THE UNIVERSITY of TENNESSEE KNOXVILLE

Lecture 21: Noise and Structured Gaussian Processes

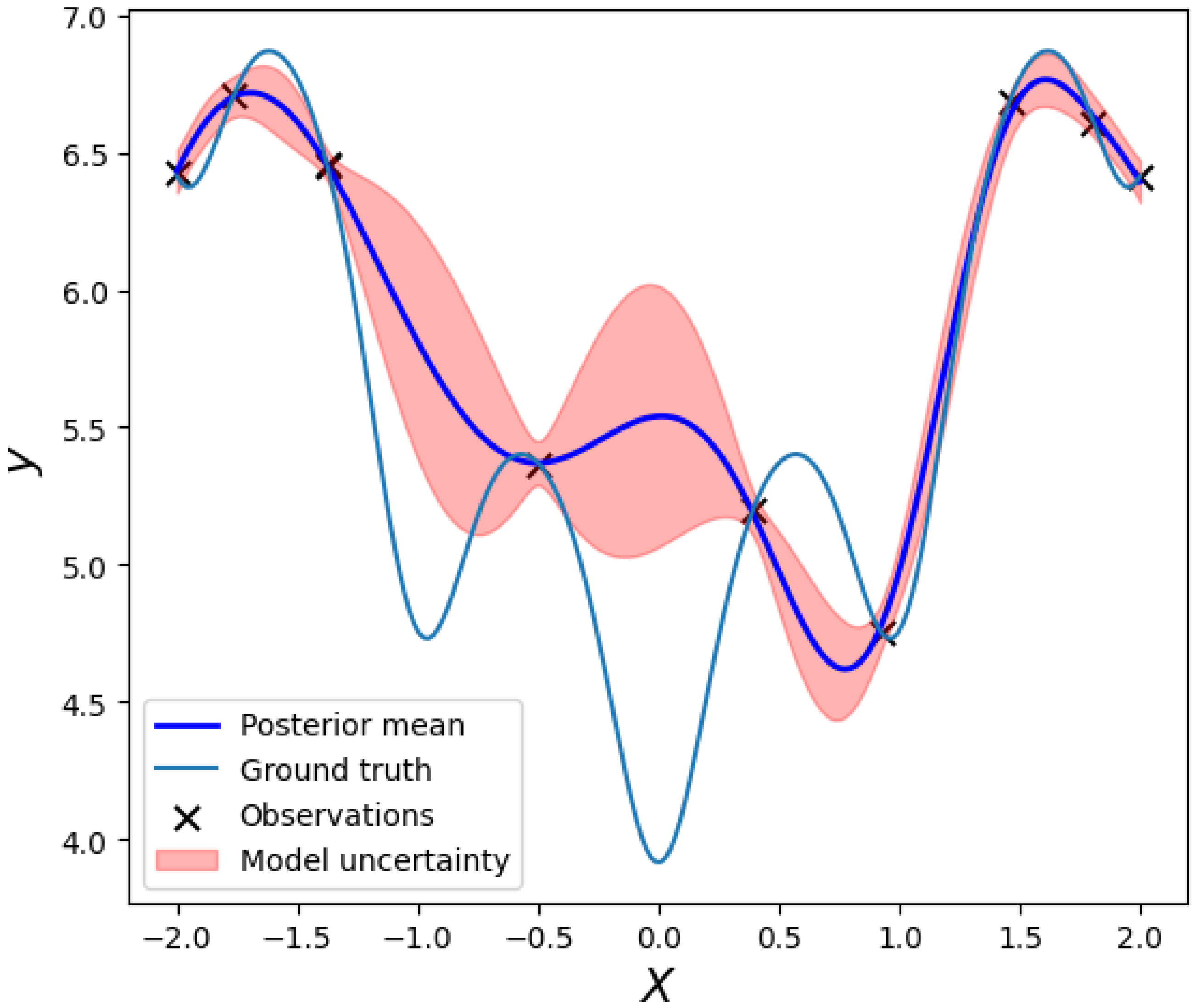
Instructor: Sergei V. Kalinin

What have we learned

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

- Bayesian Optimization
- Bayesian Optimization based on Gaussian Process

- Acquisition Functions
- Cost-award BO



… there are **known knowns**; there are things we know we know. We also know there are **known unknowns**; that is to say we know there are some things we do not know. But there are also **unknown unknowns**—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tends to be the difficult ones.

D. Rumsfeld

But what about noises?

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

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

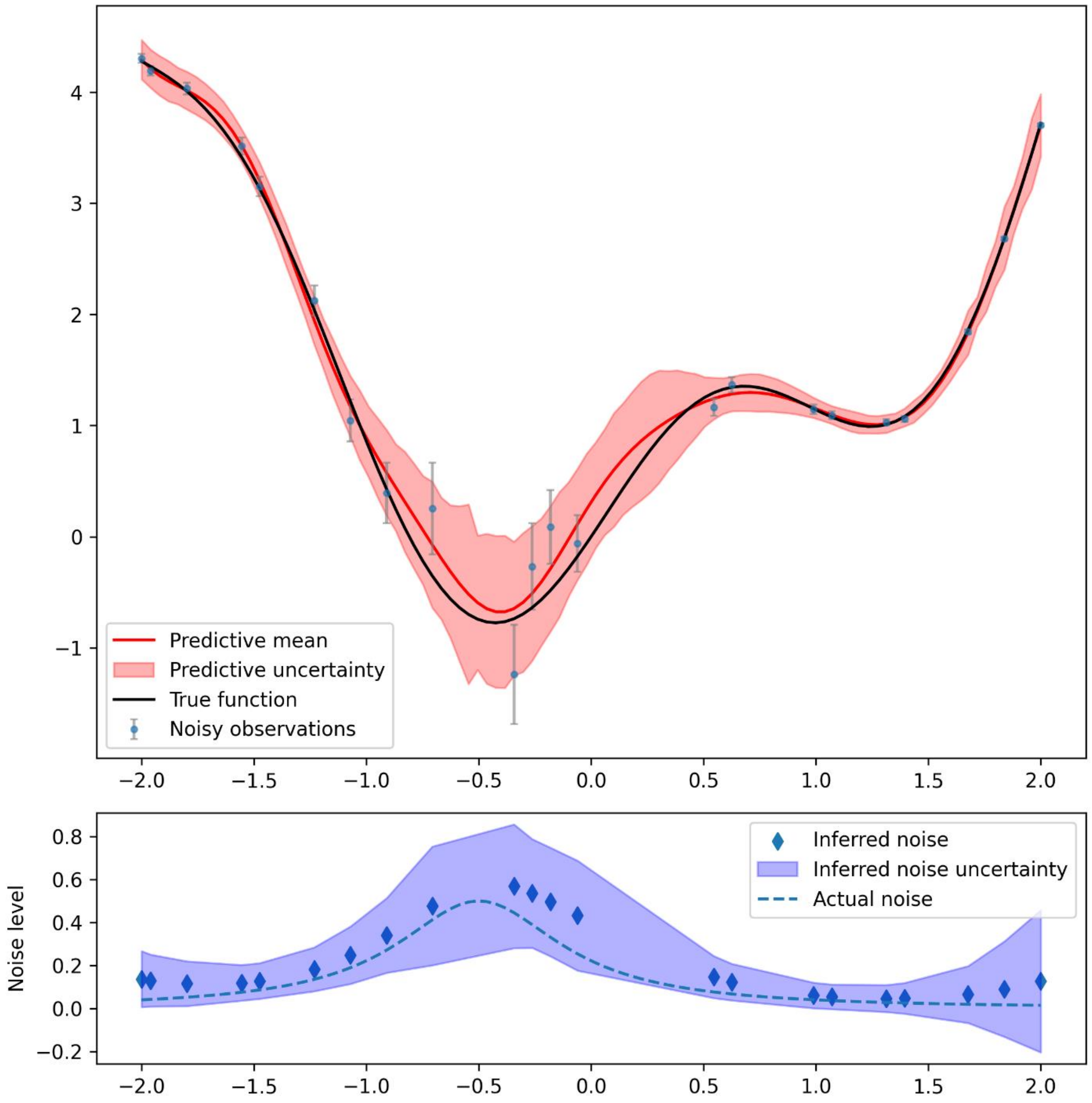
$$y(x) = f(x) + \sigma(x)$$

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).

