THE UNIVERSITY of TENNESSEE KNOXVILLE

Structured Gaussian Processes and Hypothesis Learning

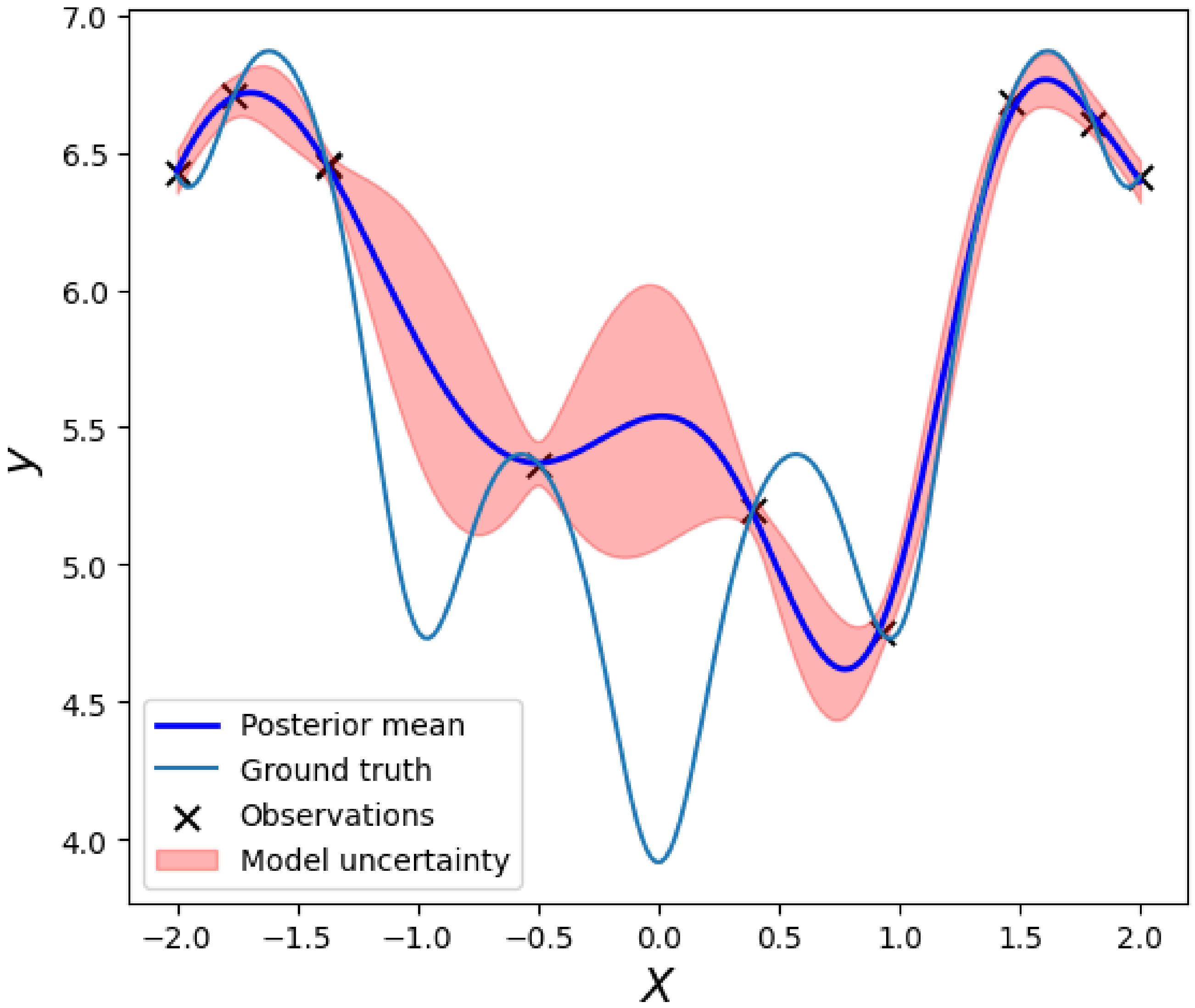
Sergei V. Kalinin

What have we learned from lecture 2

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

- Bayesian Optimization
- Bayesian Optimization based on Gaussian Process

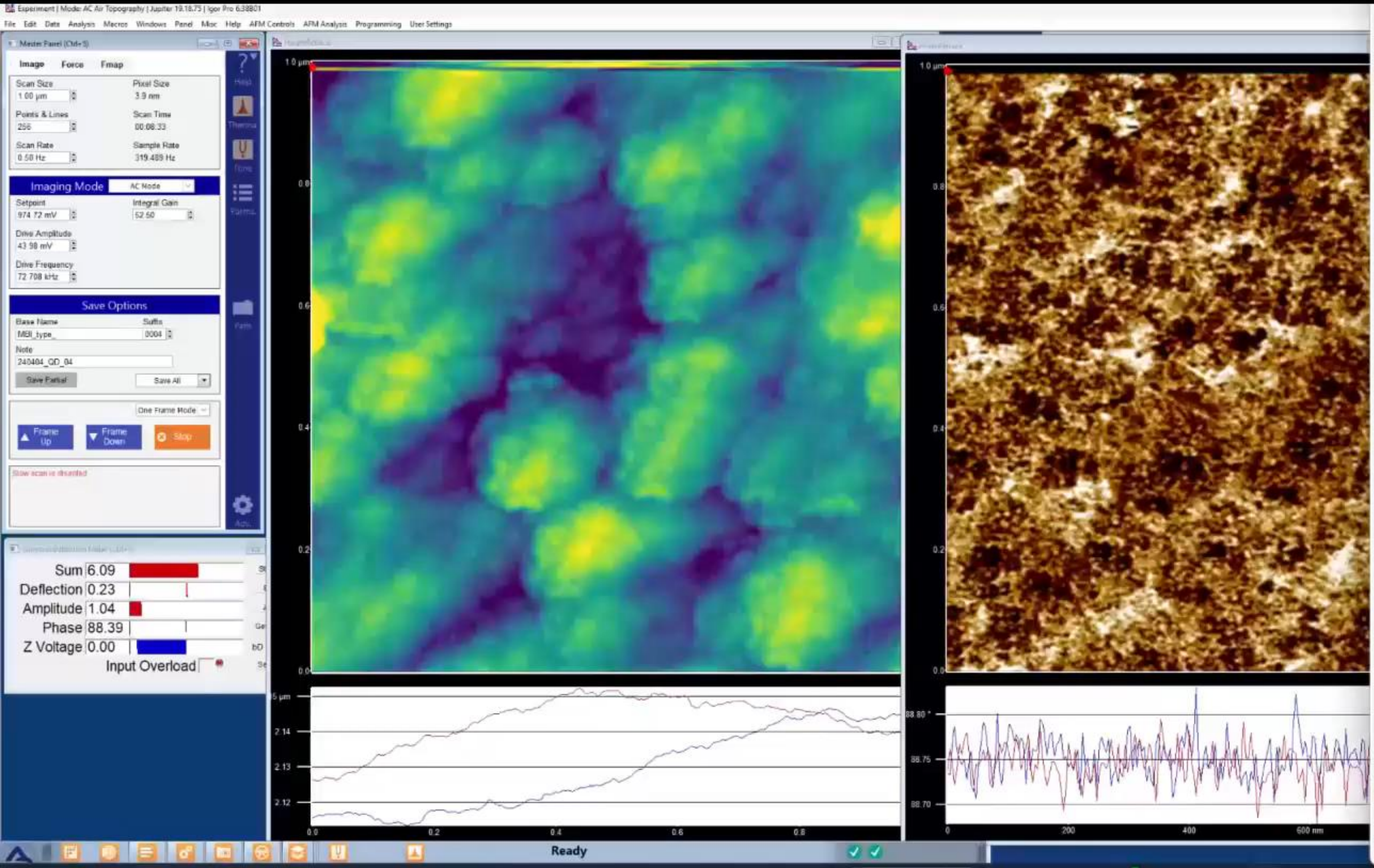
- Acquisition Functions
- Cost-award BO



We can introduce noise in GP models

- Heteroscedastic GP: learn function and noise
- Measured Noise GP: measure function and noise
  – Needs model for noise prediction at unmeasured locations
- Noise in position in parameter space

