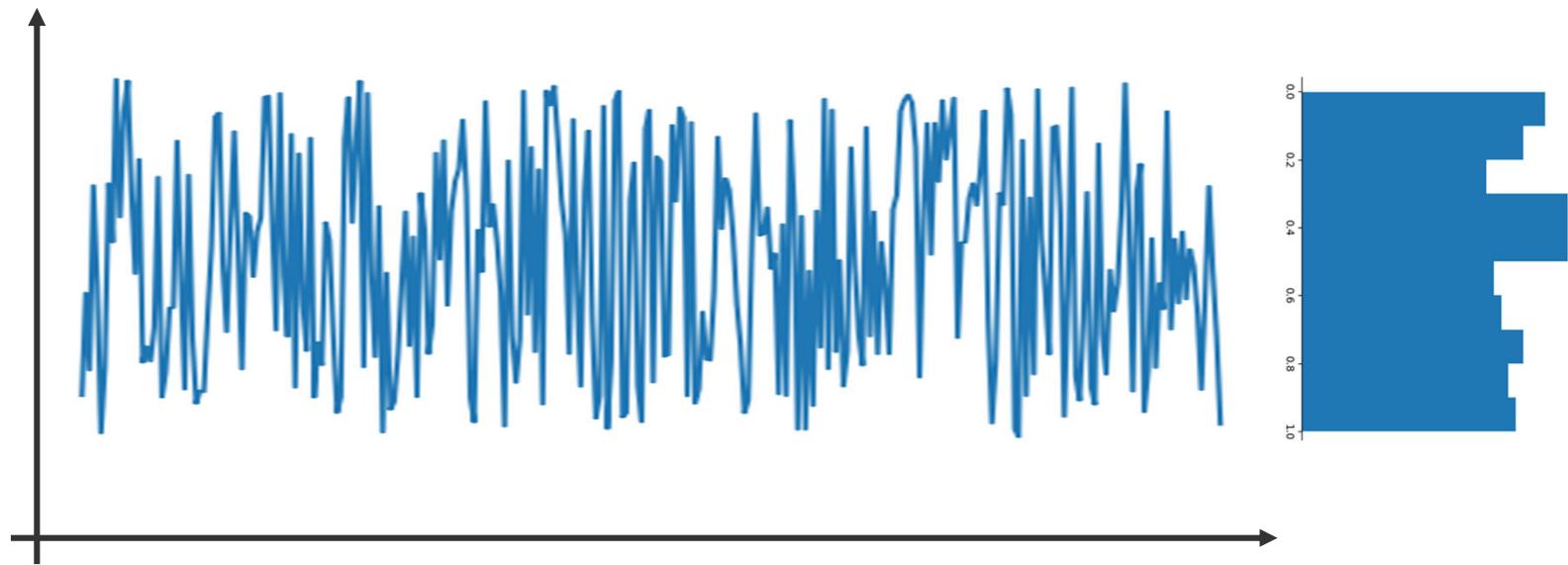


Implicitly, we postulate that **cost** and **gain** are constant along the iterative cycle

What do we know if we do not know anything?



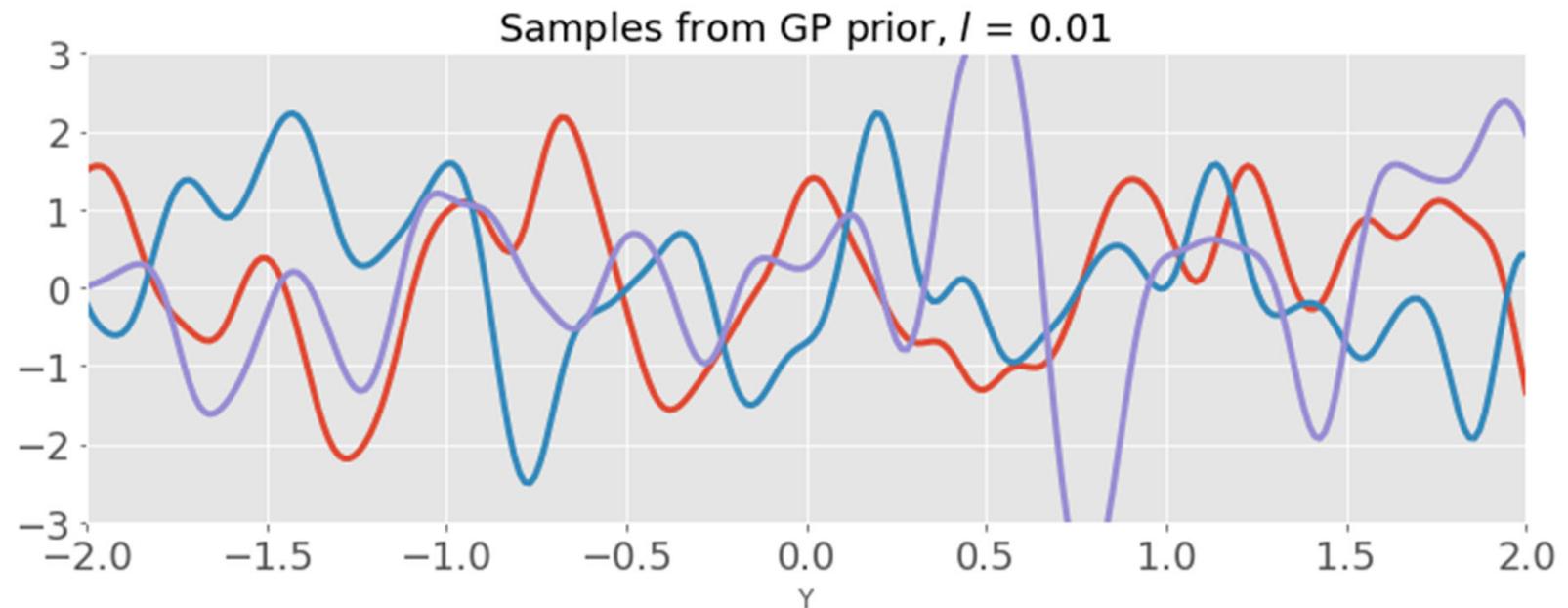
What do we know if we do not know anything?



Gaussian Process Regression

- Covariance matrix determines what type of functions we will allow.

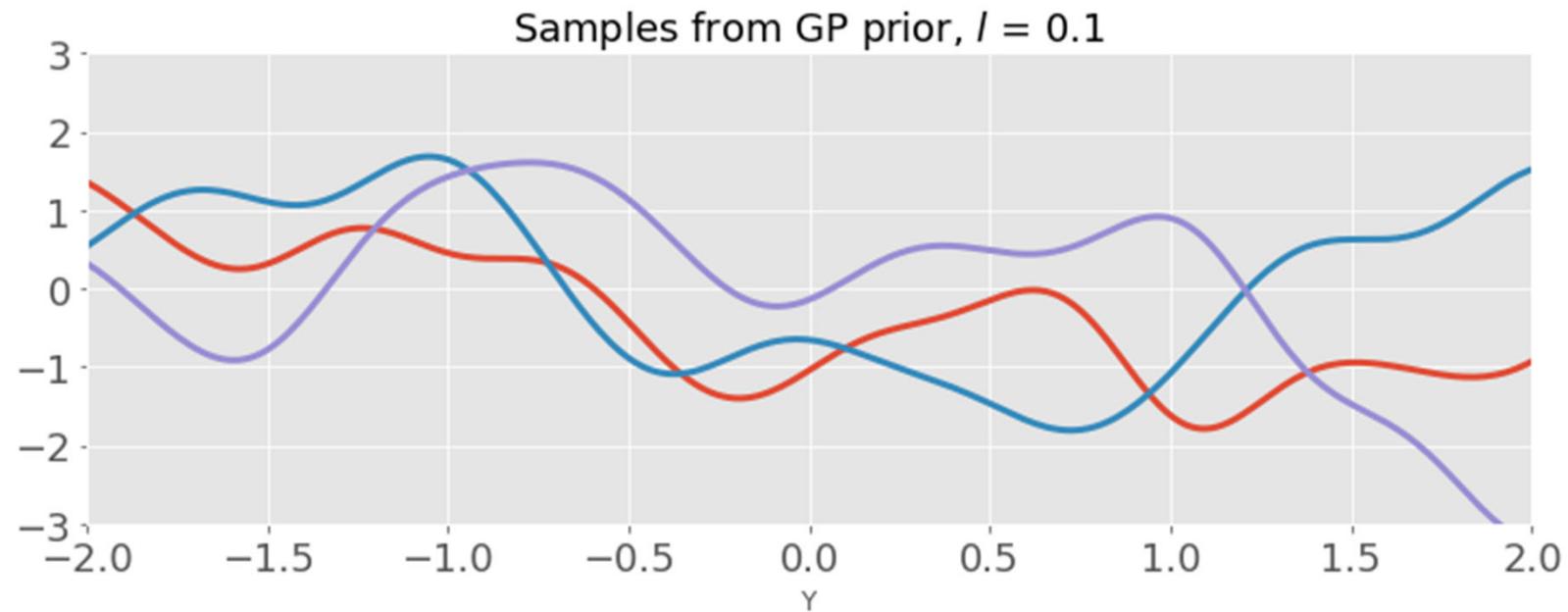
$$k(x, x') = \exp\left(-\frac{1}{2l}(x - x')^2\right)$$



Gaussian Process Regression

- Covariance matrix determines what type of functions we will allow.

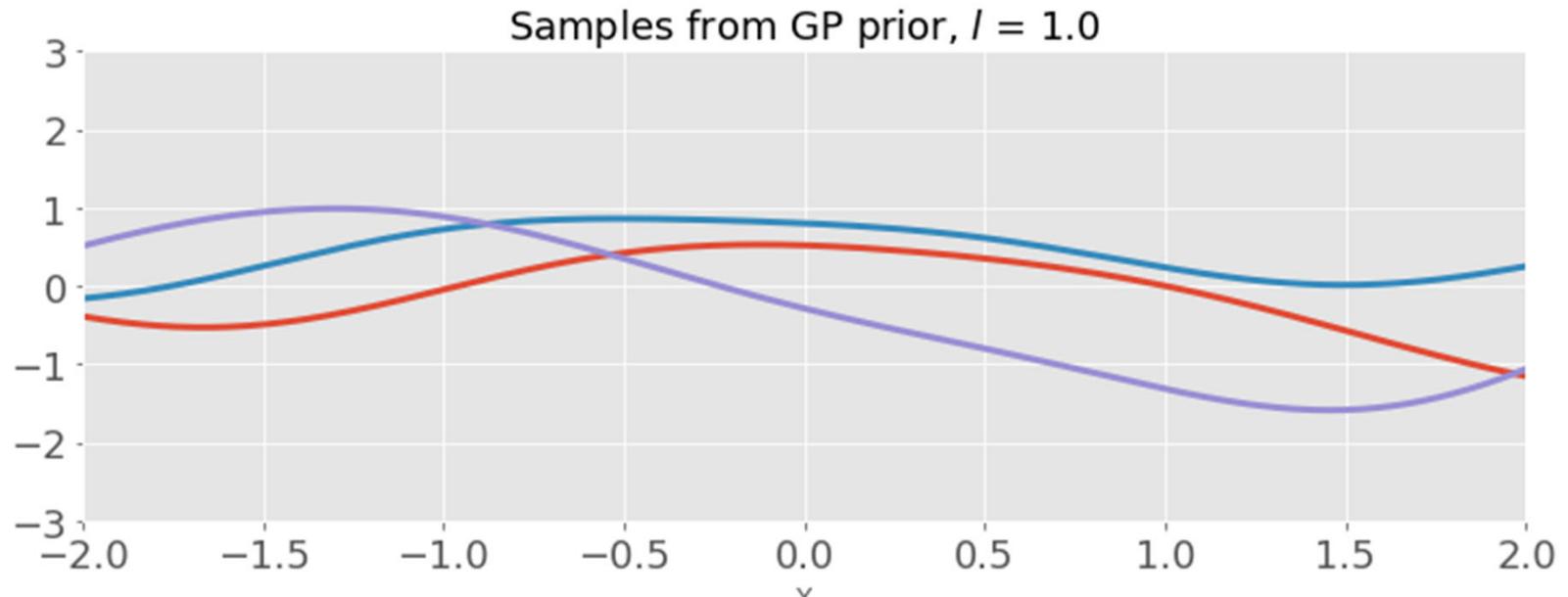
$$k(x, x') = \exp\left(-\frac{1}{2l}(x - x')^2\right)$$