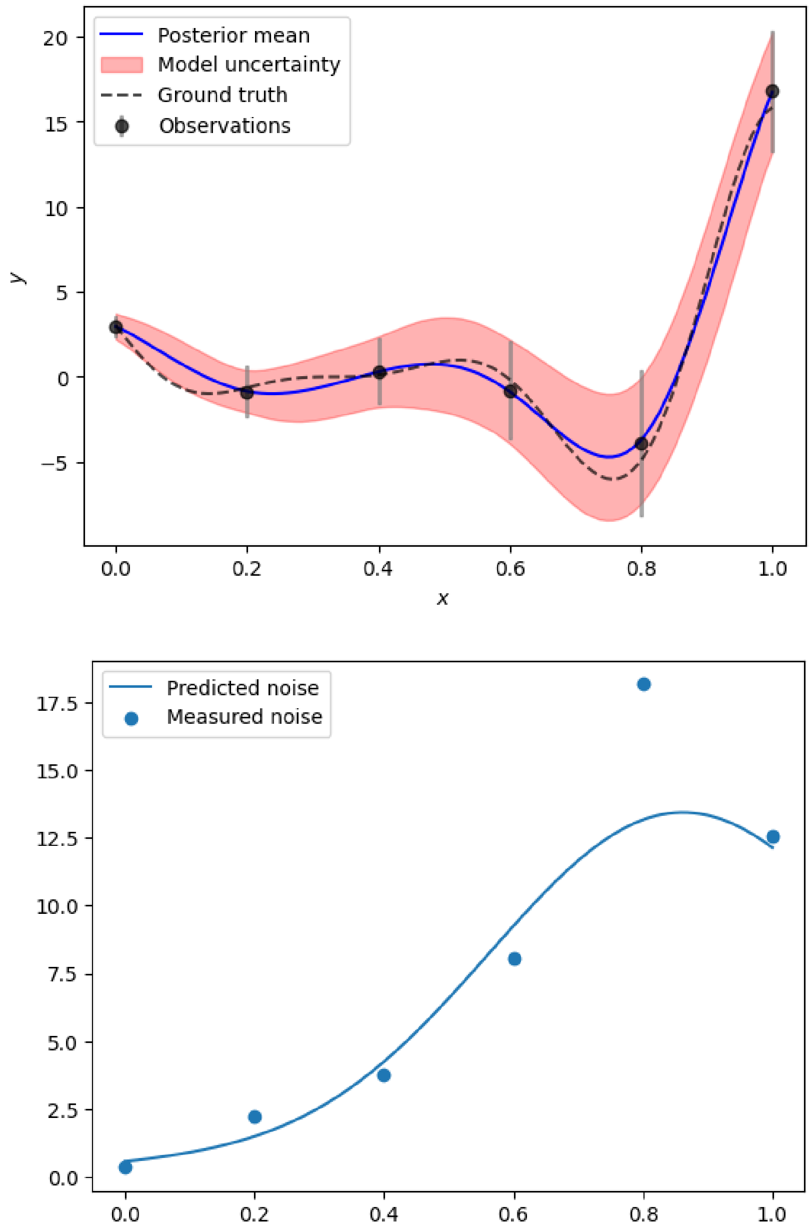
For example, it is often easier to measure multiple times (indentation curves, spectra, etc) at one location rather then move around

Alternatively, noise can be estimated from single measurement.

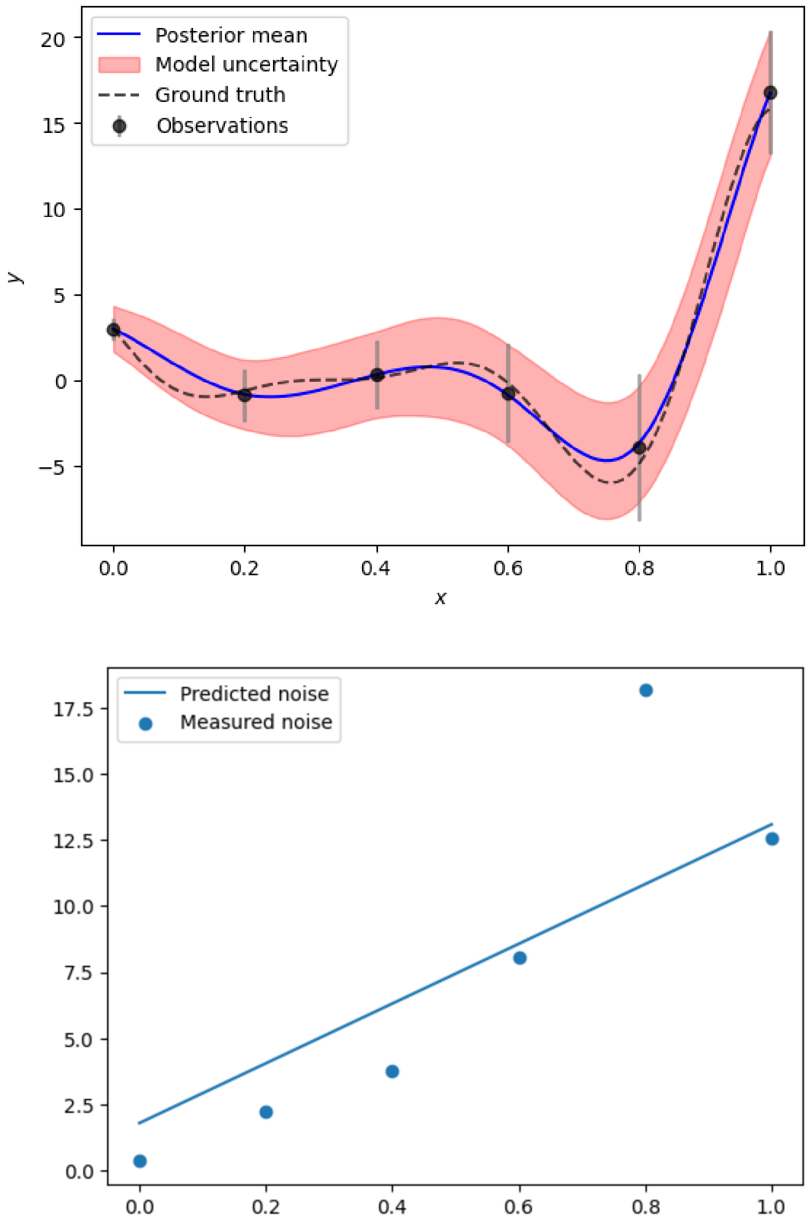
Note that we still need **noise model** even when the noise is measured, since we need to have an estimate of noise at the yet-unmeasured locations