Need to balance priors based on extant knowledge;
do **not** treat these as optimization parameters



BO functions

+ 3 cells hidden

Run the BO

[97]:

```
# Make the grid for exploration
save_name = 'AR_Height/240404_QD_04_'

x = []

# v_ac
x1_min = 0.3e-3
x1_max = 0.05

# setpoint
x2_min = 0.1
x2_max = 0.9

factor = 248 / 10.07

offset = read_igor(key=['FreeAirPhase'], connection=connection)[0] - 90
print(offset)

x1 = np.linspace(x1_min, x1_max, num=100)
x2 = np.linspace(x2_min, x2_max, num=100)

d1, d2 = len(x1), len(x2)

for i in range(len(x1)):
    for j in range(len(x2)):
        x.append([x1[i], x2[j]])

x = np.asarray(x, dtype=np.float32)
X = torch.from_numpy(x)
```

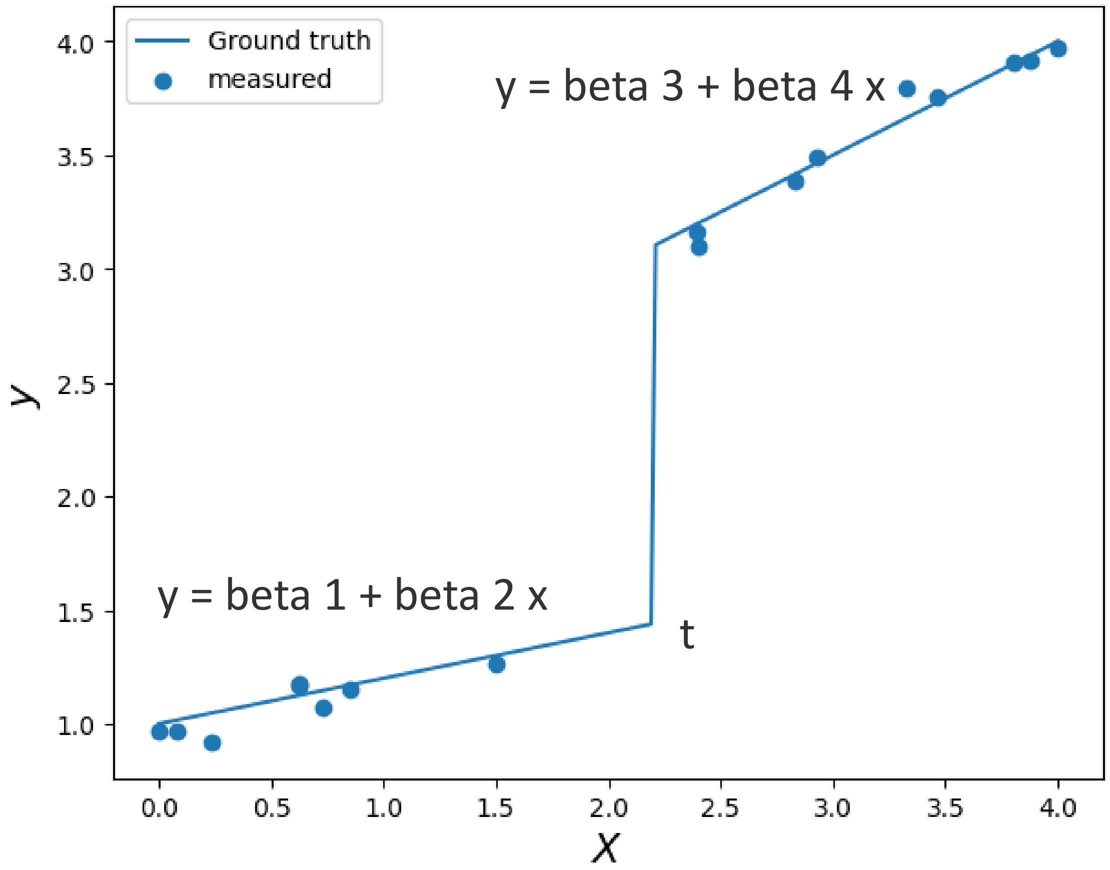
```
-0.9607010000000002
```

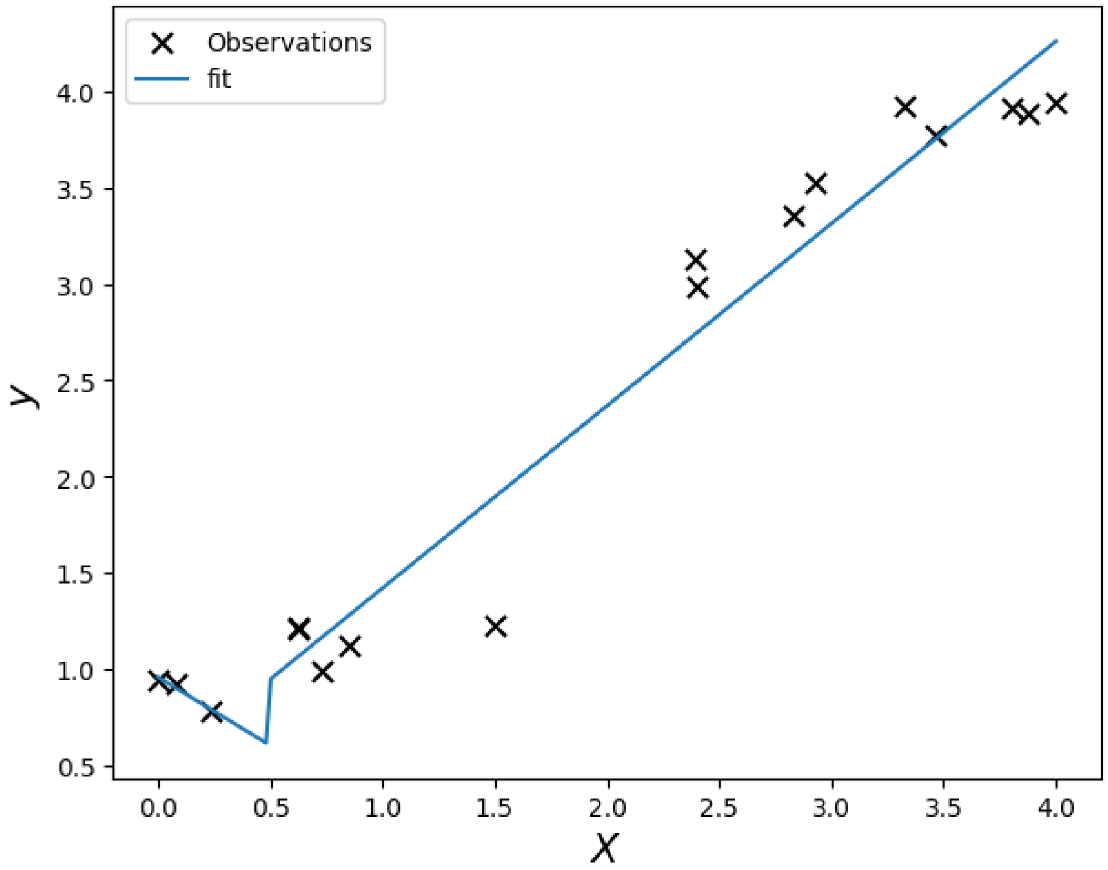
[98]:

```
X_measured, X_unmeasured, y_measured, global_min = generate_seed(num=10, factor=factor,
```