Gaussian Process Regression

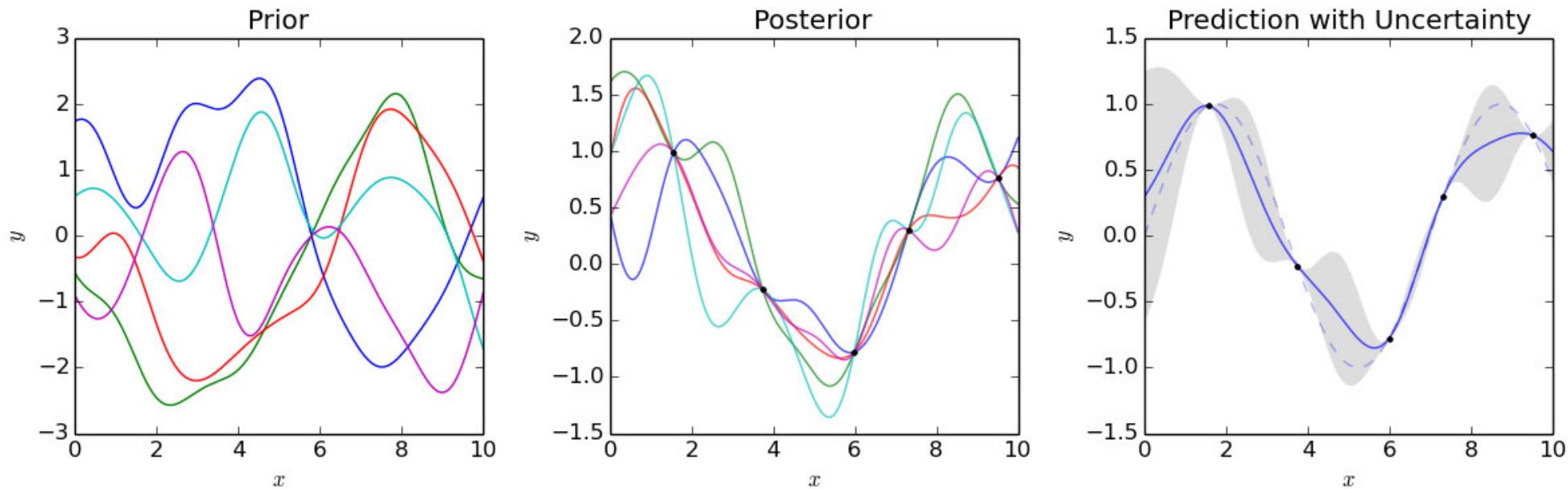
- Covariance matrix (kernel) determines what type of functions we will allow.

$$k(x, x') = \exp\left(-\frac{1}{2l}(x - x')^2\right)$$



l controls the length scale – sort of how far points should be to make them independent of each other.

Gaussian Process Regression



Prior:

What can the function be before the measurement
Measurement

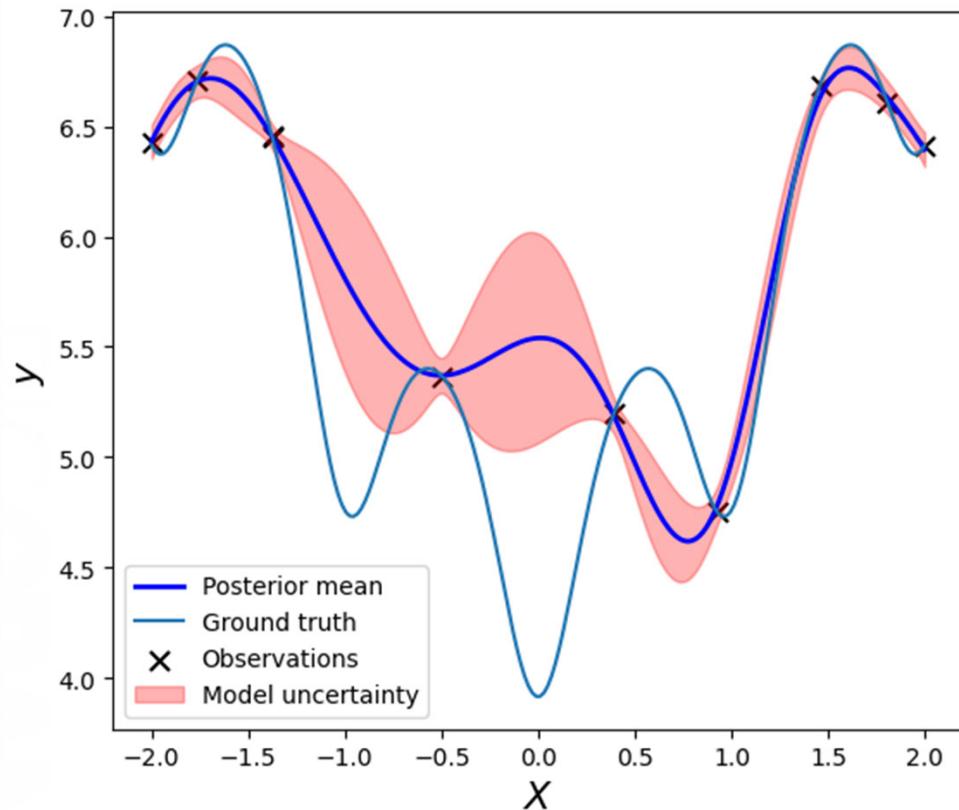
Posterior:

What can the function be after measurement

Policy:

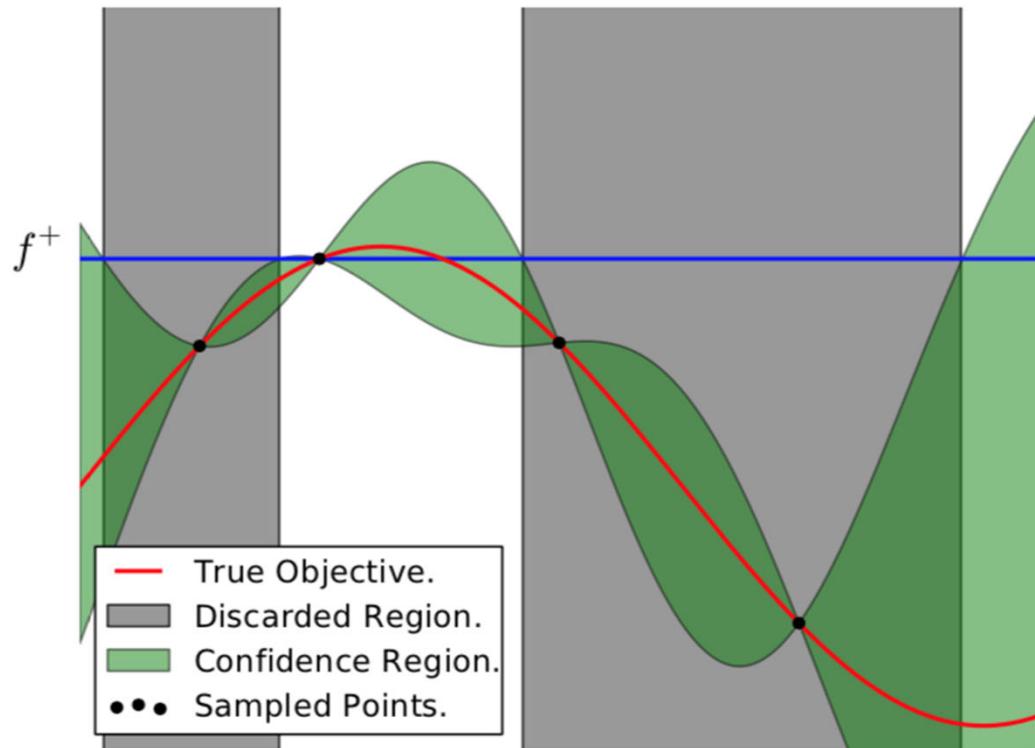
How do we balance exploration and exploitation (acquisition function)

GP Vocabulary



- Gaussian Process
- Kernel and kernel parameters
- Kernel Priors
- Noise Priors
- Posteriors

Bayesian Optimization

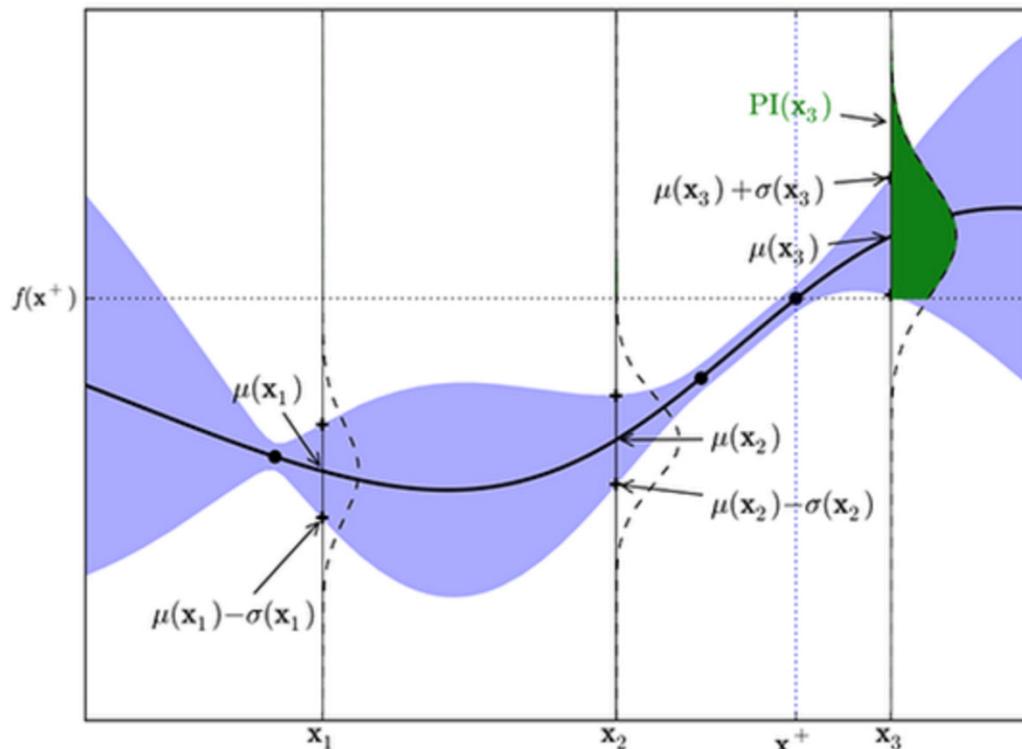


- We have some measurements in space X , and we want to maximize some property $f(X)$.
- How can we decide what point to measure next to best maximize f ?
- We need to balance the exploration of the space with exploitation of regions near we have already know