Measured Noise GP

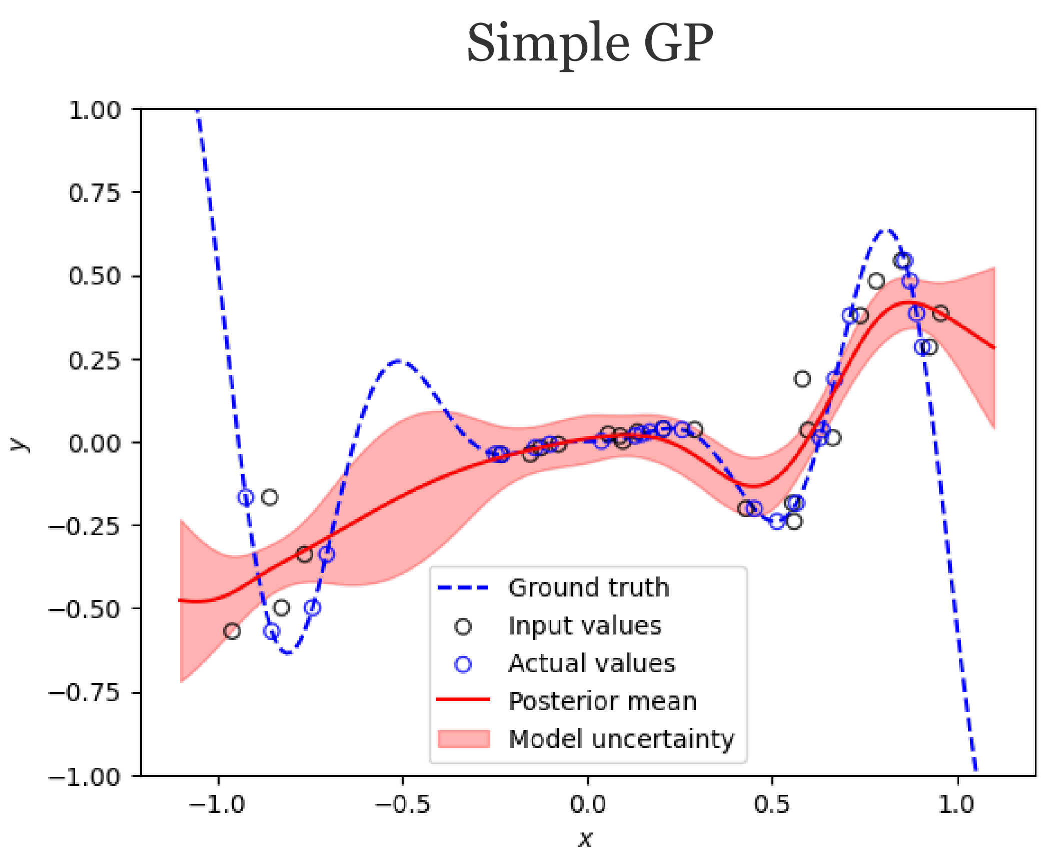
GP regression

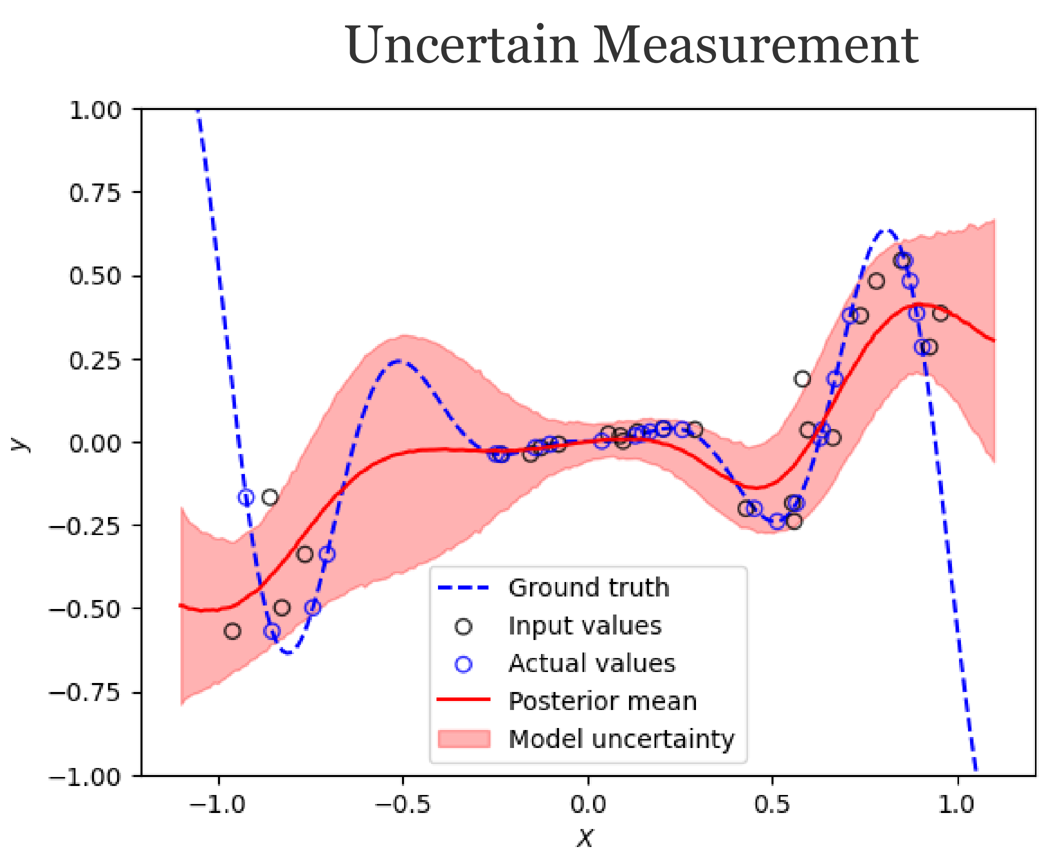


Linear regression



What if the measurement location in uncertain?





Automated Experiment:
… as a scientist…

Bayesian optimization:

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)

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 **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_*:

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

replaced with

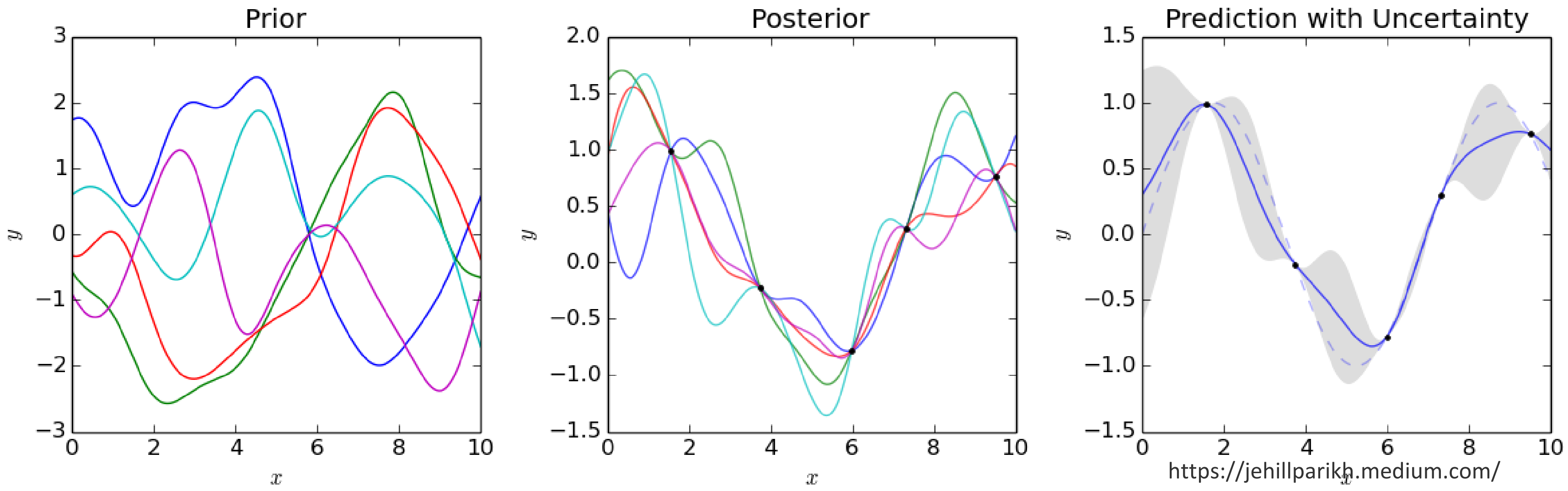
$$\mu_{\boldsymbol{\theta}^i}^{\text{post}} = \mathbf{m}(X_*) + \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}(X, X|\boldsymbol{\theta}^i)^{-1}(\mathbf{y} - \mathbf{m}(X)) \rightarrow \mu_{\boldsymbol{\Omega}^i}^{\text{post}} = \mathbf{m}(X_*|\phi^i) + \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}(X,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

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

$\boldsymbol{\Omega}^{\boldsymbol{i}} = \{\phi^i, \boldsymbol{\theta}^i\}$ 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

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

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

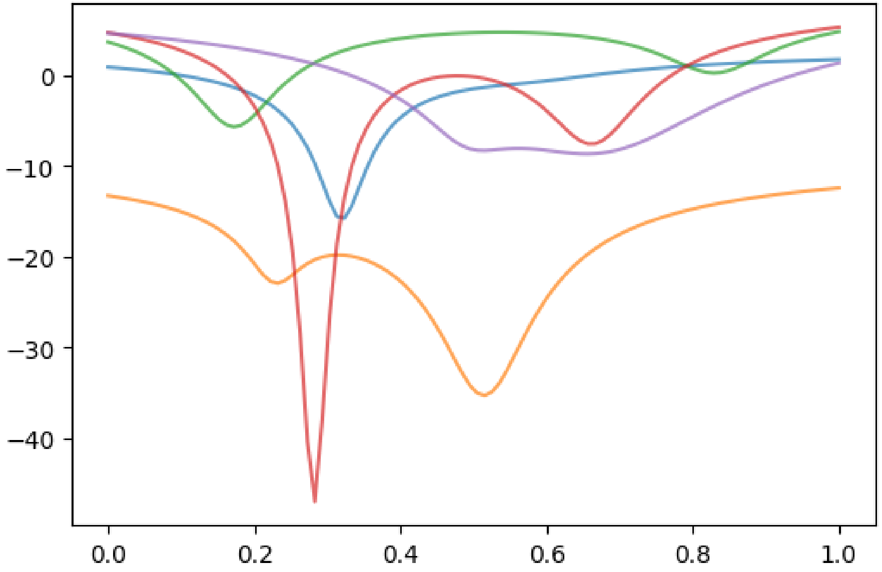
$$L_n \sim \frac{A_n}{\sqrt{(x - x_n^0)^2 + w_n^2}}$$