Python notebook is running on supercomputer -- ISSACs at UTK (x3 speed).

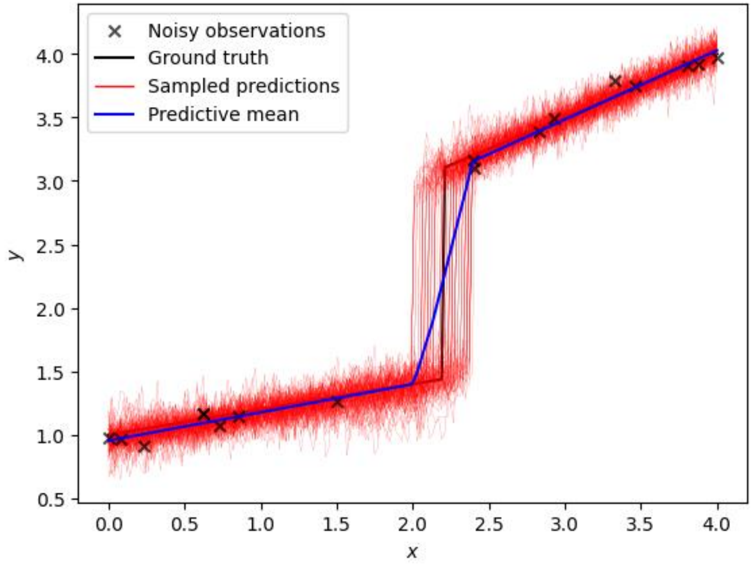
Least Square Fits can be a problem

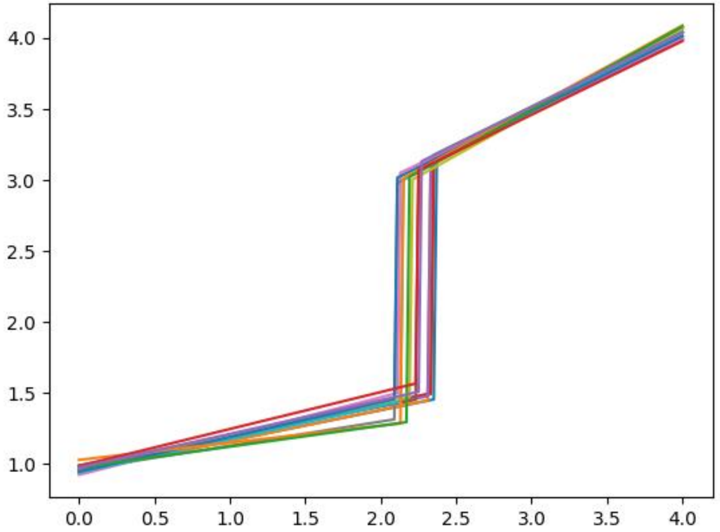


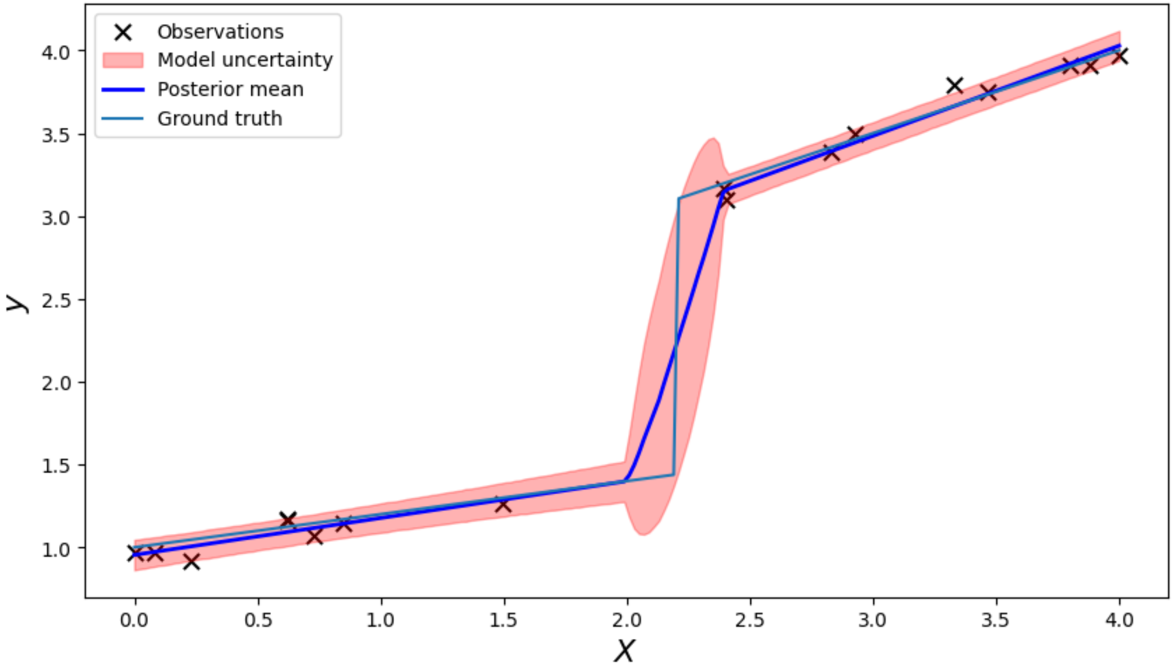


- No way to incorporate prior knowledge
- Convergence to metastable minima
- No feedback when gradient descent gets stuck