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

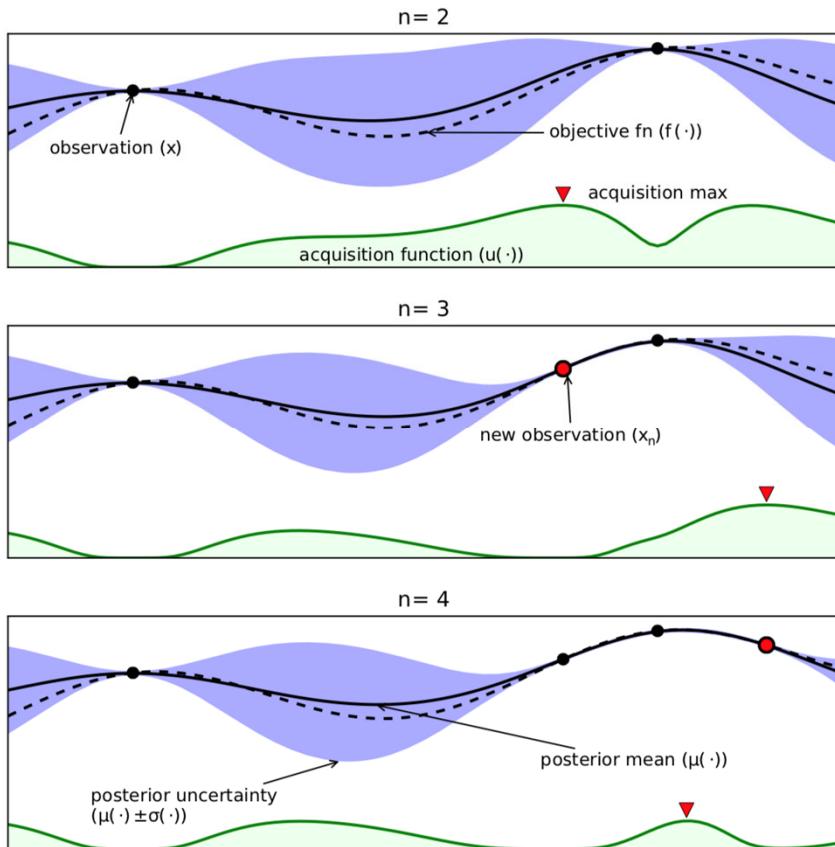
Acquisition Functions

Probability of Improvement Acquisition Function

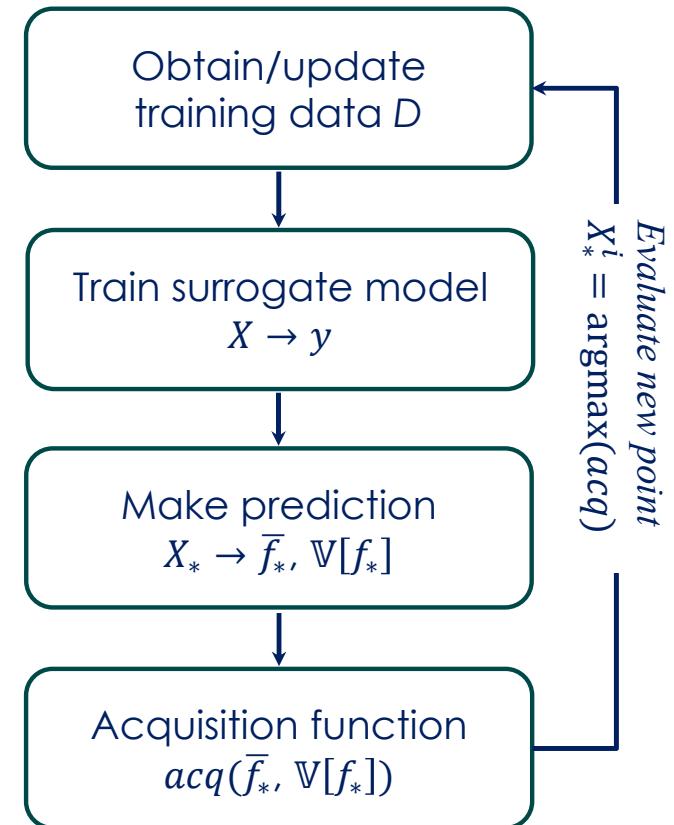


- 1. Upper confidence bound:** simplest possible - just take the upper confidence bound from the prediction
- 2. Probability of Improvement:** Integral from current functional maximum to upper limit of distribution as test point
- 3. Expected Improvement:** Instead of probability of improvement, we want to maximize the expected increase in the function value
- 4. There are (always) more...**

The basics: Bayesian Optimization



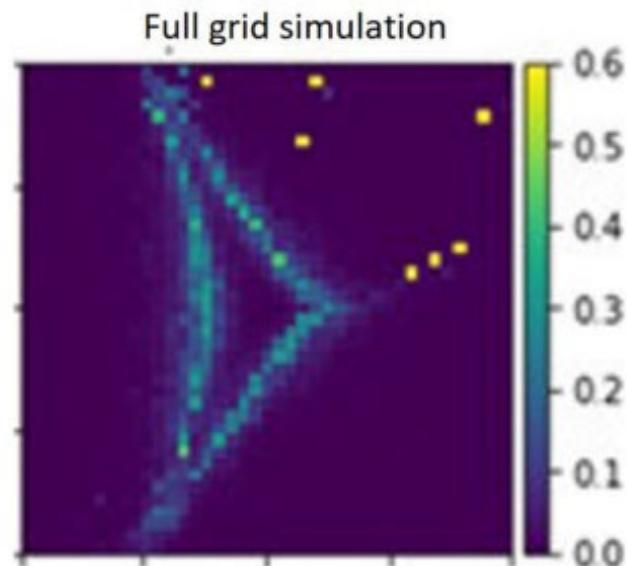
X, y : (sparse) Training data
 X_* : New (not yet evaluated) points



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

Bayesian Optimization for physical discovery

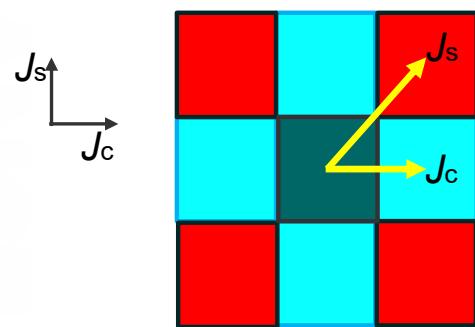
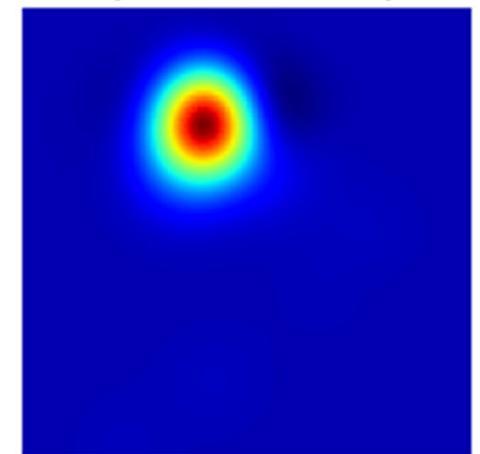
Discovering regions where heat capacity is maximized in NNN Ising model