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

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

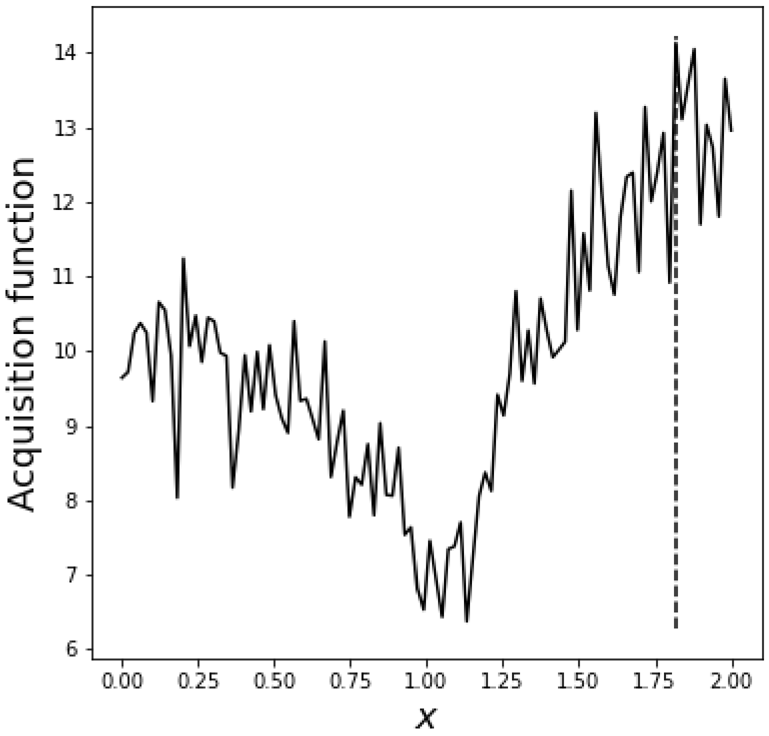
$$x_n^0 \sim Uniform(0, 1)$$

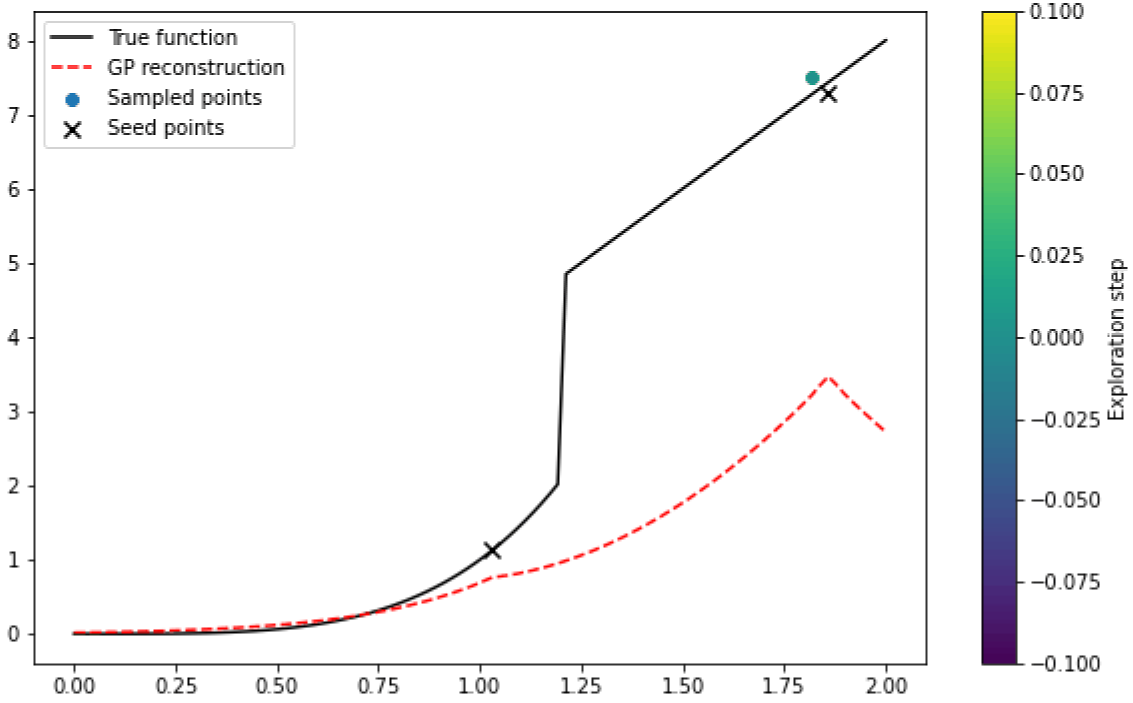
Prior predictive distribution



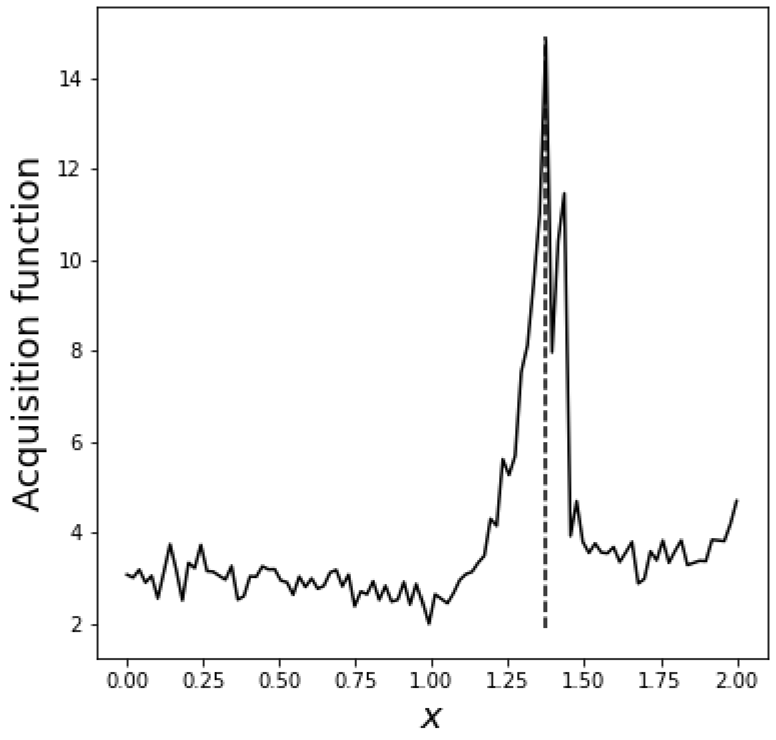
This model simply tells us that there are two minima in our data but does not assume to have any prior knowledge about their relative depth, width, or distance