Bayesian Inference



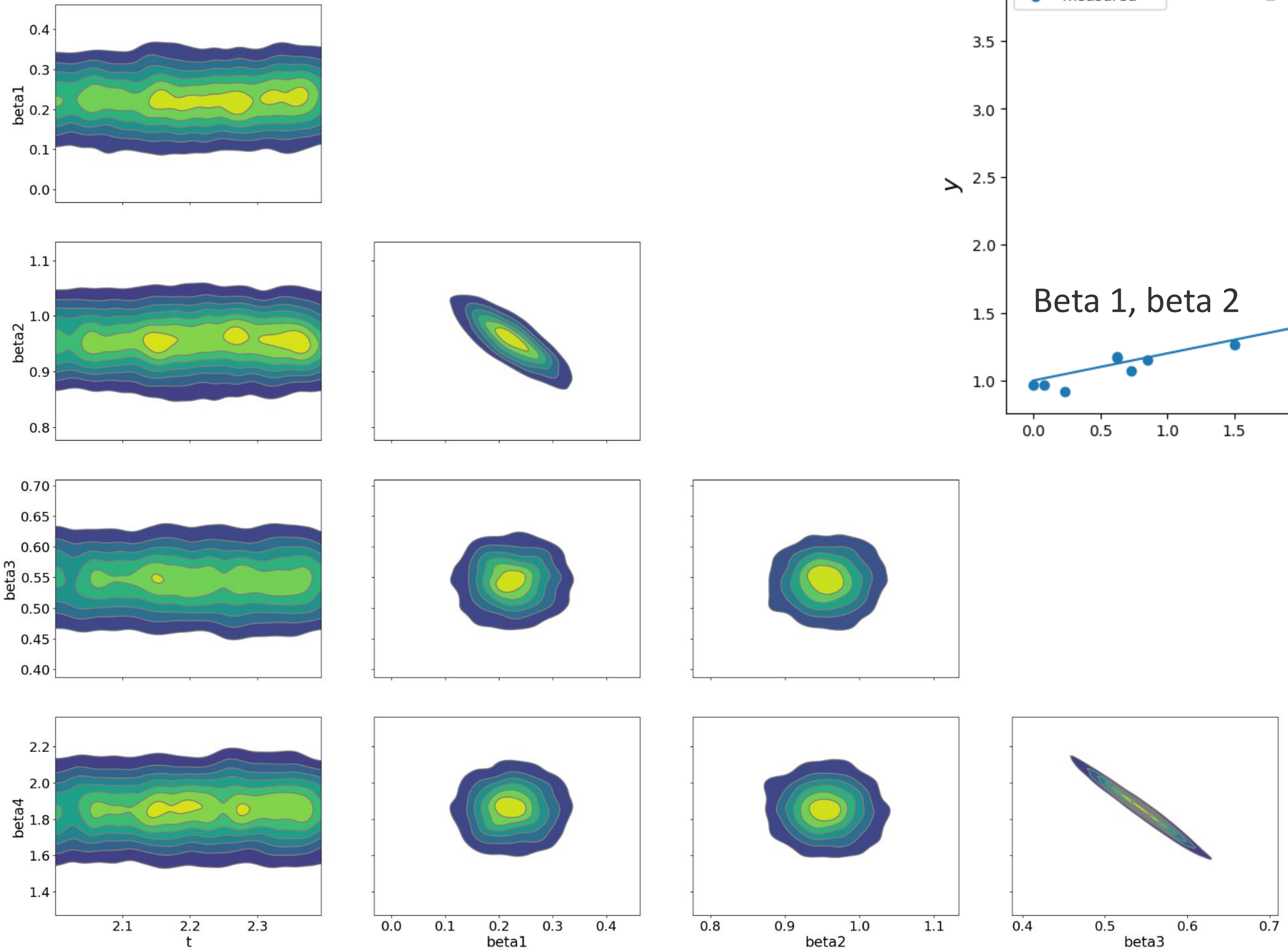


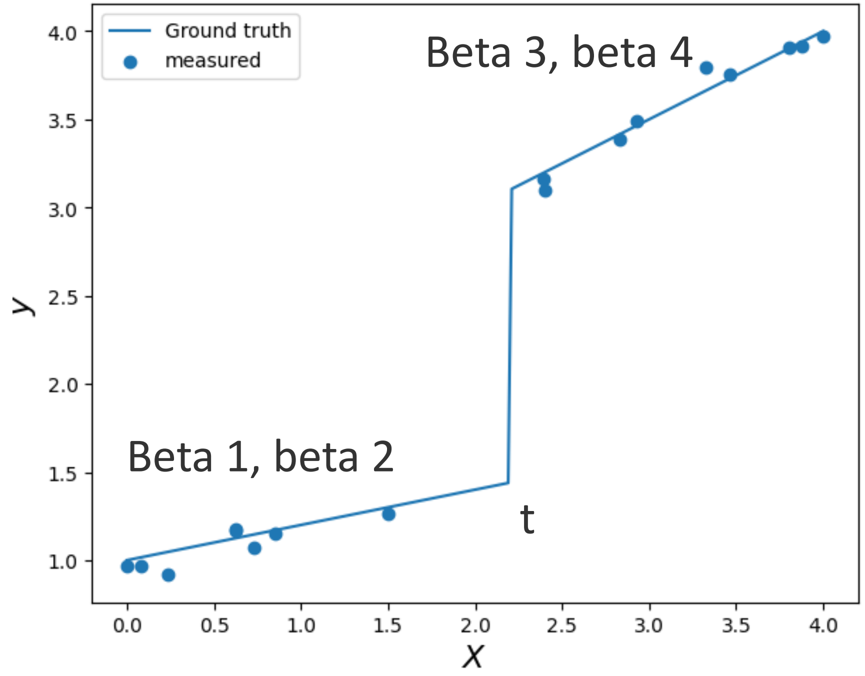


Bayesian inference:

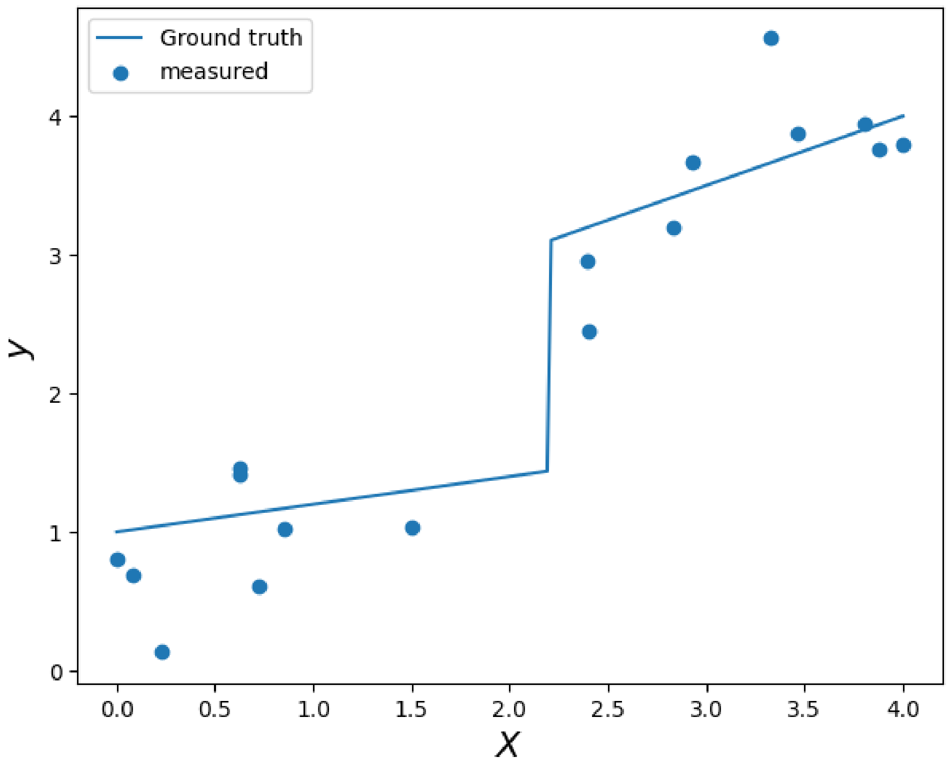
- Priors on parameter values
- Sampling values or mean functions
- Predictive mean and uncertainty
- Posteriors on parameter values