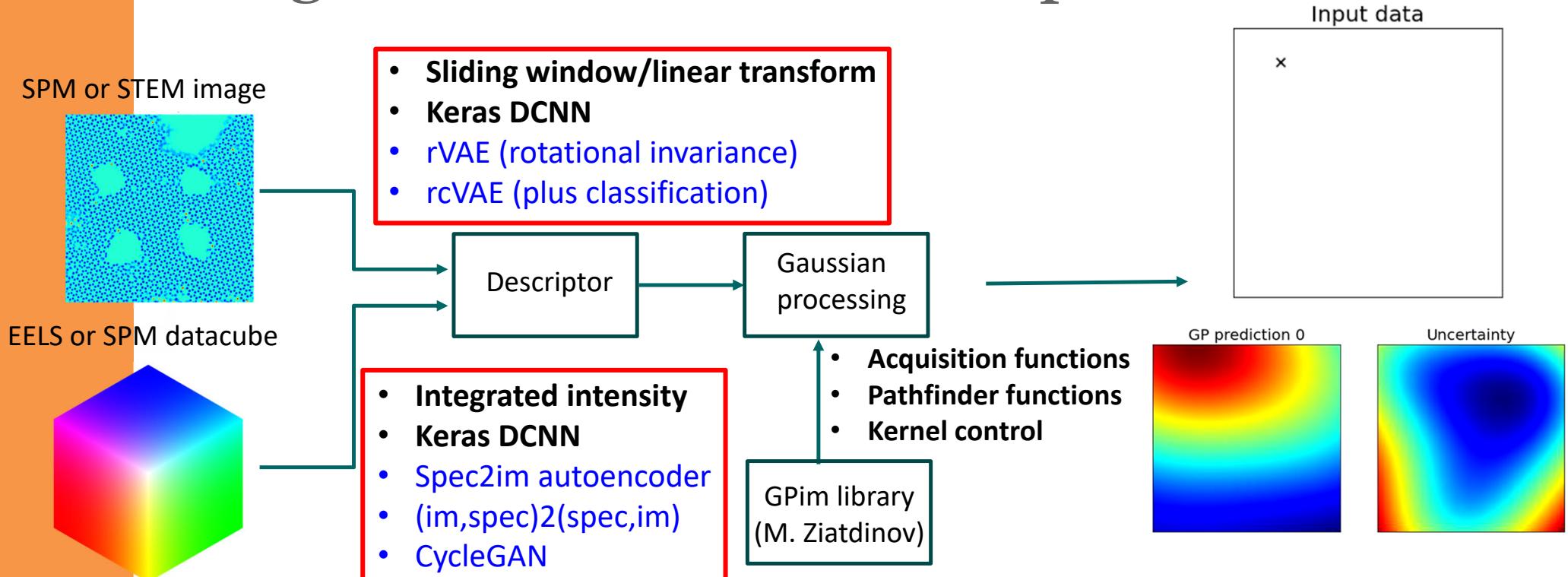
Explored points at step 0



GP prediction at step 0

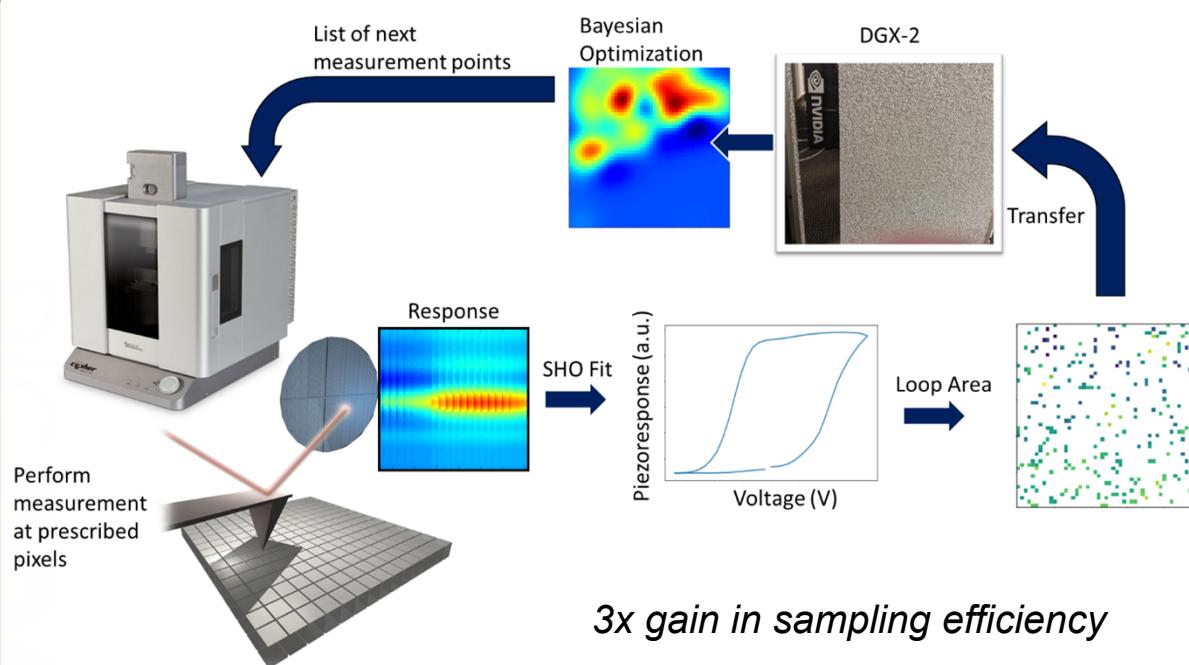


Going real time: automated experiment

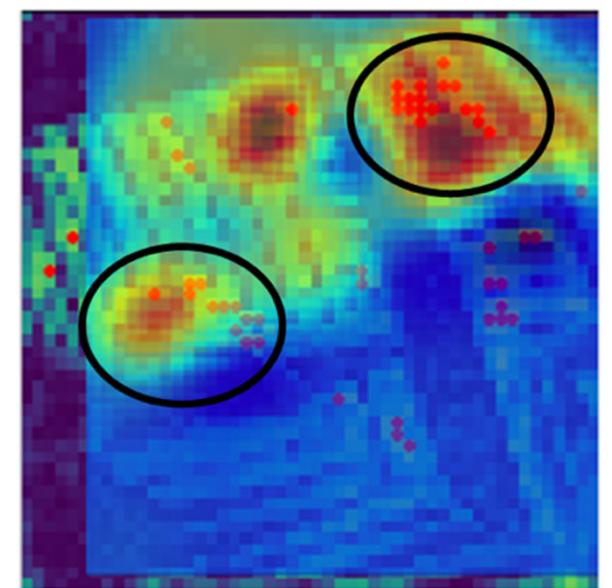


- AE based on structural analysis for STEM data
- AE based on spectral data in PFM
- AE based on DL for EELS data
- Feature of interest finding for mesoscopic images

BO for Self-Driving Microscopy



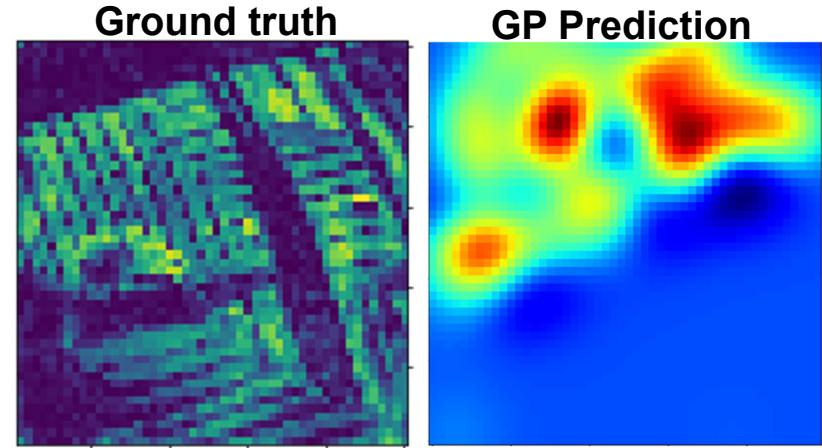
Comparison with “ground truth”



R. K. Vasudevan, K. Kelley, H. Funakubo, S. Jesse, S. V. Kalinin, M. Ziatdinov, **ACS Nano** (2021) <https://doi.org/10.1021/acsnano.oc10239>

Classical GP based BO

- Purely data-driven: limited advantage in high dimensional spaces
- Predicts scalar functions
- Typically used assuming equal cost of measurements
- And targeting fully automated process



Vasudevan et al, ACS Nano 2021

These assumptions rarely comport to real world scenarios

- We typically have ample (but partial) physical knowledge
- Multiple proxy signals
- Our observed data is very often high-dimensional (spectra, images)
- Cost and latencies of measurements is determined by physical equipment
- We can co-orchestrate measurements
- Humans are a part of the process (if process is slow)

But what about noises?

Gaussian Process learns the noise and kernel function while exploring parameter space. What if the noise level is not constant?

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

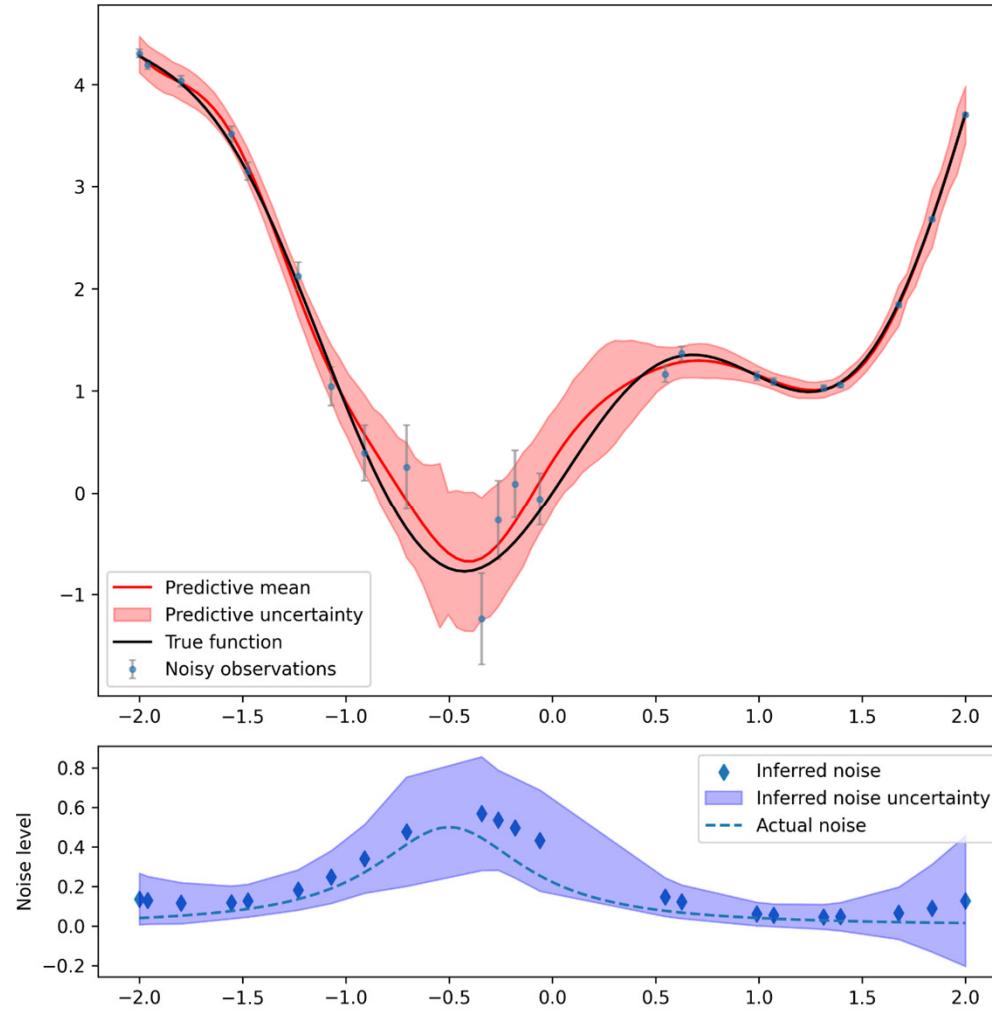
$$y(x) = f(x) + \sigma(\textcolor{violet}{x})$$

Solution: heteroscedastic GP uses

- one GP for function, and
- another GP for the noise

Note that we can **create models** for function and noise (structured GP)

Heteroscedastic GP



But what if the noise can be measured?

In many experimental scenarios, the experiment can be configured so that the noise can be measured (or estimated).

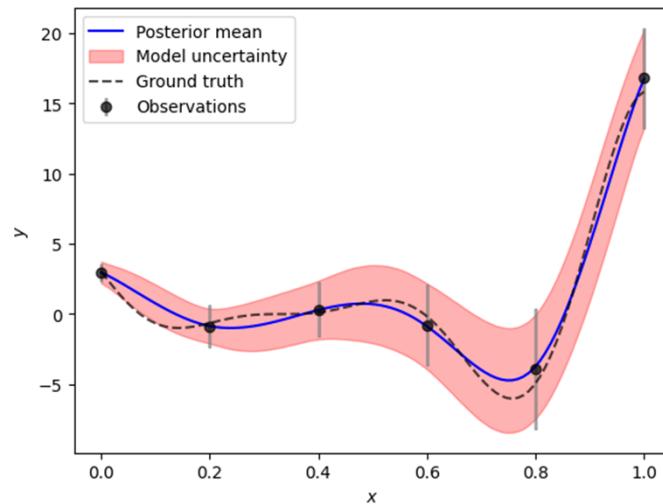
For example, it is often easier to measure multiple times (indentation curves, spectra, etc) at one location rather than move around

Alternatively, noise can be estimated from single measurement.