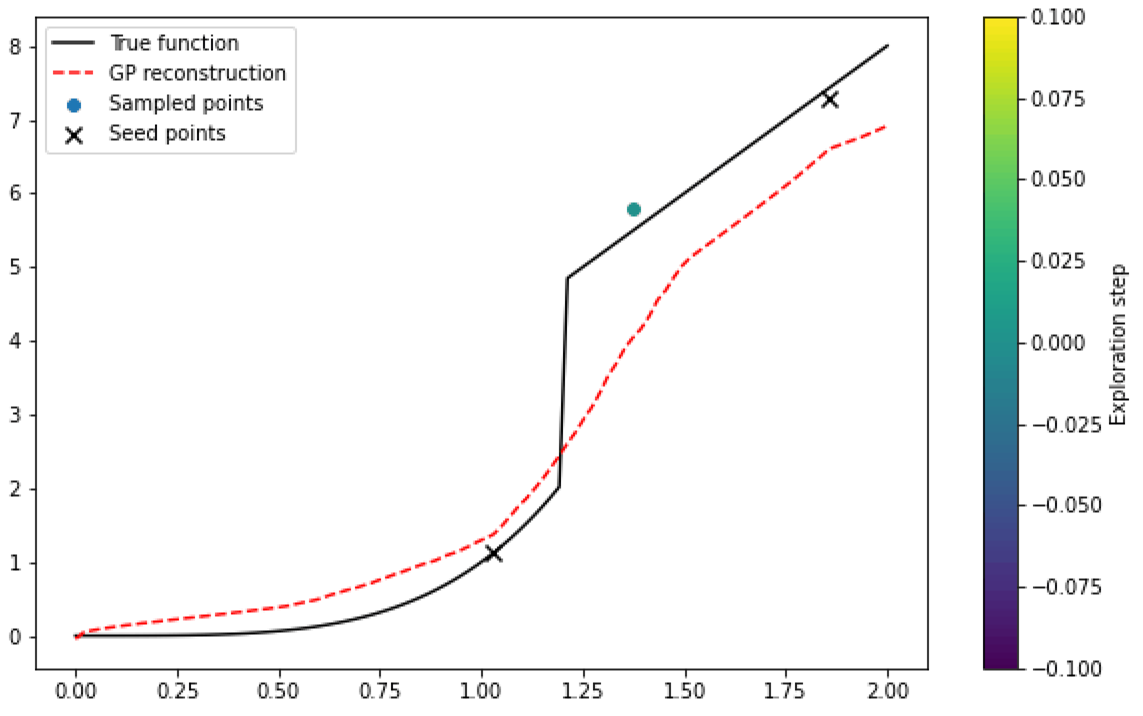
Simple GP search





Structured GP search



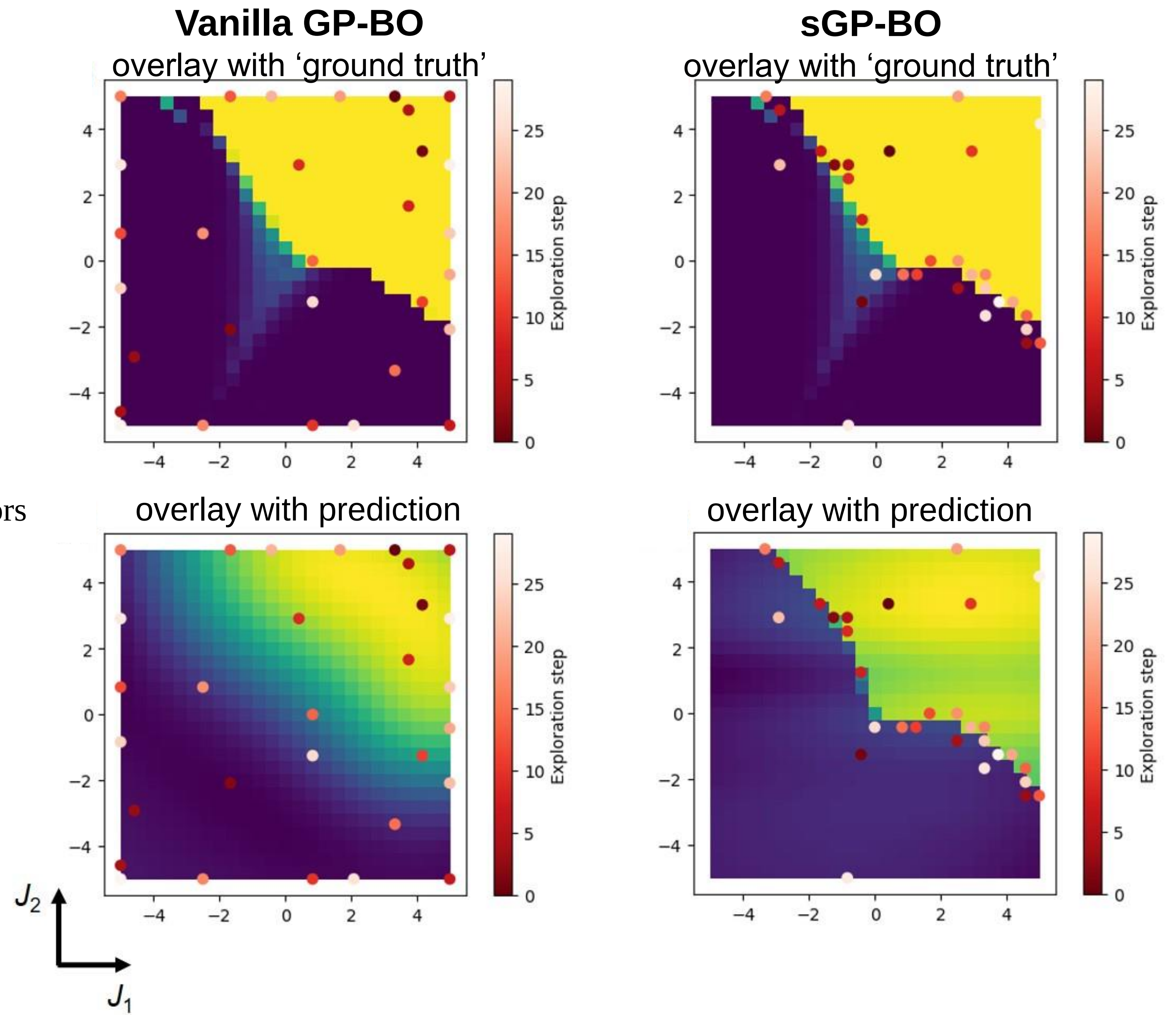


Application to Ising model

Probabilistic model

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

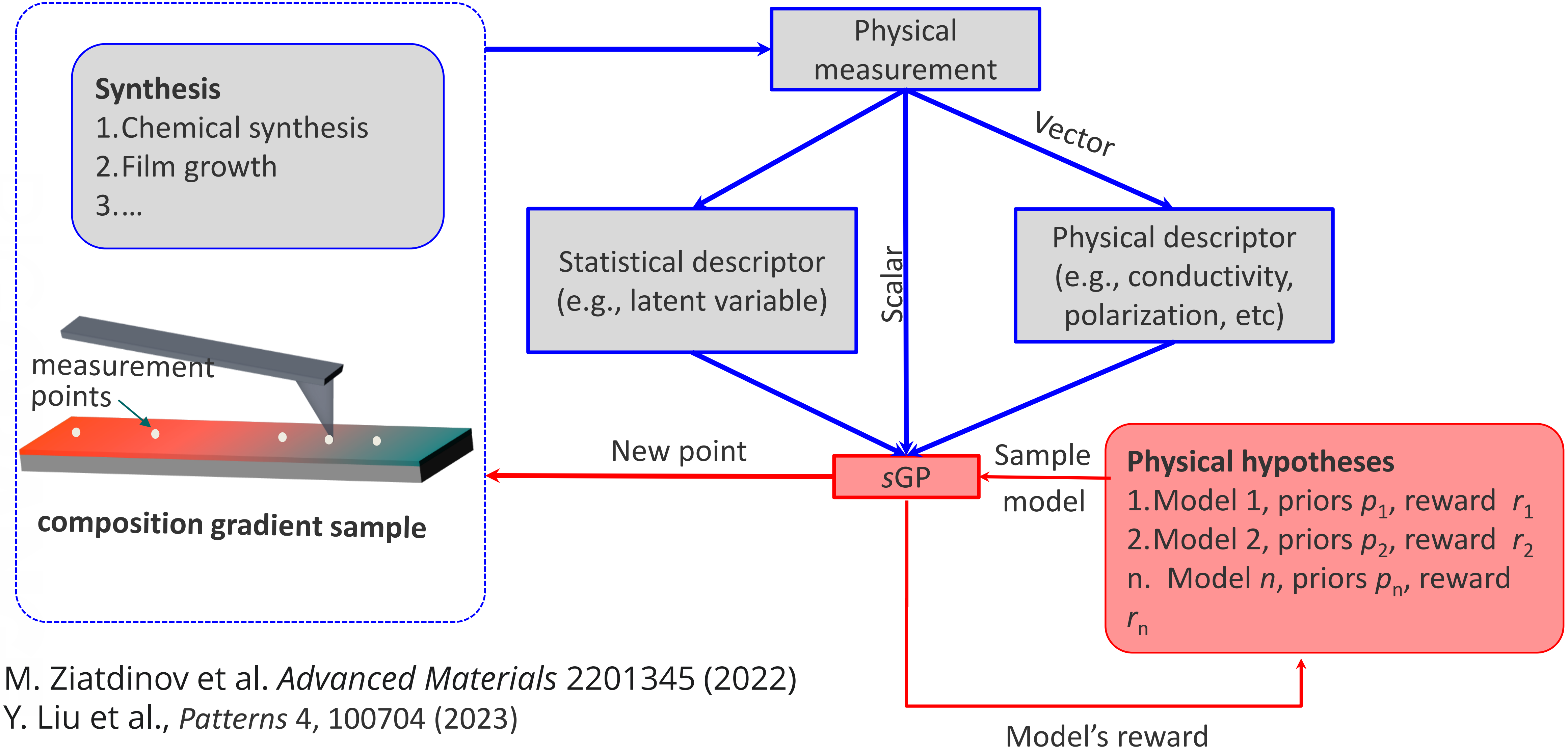
where $f(J)$ is a third-degree polynomial with normal priors on its parameters