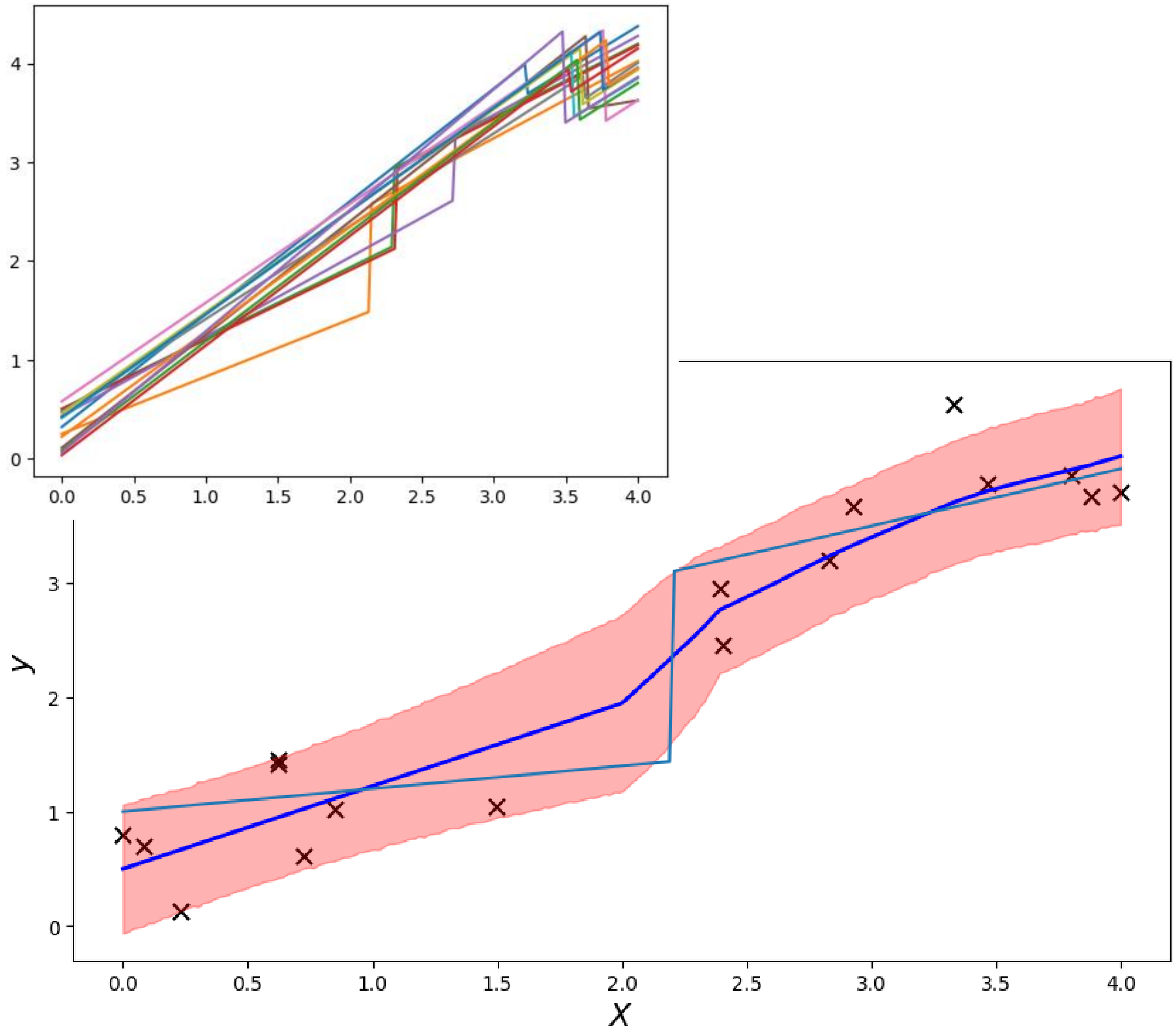
Bayesian Inference



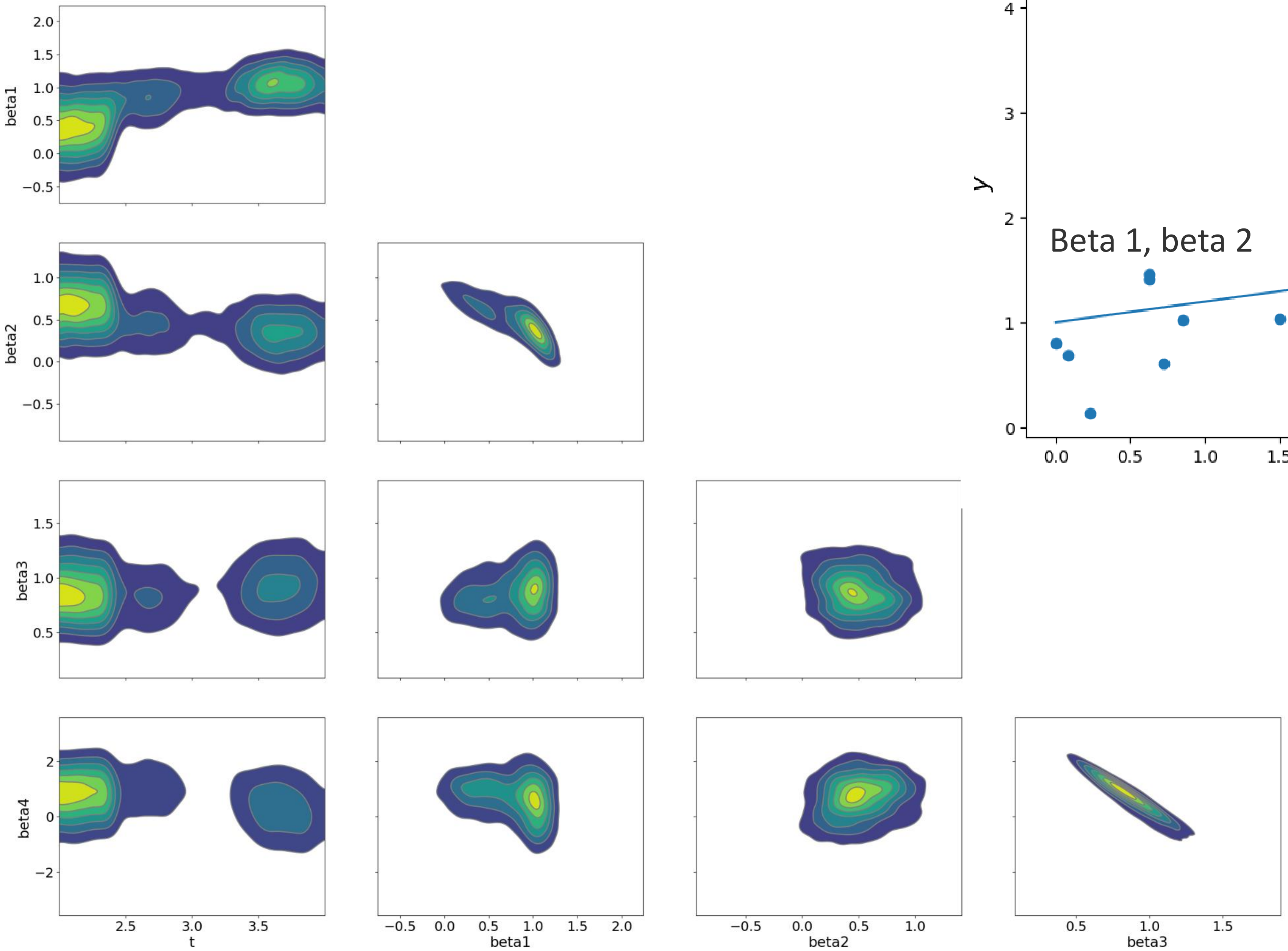


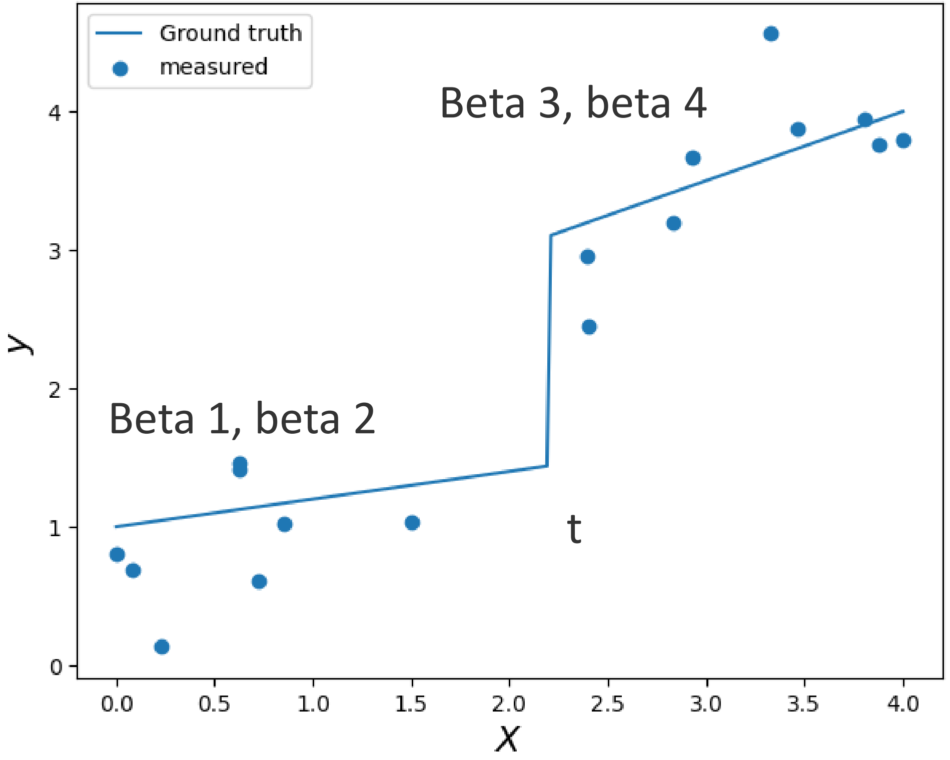
Bayesian Inference



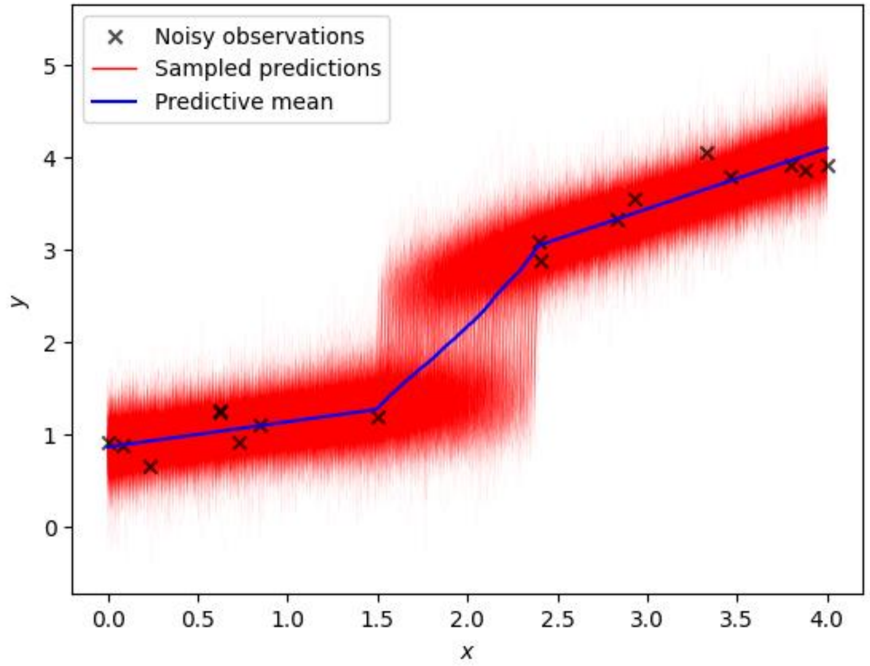


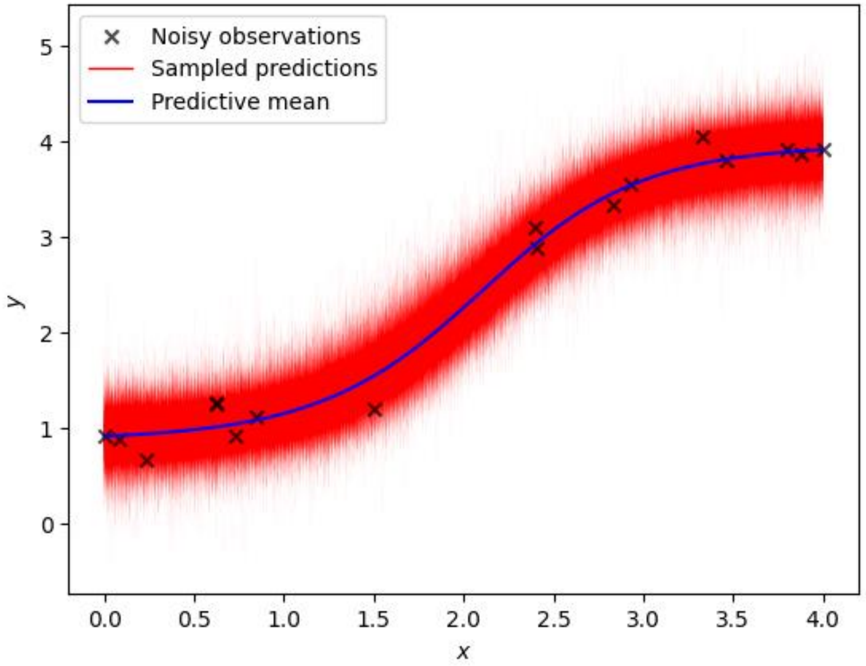
Bayesian Inference





Comparison of models: WAIC






```
1 az.compare(mcmc_all, ic="waic")
```

	rank	elpd_waic	p_waic	elpd_diff	weight	se	dse	warning	scale
piecewise	0	1.305878	3.207341	0.000000	5.027393e-01	2.217737	0.000000	True	log
sigmoidal	1	0.845156	4.073488	0.460722	4.972607e-01	2.829716	2.693397	True	log
linear2	2	-6.792783	2.312571	8.098661	0.000000e+00	2.842782	3.265961	True	log
power_law2	3	-7.820722	3.031942	9.126600	0.000000e+00	2.667691	3.113230	True	log
exponential1	4	-9.072673	2.661915	10.378551	0.000000e+00	2.251935	2.767500	True	log
linear1	5	-9.886713	1.539532	11.192591	9.836576e-14	3.207155	3.532880	True	log