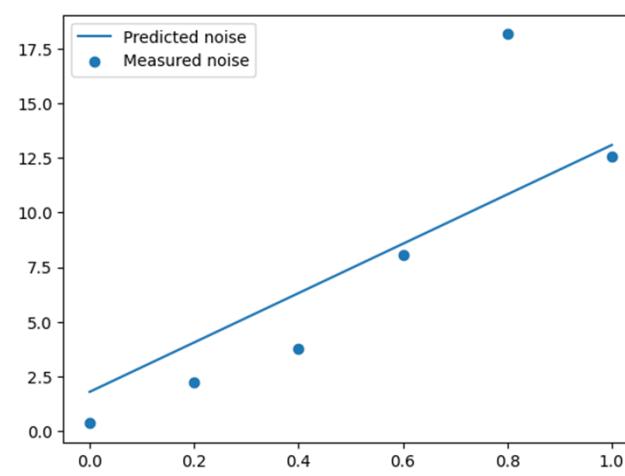
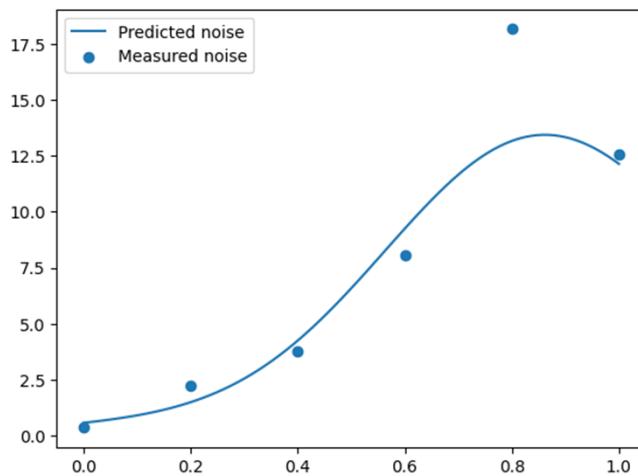
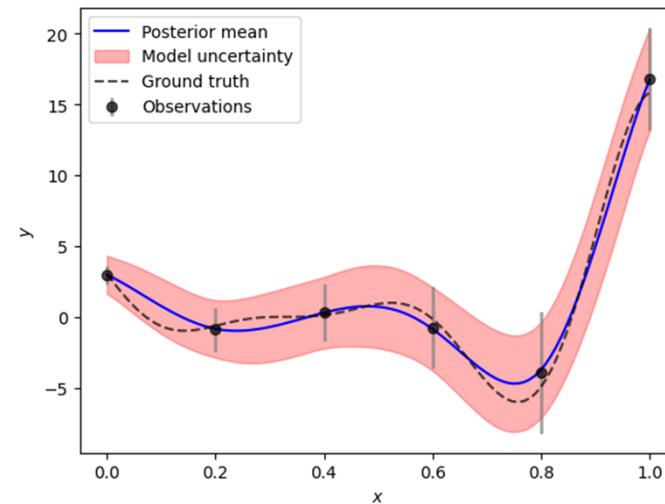
Note that we still need **noise model** even when the noise is measured, since we need to have an estimate of noise at the yet-unmeasured locations

Measured Noise GP

GP regression

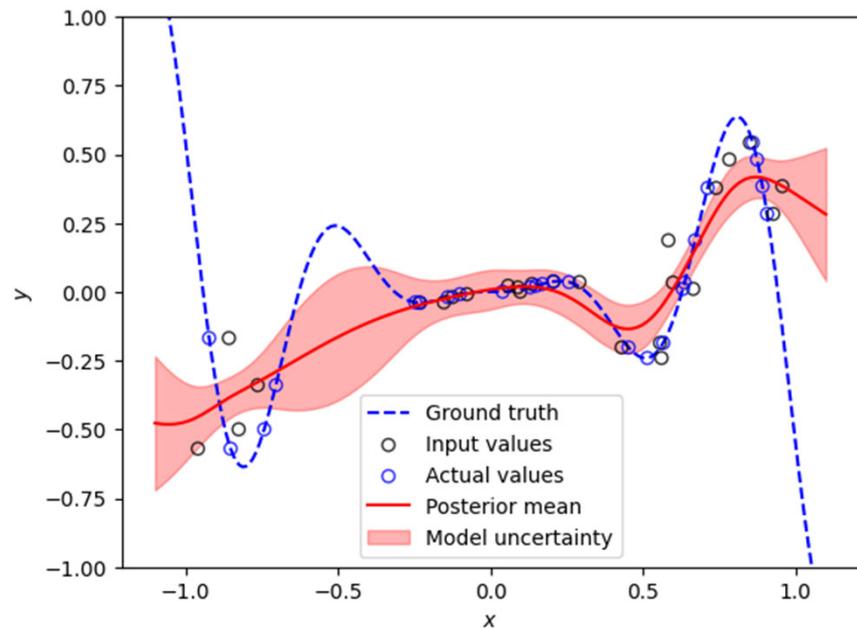


Linear regression

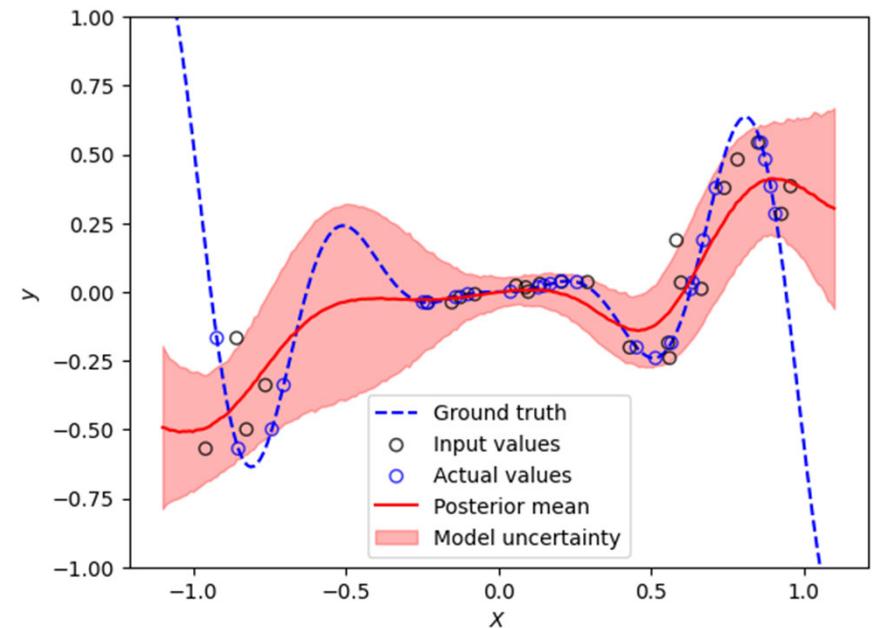


What if the measurement location is uncertain?

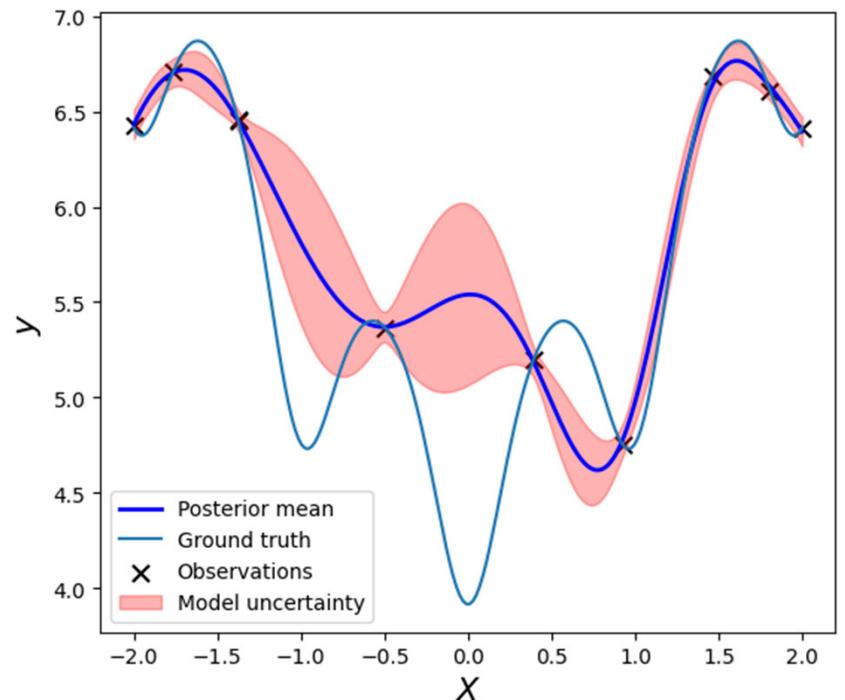
Simple GP



Uncertain Measurement



- Classical Bayesian Optimization is useful for microscope tuning and imaging optimization, but almost useless for exploration in image plane
- Limited to low D: we need Deep Kernel Learning for Structure-Property relationship discovery
- No physics priors: we need structured Gaussian Processes to learn physics



GP Augmented with Structural model

Define a probabilistic model:

$$\mathbf{y} \sim MVNormal(\mathbf{m}, \mathbf{K})$$

$$K_{ij} = \sigma^2 \exp(0.5(x_i - x_j)^2 / l^2)$$

$$\sigma \sim LogNormal(0, s_1)$$

$$l \sim LogNormal(0, s_2)$$

- We substitute a constant GP prior mean function \mathbf{m} with a structured probabilistic model of the expected behavior.
- This probabilistic model reflects our prior knowledge about the system, but it does not have to be precise.
- The model parameters are inferred together with the kernel parameters via the Hamiltonian Monte Carlo.
- The fully Bayesian treatment of the model allows additional control over the optimization via the selection of priors for the model parameters.