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

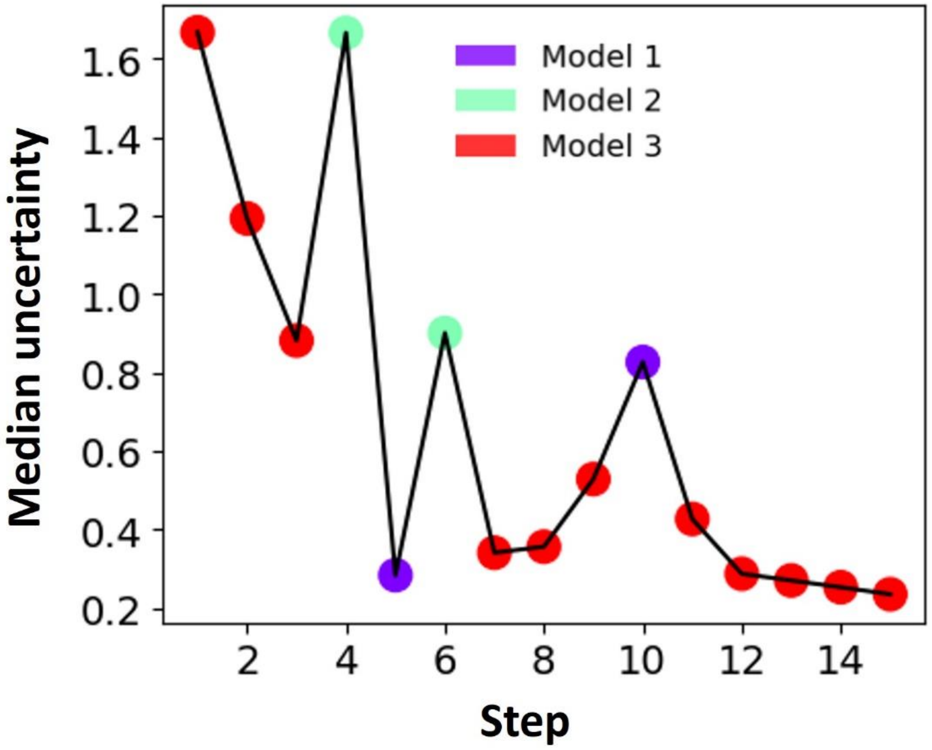


M. Ziatdinov et al. *Advanced Materials* 2201345 (2022)
Y. Liu et al., *Patterns* 4, 100704 (2023)

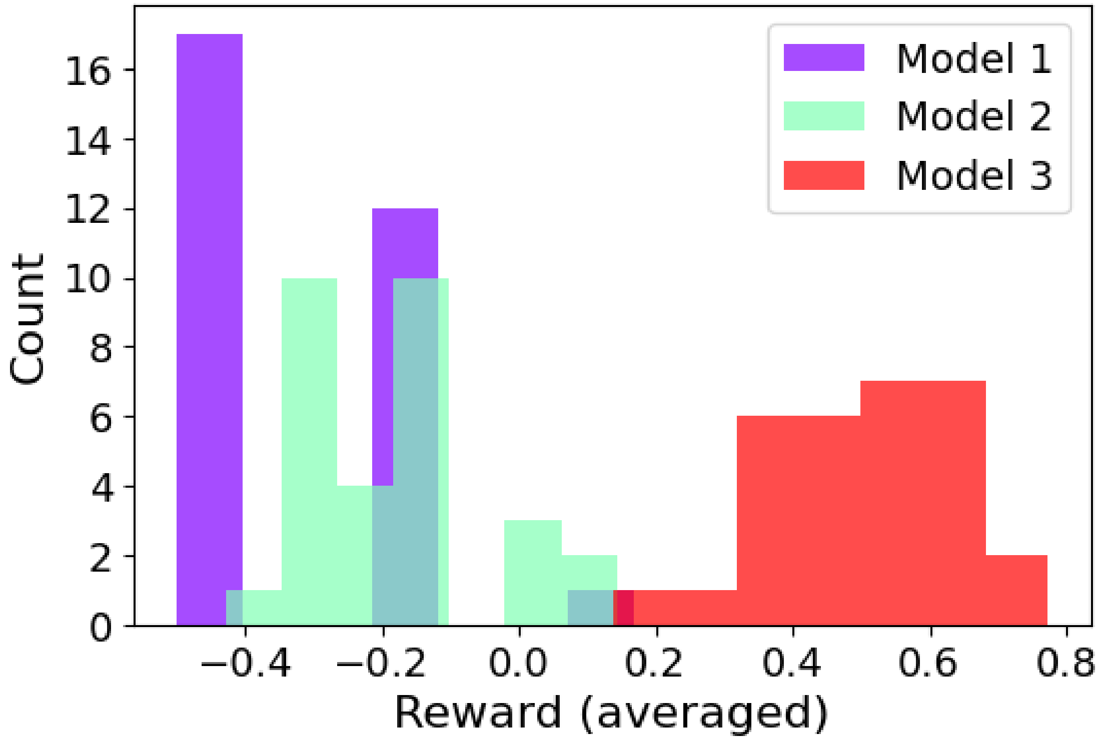
Hypothesis Learning: Synthetic data

Synthetic data represents a 1D discontinuous phase transition

Evolution of uncertainty for a single seed



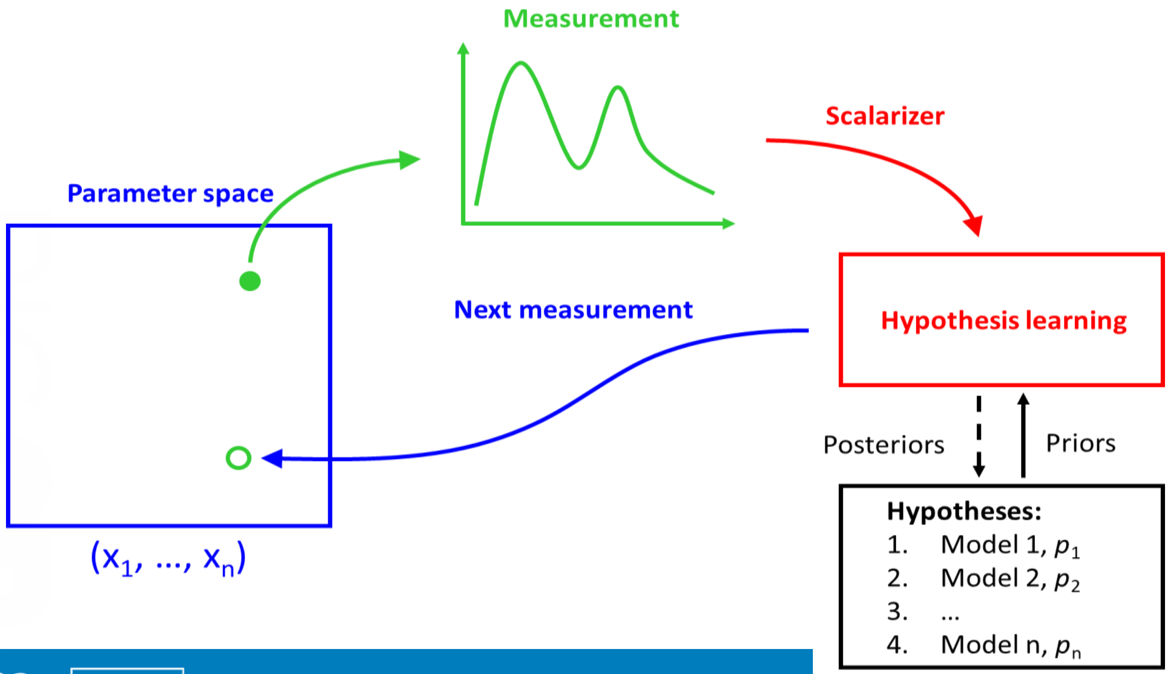
Results for 30 random seeds

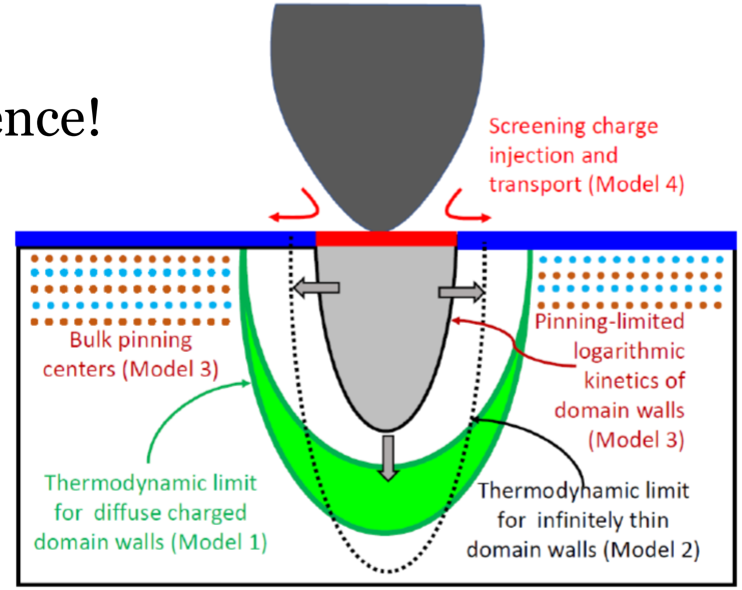


The hypothesis learning learns a correct data distribution with a small number of sparse measurements while also identifying a correct model that describes the system's behavior