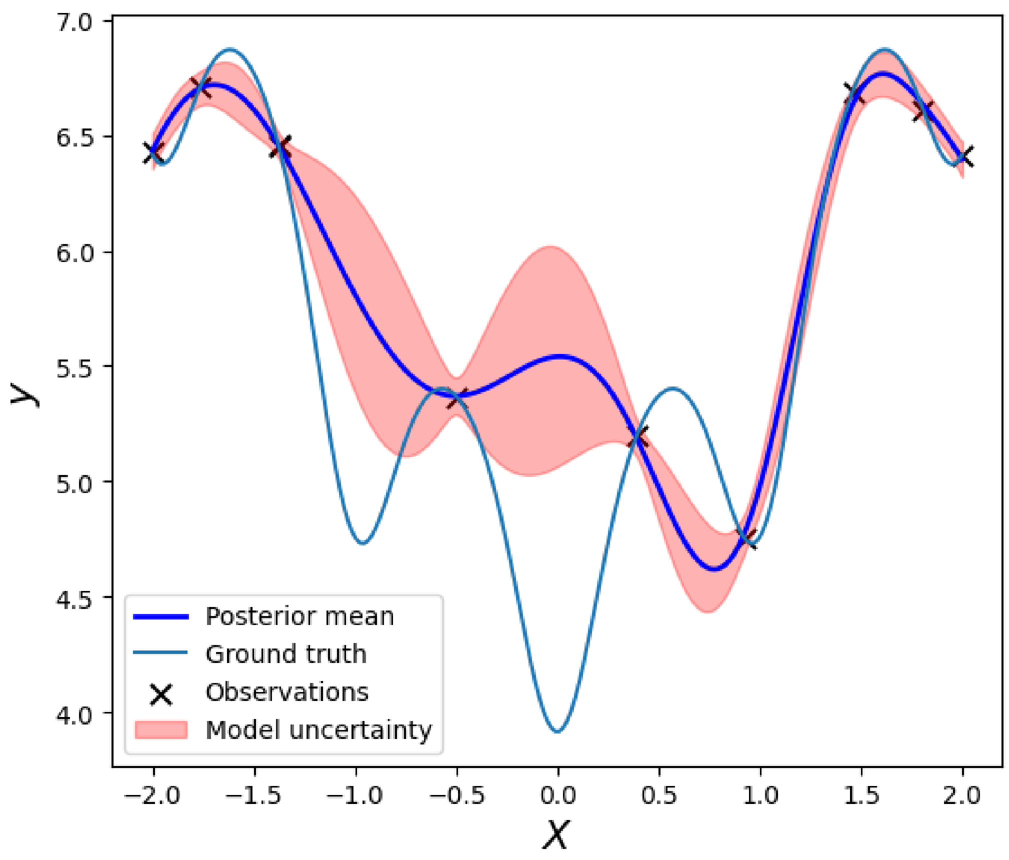
Automated Experiment:

… 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)

- 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 **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

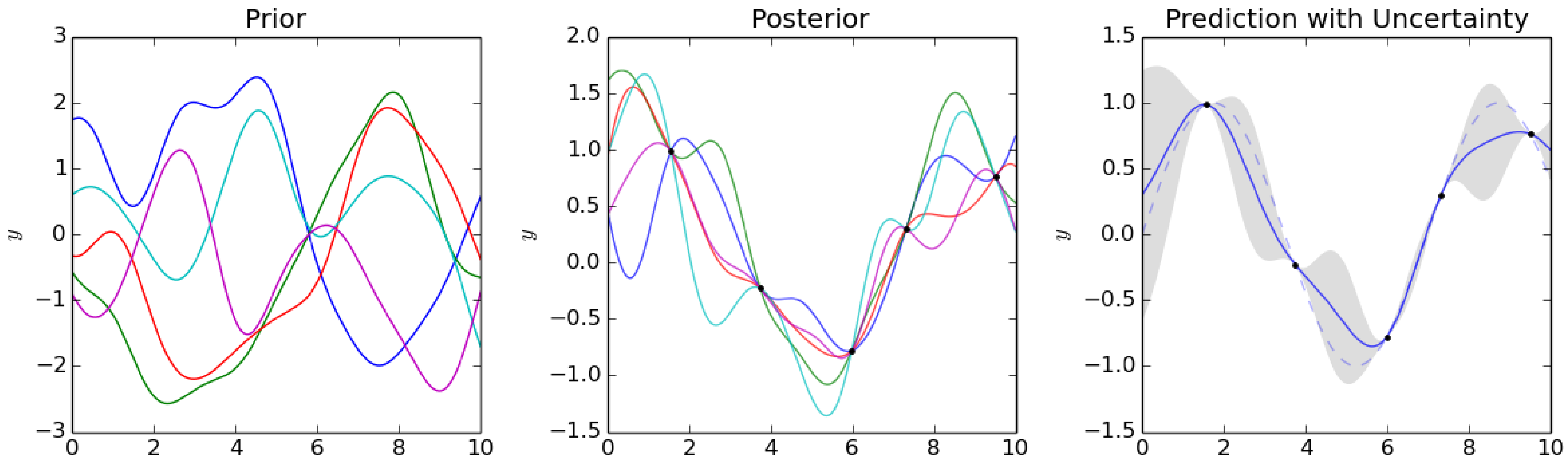
$$\mu_{\boldsymbol{\theta}^i}^{\text{post}} = \mathbf{m}(X_*) + \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}(X, X|\boldsymbol{\theta}^i)^{-1}(\mathbf{y} - \mathbf{m}(X)) \rightarrow \mu_{\boldsymbol{\Omega}^i}^{\text{post}} = \mathbf{m}(X_*|\phi^i) + \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}(X,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

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

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



https://jehillparikh.medium.com/

Probabilistic model

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

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

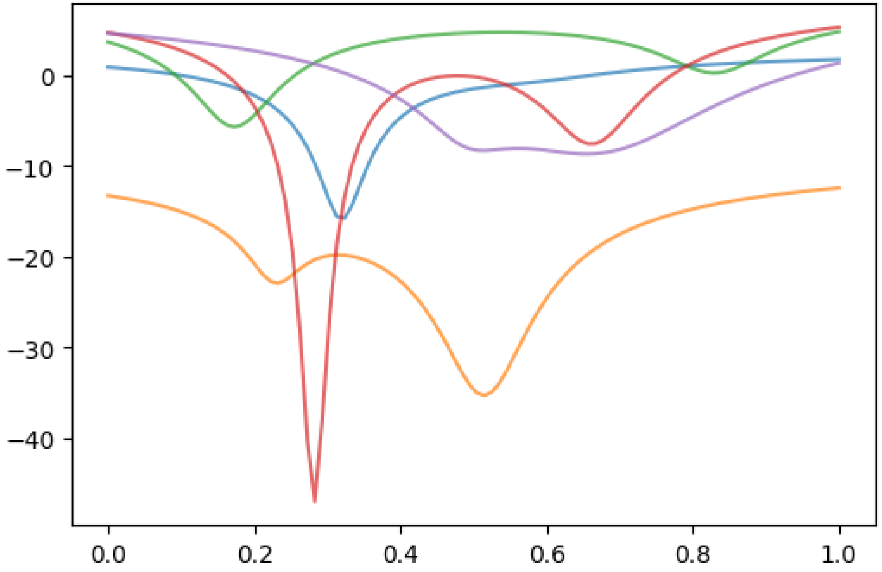
$$L_n \sim \frac{A_n}{\sqrt{(x - x_n^0)^2 + w_n^2}}$$