Prediction on new data X_* :

$$\mathbf{f}_*^i \sim MVNormal\left(\mu_{\theta^i}^{\text{post}}, \Sigma_{\theta^i}^{\text{post}}\right)$$

replaced with

$$\mu_{\theta^i}^{\text{post}} = \mathbf{m}(X_*) + \mathbf{K}(X_*, X | \boldsymbol{\theta}^i) \mathbf{K}(X, X | \boldsymbol{\theta}^i)^{-1} (\mathbf{y} - \mathbf{m}(X)) \rightarrow \mu_{\Omega^i}^{\text{post}} = \mathbf{m}(X_* | \phi^i) + \mathbf{K}(X_*, X | \boldsymbol{\theta}^i) \mathbf{K}(X, X | \boldsymbol{\theta}^i)^{-1} (\mathbf{y} - \mathbf{m}(X | \phi^i))$$

$$\Sigma_{\theta^i}^{\text{post}} = \mathbf{K}(X_*, X_* | \boldsymbol{\theta}^i) - \mathbf{K}(X_*, X | \boldsymbol{\theta}^i) \mathbf{K}(X, X | \boldsymbol{\theta}^i)^{-1} \mathbf{K}(X, X_* | \boldsymbol{\theta}^i)$$

$\Omega^i = \{\phi^i, \boldsymbol{\theta}^i\}$ is a single HMC posterior sample with the kernel and prob model parameters

GP Augmented with Structural Model

Standard Gaussian process aims to discover function based on learned correlations (kernel)

Probabilistic model

$$m = y_0 - \sum_{n=1}^N L_n \quad (N=2)$$

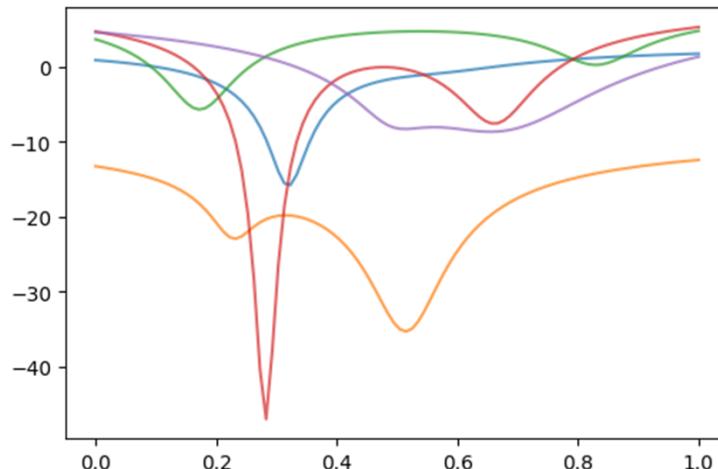
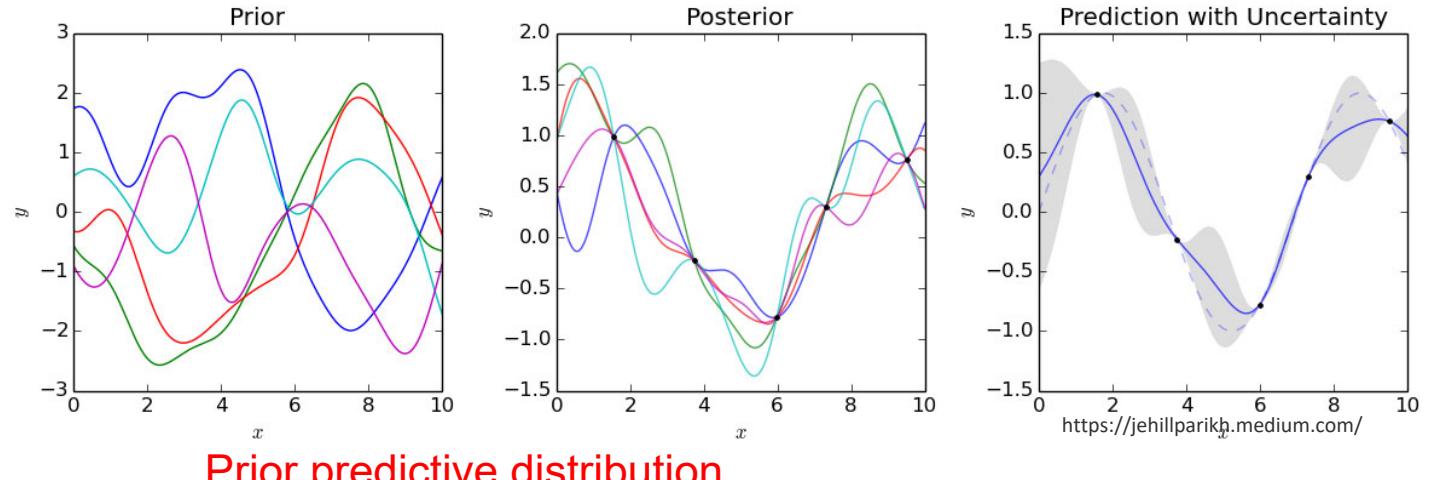
$$y_0 \sim Uniform(-10, 10)$$

$$L_n \sim \frac{A_n}{\sqrt{(x-x_n^0)^2+w_n^2}}$$

$$A_n \sim LogNormal(0, 1)$$

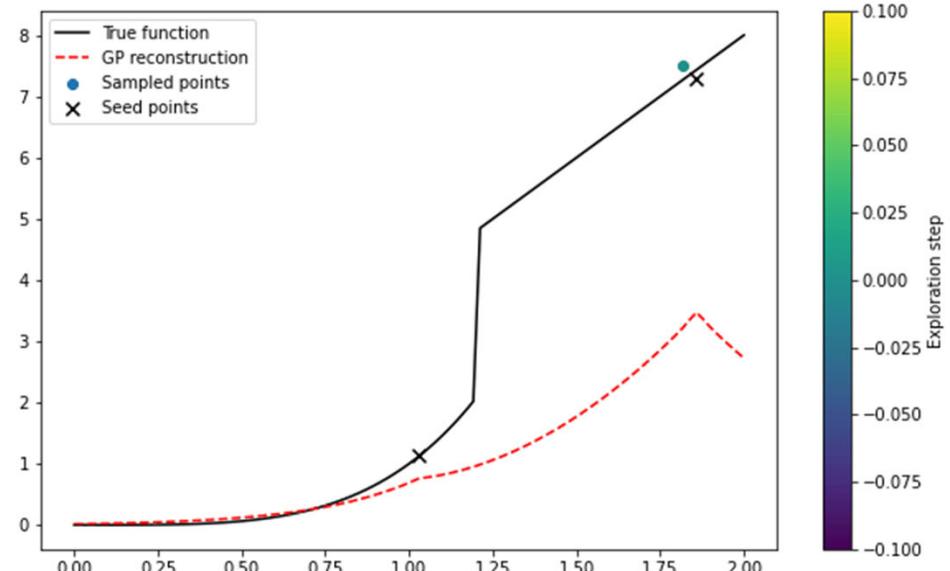
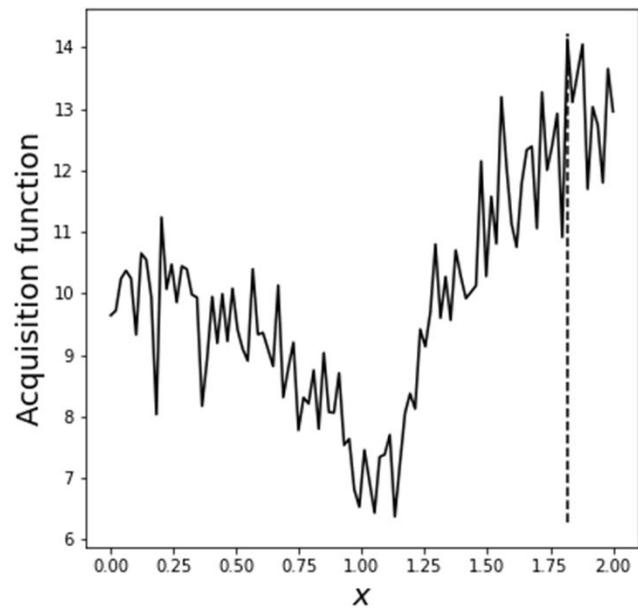
$$w_n \sim HalfNormal(.1)$$

$$x_n^0 \sim Uniform(0, 1)$$

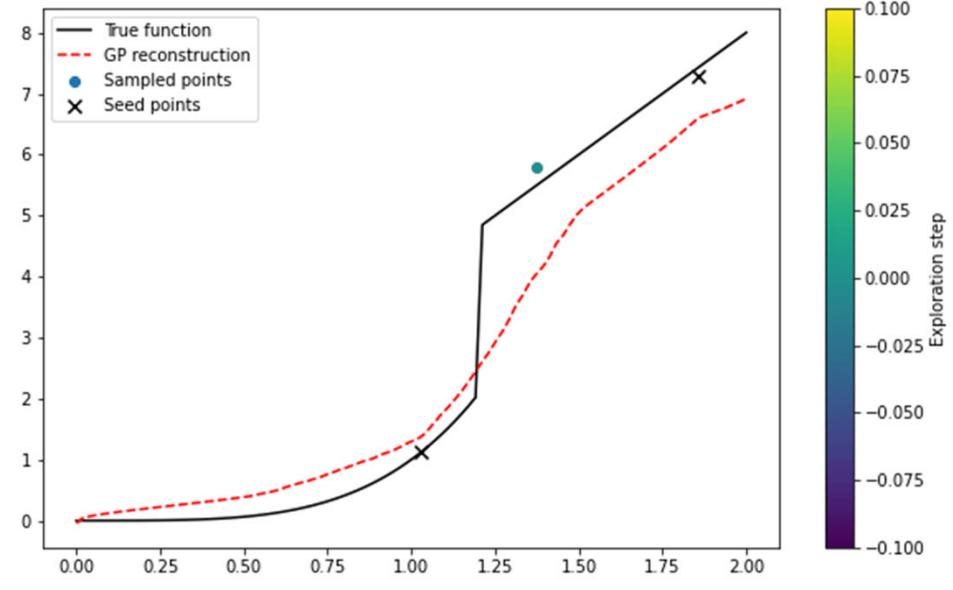
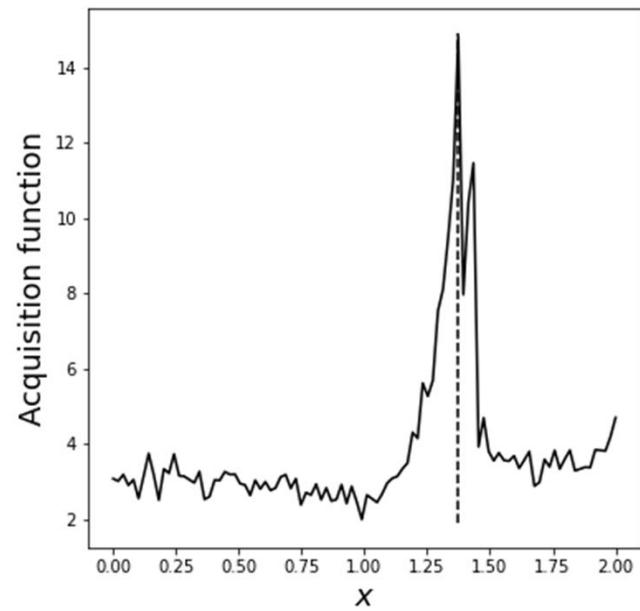


This model simply tells us that there are two minima in our data but does not assume to have any prior knowledge about their relative depth, width, or distance