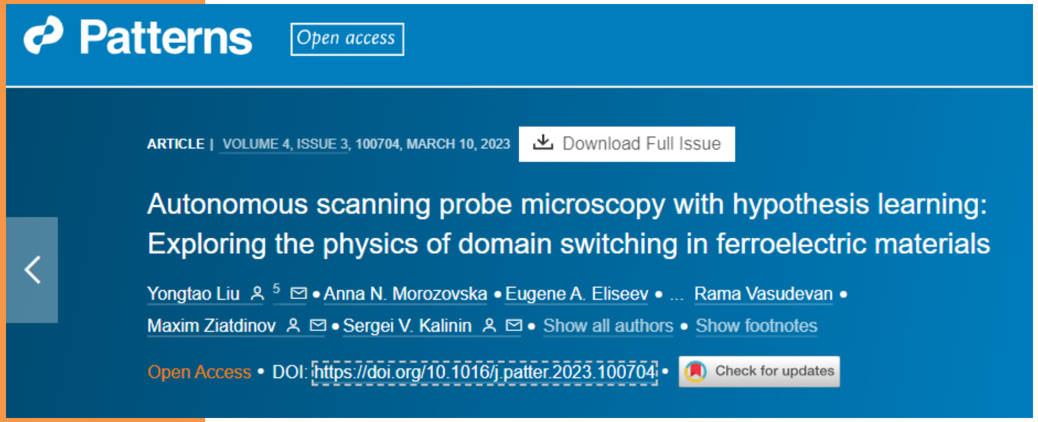
Hypothesis Learning

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





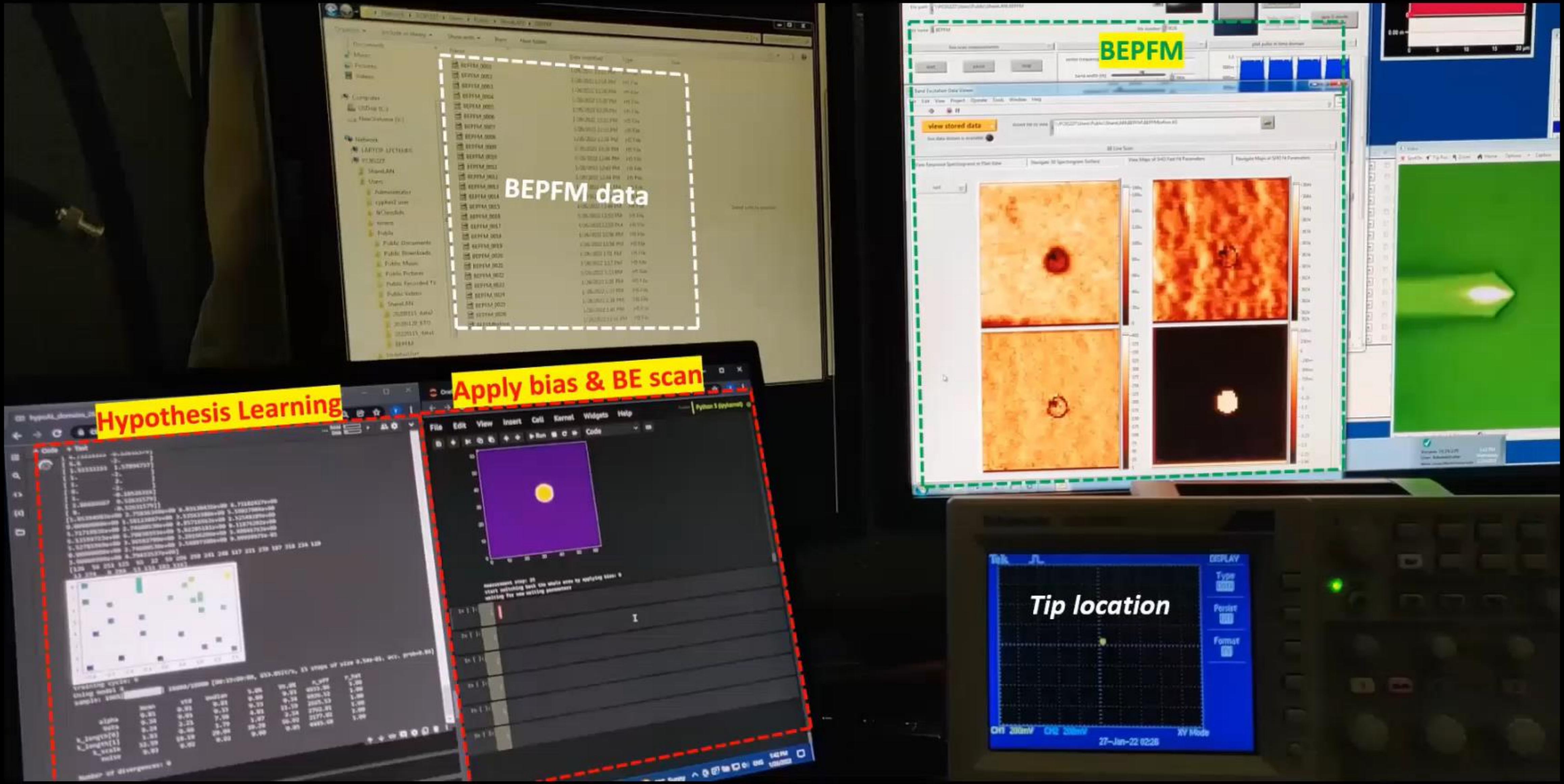
		Model Equation
Thermodynamic 1	Model I	$r(V) = r_{cr} + r_0\sqrt{\left(\frac{V}{V_c}\right)^{2/3} - 1}$
Thermodynamic 2	Model II	$r(V) = r_{cr} + r_0\sqrt[3]{\left(\frac{V}{V_c}\right)^{2} - 1}$
Wall pinning	Model III	$r(V,t) = V^{\alpha}\log\tau$
Charge injection	Model IV	$r(V,t) = V^{\alpha}\tau^{\beta}$


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

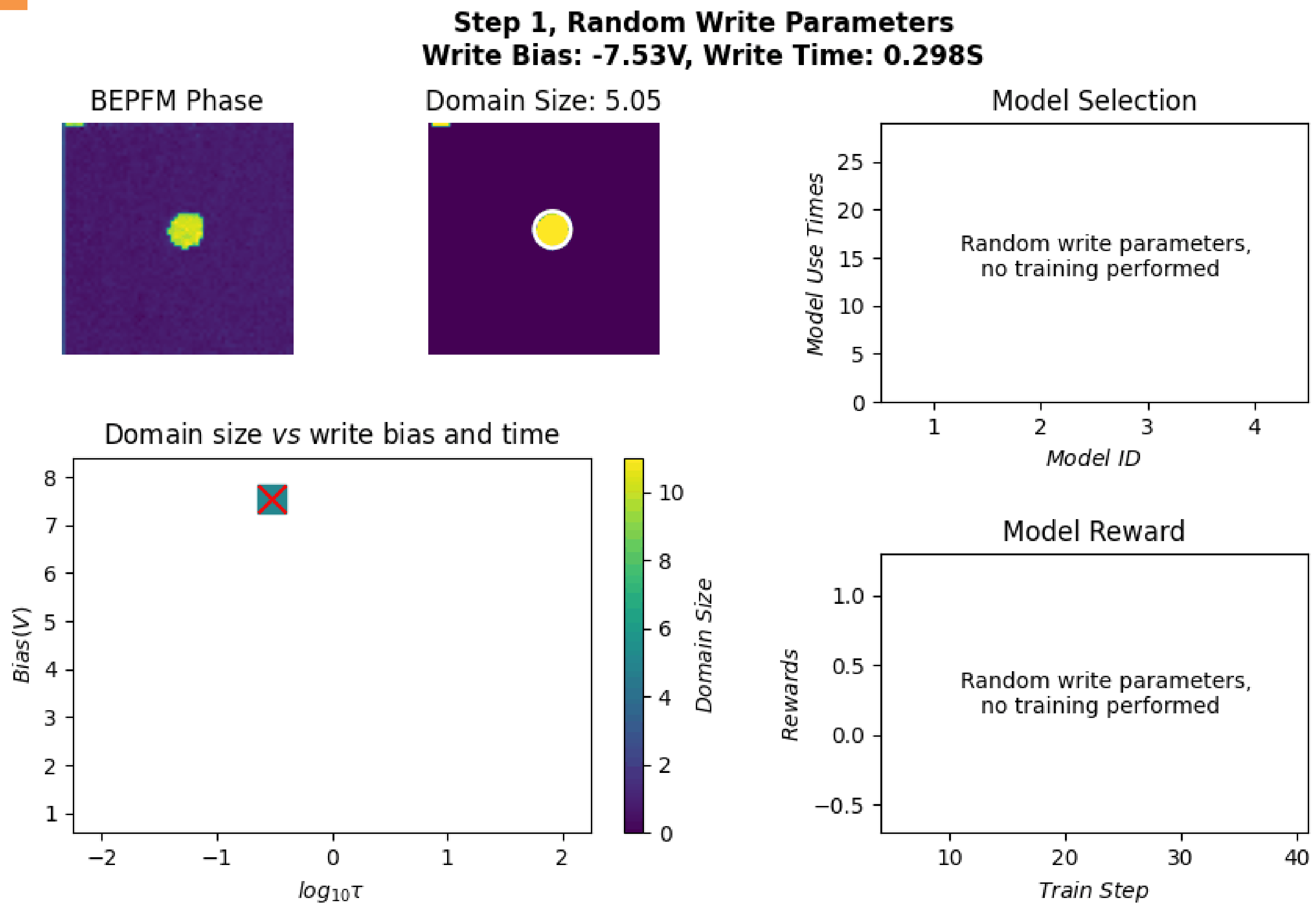
Yongtao Liu [5] • Anna N. Morozovska • Eugene A. Eliseev • … Rama Vasudevan • Maxim Ziatdinov • Sergei V. Kalinin • Show all authors • Show footnotes

Open Access • DOI: https://doi.org/10.1016/j.patter.2023.100704 • Check for updates



BEPFM data
BEPFM
Hypothesis Learning
Apply bias & BE scan
Tip location

Hypothesis learning in action



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

- 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!

Combinatorial Synthesis

ADVANCED MATERIALS

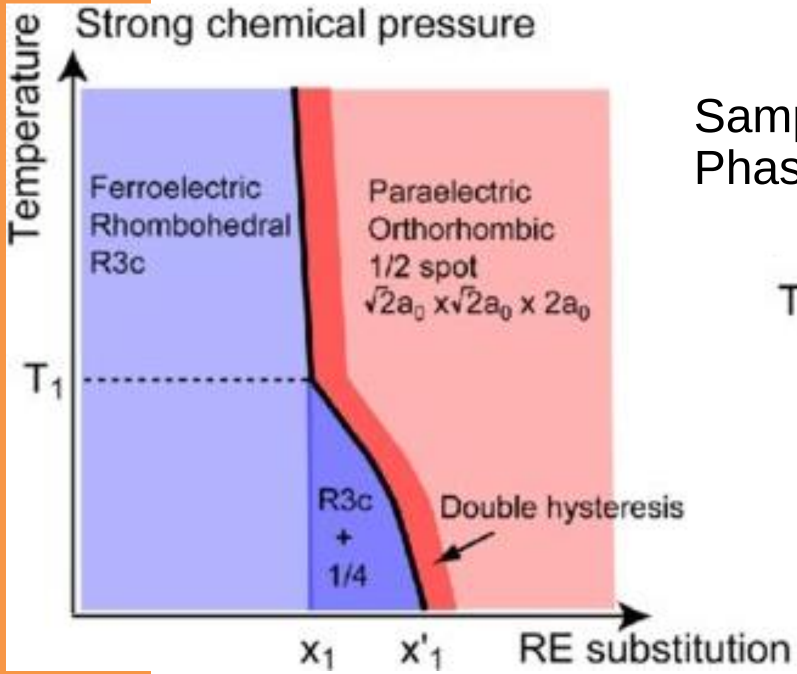
Research Article

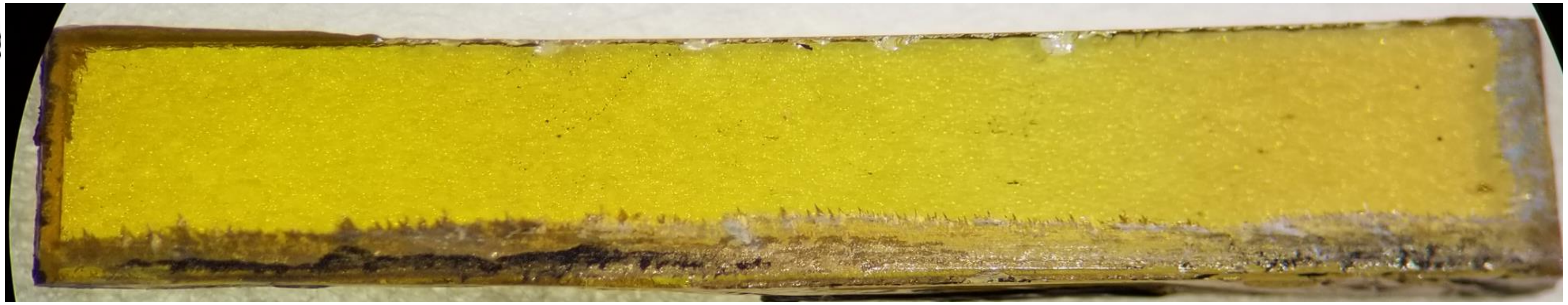
Hypothesis Learning in Automated Experiment: Application to Combinatorial Materials Libraries