Simple GP search



Structured GP search

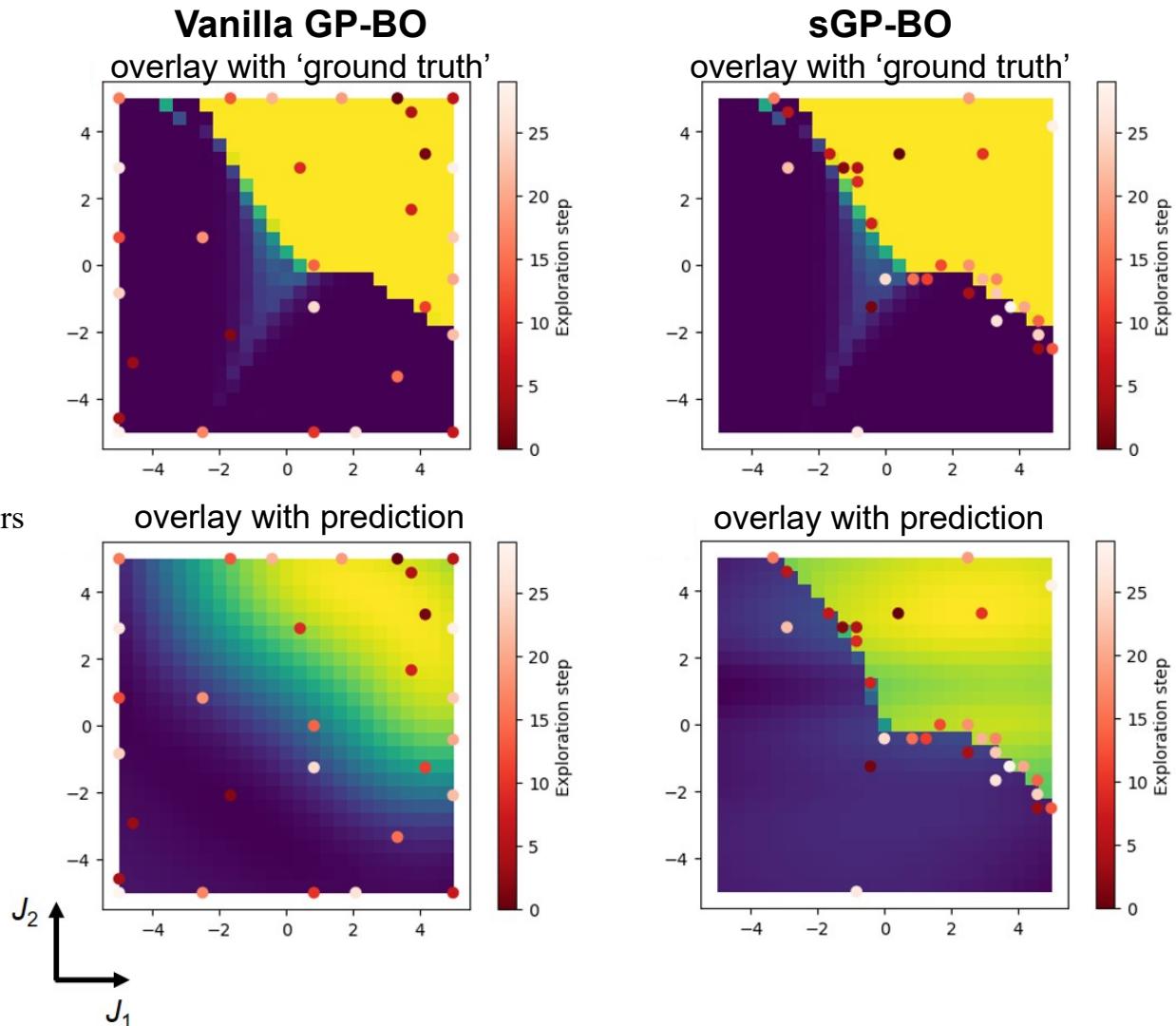


Application to Ising model

Probabilistic model

$$A/\tanh\left(\frac{f(J_1)+f(J_2)}{w}\right)$$

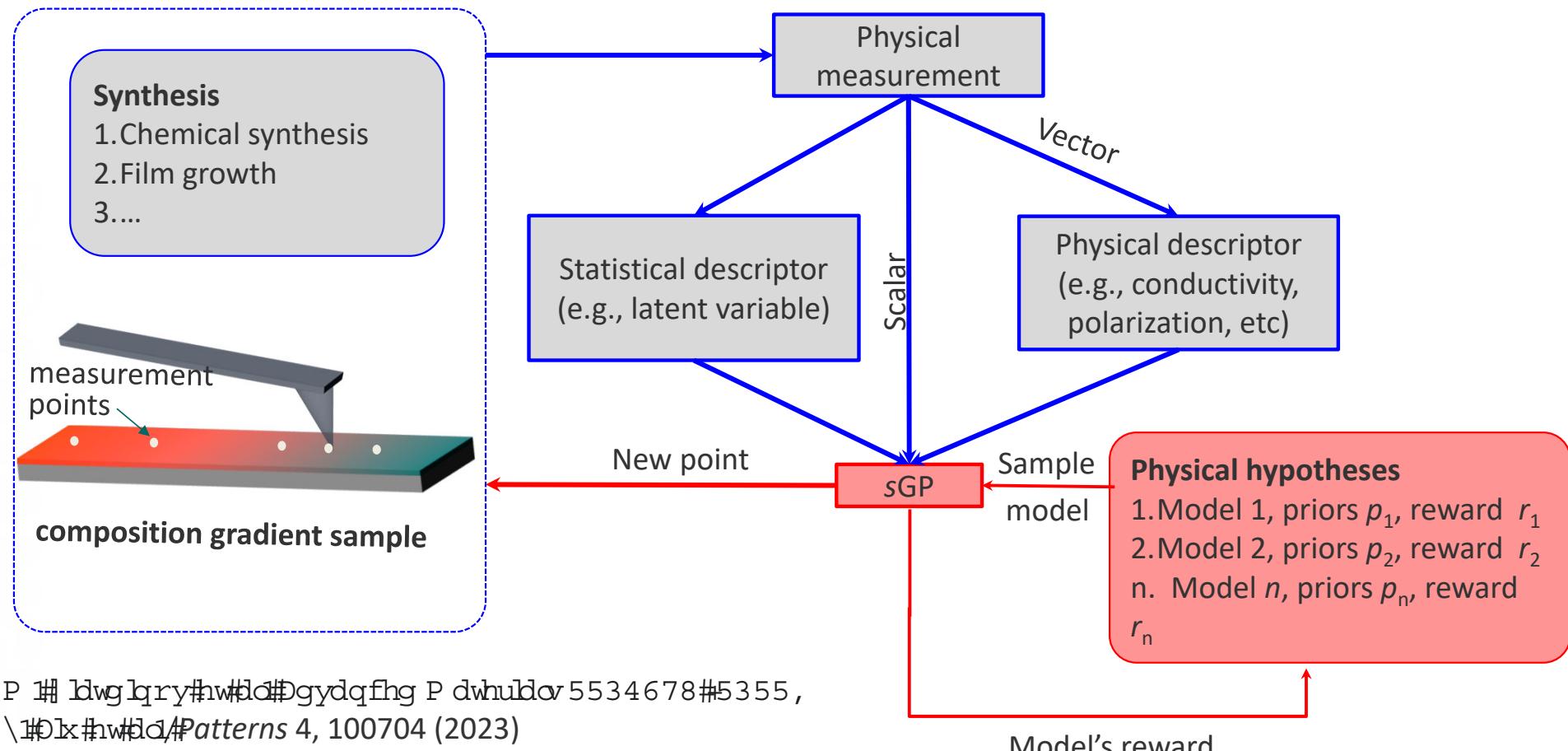
where $f(J)$ is a third-degree polynomial with normal priors on its parameters



Hypothesis Active Learning

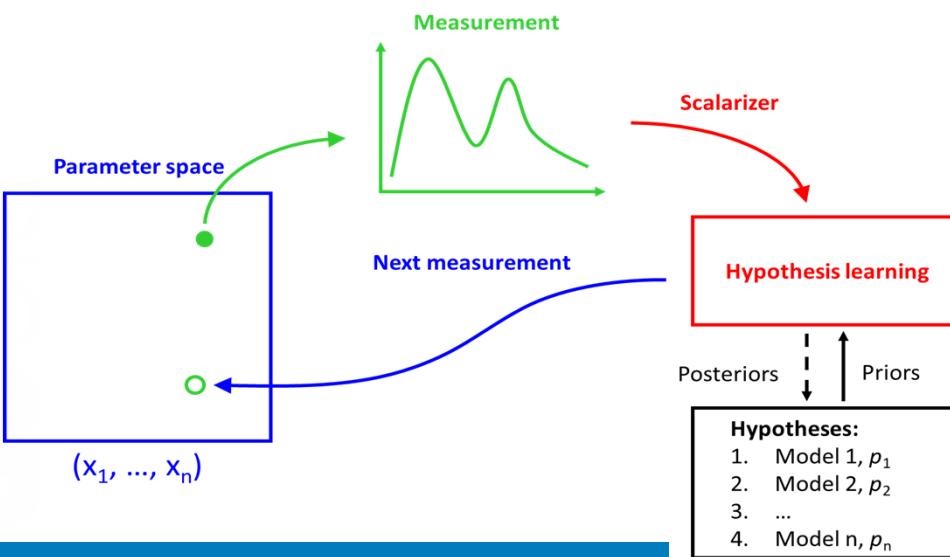
Co-navigation of experimental and hypothesis spaces

Goal: Learn (1) physical property distribution and (2) a correct model of system's behavior

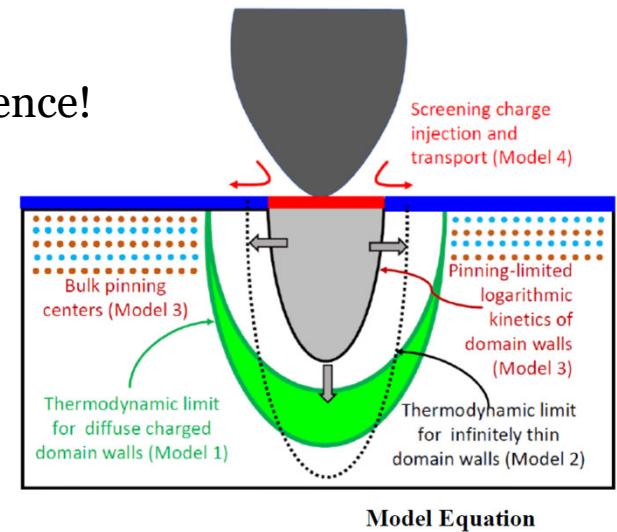


Hypothesis Learning

- Can ML algorithm think like a scientist?
- Yes – automated experiment can pursue hypothesis-driven science!