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

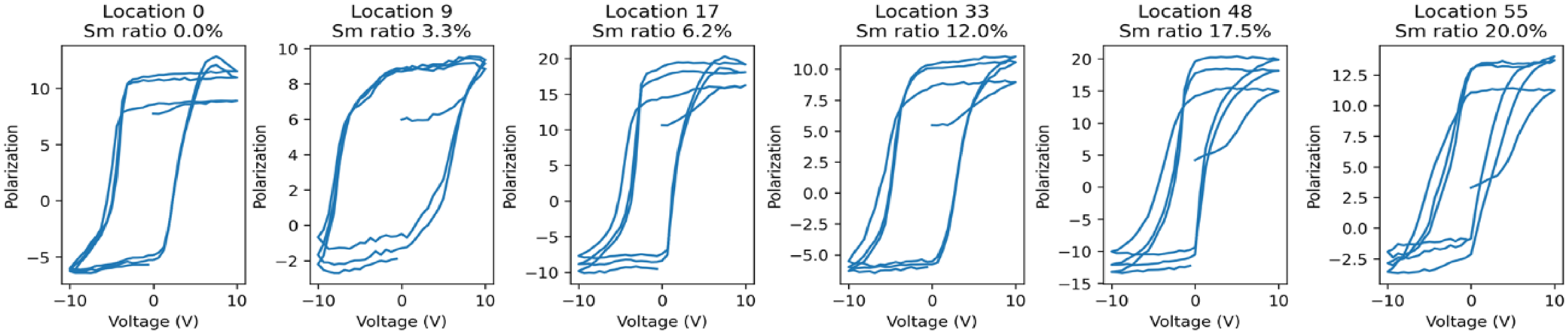
First published: 12 March 2022 | https://doi.org/10.1002/adma.202201345 | Citations: 17

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





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



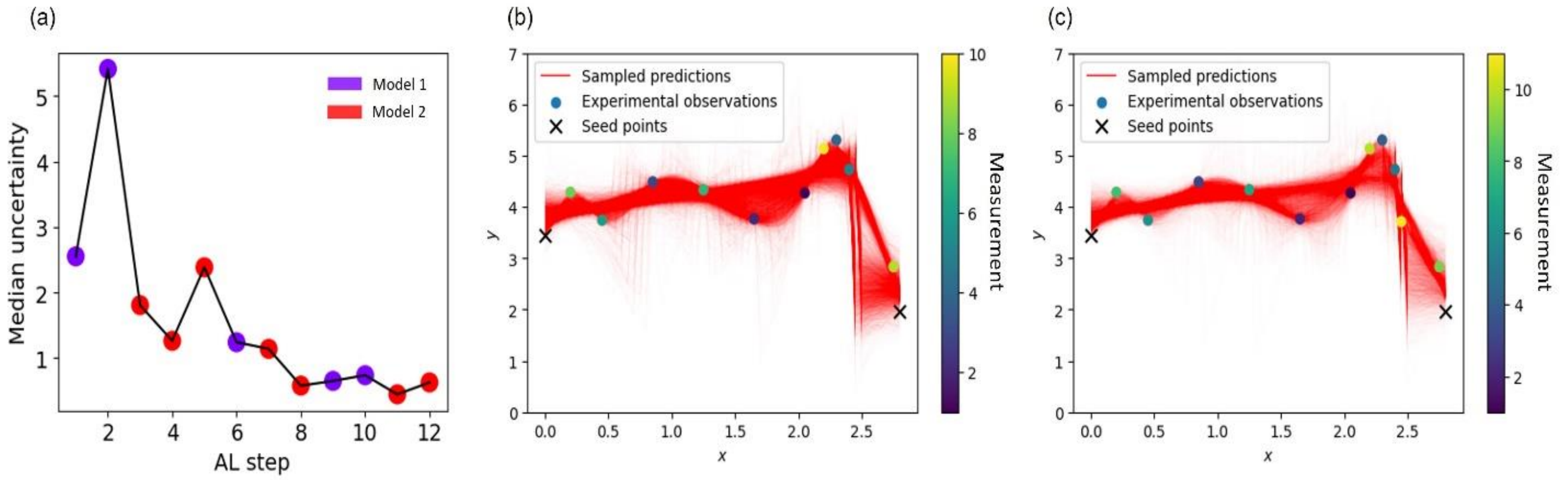
Hypothesis Selection for Ferroelectric

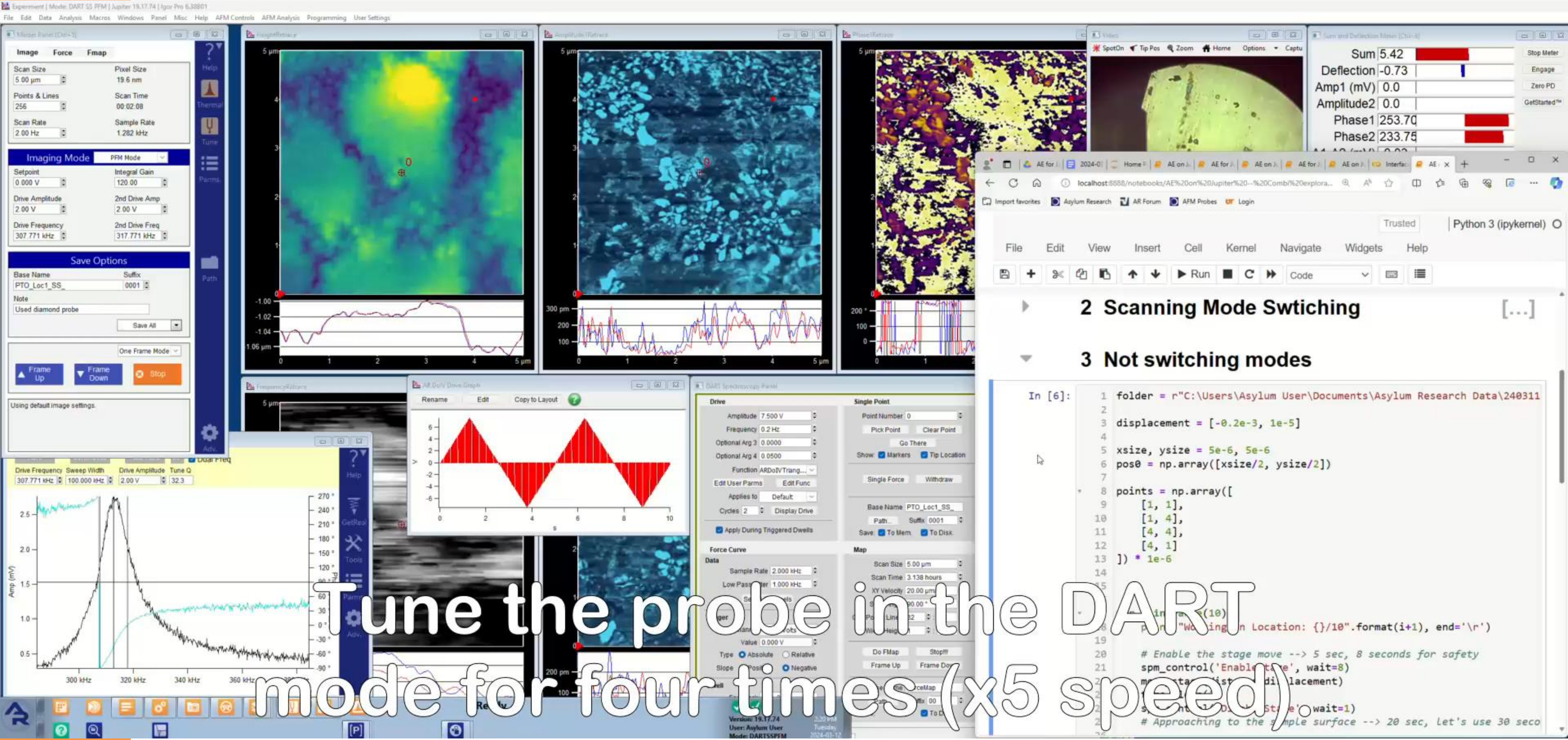
Model 1 (second order phase transition):

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

Model 2 (first order phase transition):

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




2 Scanning Mode Swtiching
3 Not switching modes
Tune the probe in the DART mode for four times (x5 speed)