THE UNIVERSITY OF TENNESSEE KNOXVILLE



Thermodynamic 1

Model I

$$r(V) = r_{cr} + r_0 \sqrt{\left(\frac{V}{V_c}\right)^{2/3} - 1}$$

Thermodynamic 2

Model II

$$r(V) = r_{cr} + r_0 \sqrt[3]{\left(\frac{V}{V_c}\right)^2 - 1}$$

Wall pinning

Model III

$$r(V, t) = V^\alpha \log \tau$$

Charge injection

Model IV

$$r(V, t) = V^\alpha \tau^\beta$$

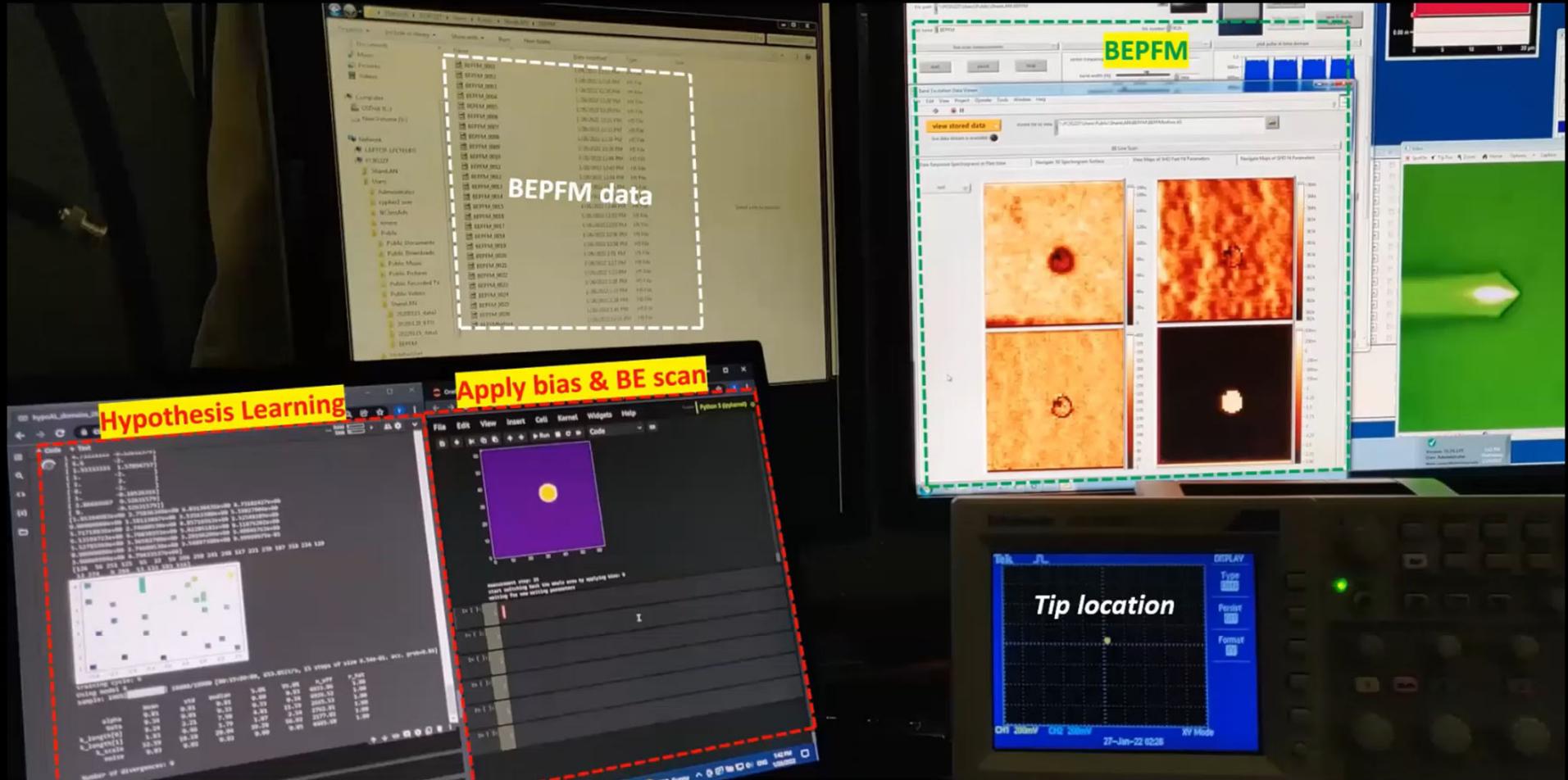
Patterns Open access

ARTICLE | VOLUME 4, ISSUE 3, 100704, MARCH 10, 2023 Download Full Issue

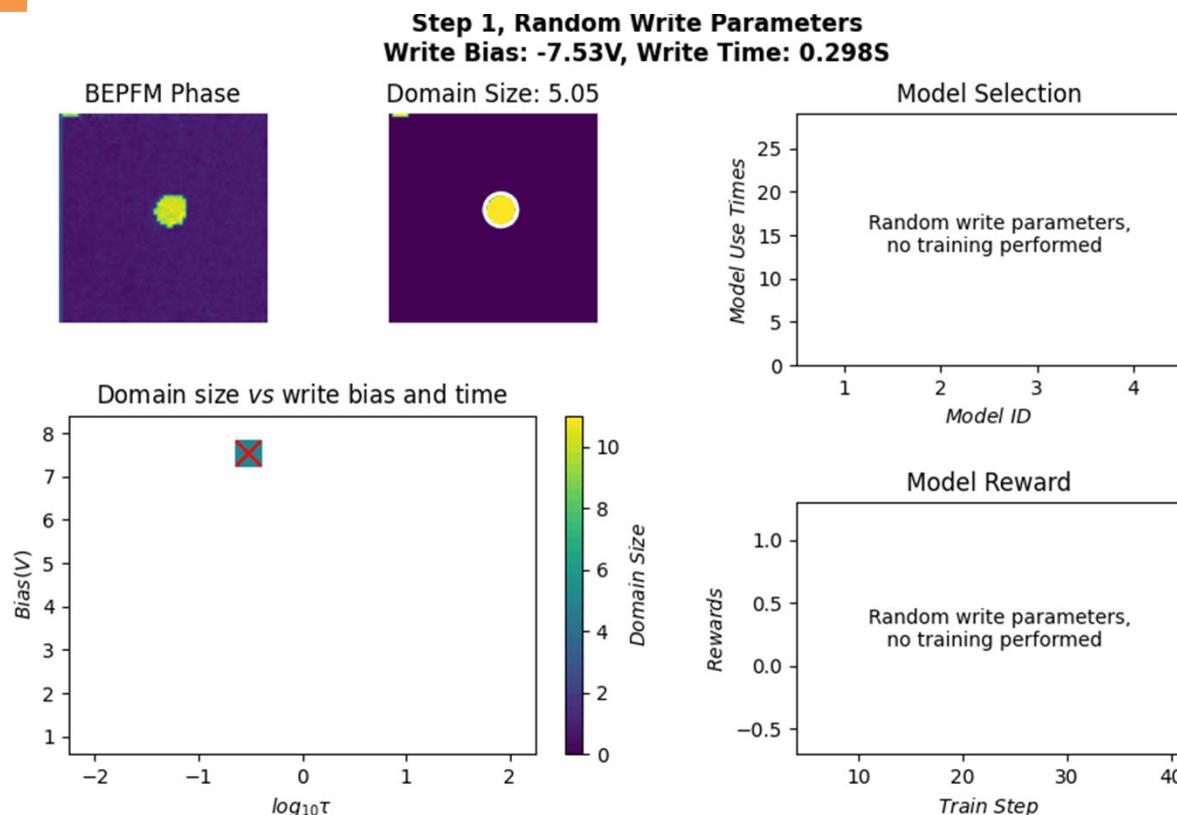
Autonomous scanning probe microscopy with hypothesis learning:
Exploring the physics of domain switching in ferroelectric materials

Yongtao Liu ⁵ • Anna N. Morozovska • Eugene A. Eliseev • ... Rama Vasudevan •
Maxim Ziatdinov • Sergei V. Kalinin • Show all authors • Show footnotes

Open Access • DOI: <https://doi.org/10.1016/j.patter.2023.100704> •



Hypothesis learning in action



- ML algorithm has 4 competing hypothesis on domain switching mechanisms
- These hypothesis represent full set of possibilities for this system
- The microscope chooses experimental parameters in such a way as to establish which hypothesis is correct fastest
- Important: the same approach can be implemented in synthesis and electrical characterization
- Machine learning meets hypothesis-driven scientific discovery!

Y. Liu, arxiv 2202.01089
Y. Liu, arxiv 2112.06649

Combinatorial Synthesis

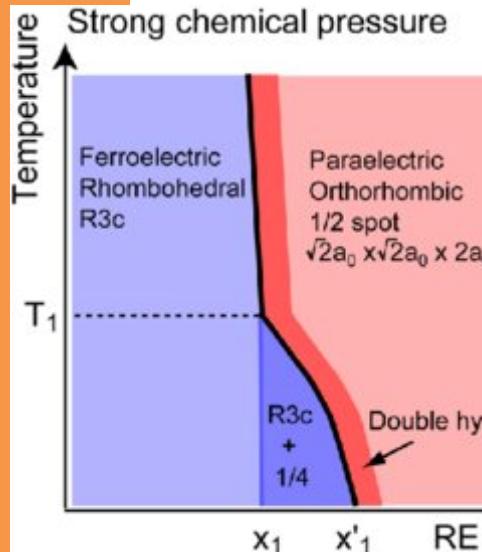
ADVANCED MATERIALS

Research Article

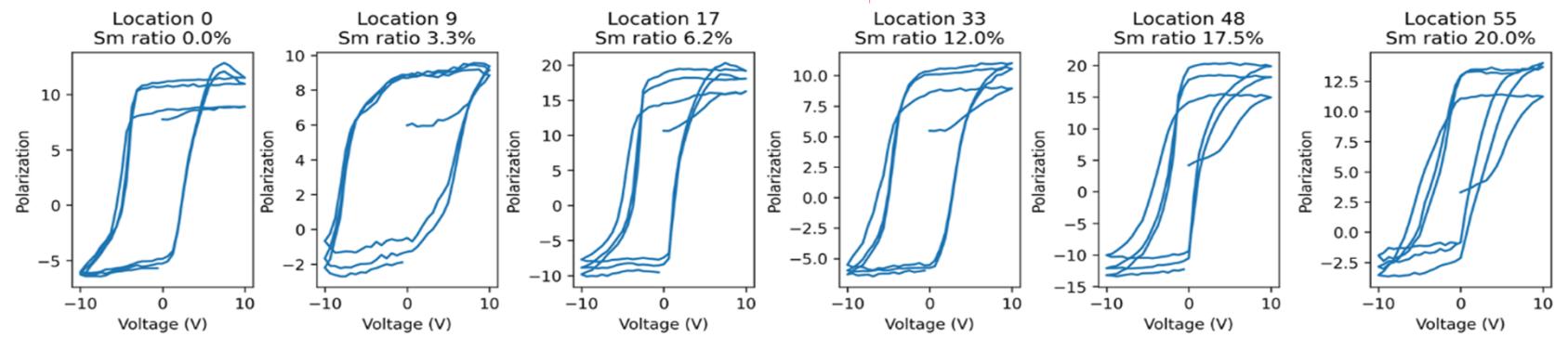
Hypothesis Learning in Automated Experiment: Application to Combinatorial Materials Libraries

Maxim A. Zlatdinov ✉, Yongtao Liu, Anna N. Morozovska, Eugene A. Eliseev, Xiaohang Zhang, Ichiro Takeuchi, Sergei V. Kalinin ✉

First published: 12 March 2022 | <https://doi.org/10.1002/adma.202201345> | Citations: 17



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



Hypothesis Selection for Ferroelectric

Model 1 (second order phase transition):

$$S = \begin{cases} S_0 \left(1 - \frac{x}{x_0} \right)^2 + C, & x \leq x_c, \\ C, & x > x_c \end{cases}$$

Model 2 (first order phase transition):

$$S = \begin{cases} S_0 \left(1 - \frac{x}{x_0} \right)^{\frac{5}{4}} + C_0, & x \leq x_c, \\ C_1, & x > x_c \end{cases}$$