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

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

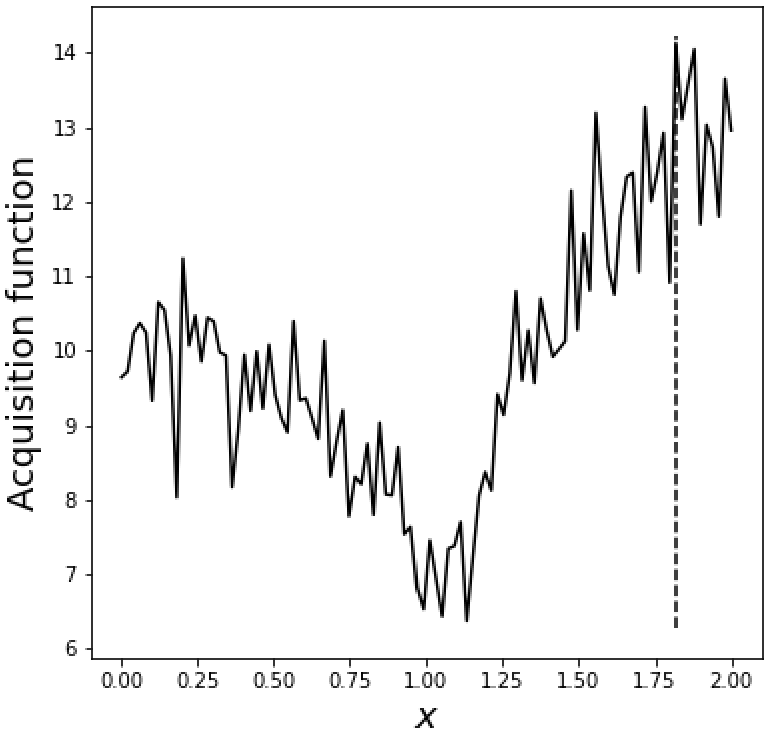
$$x_n^0 \sim Uniform(0, 1)$$

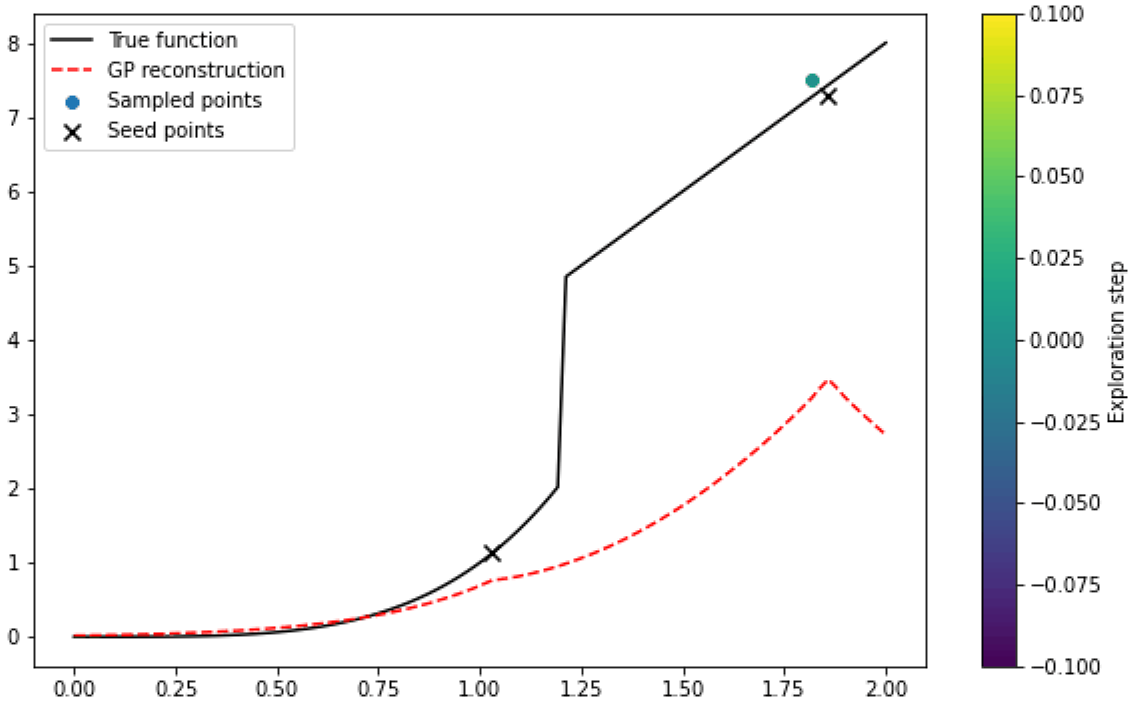
Prior predictive distribution



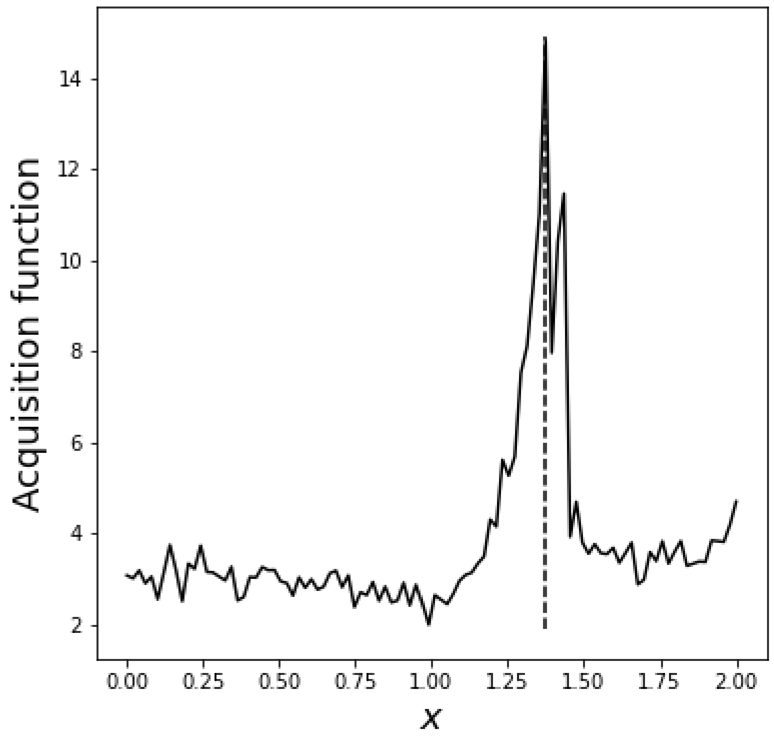
This model simply tells us that there are two minima in our data but does not assume to have any prior knowledge about their relative depth, width, or distance

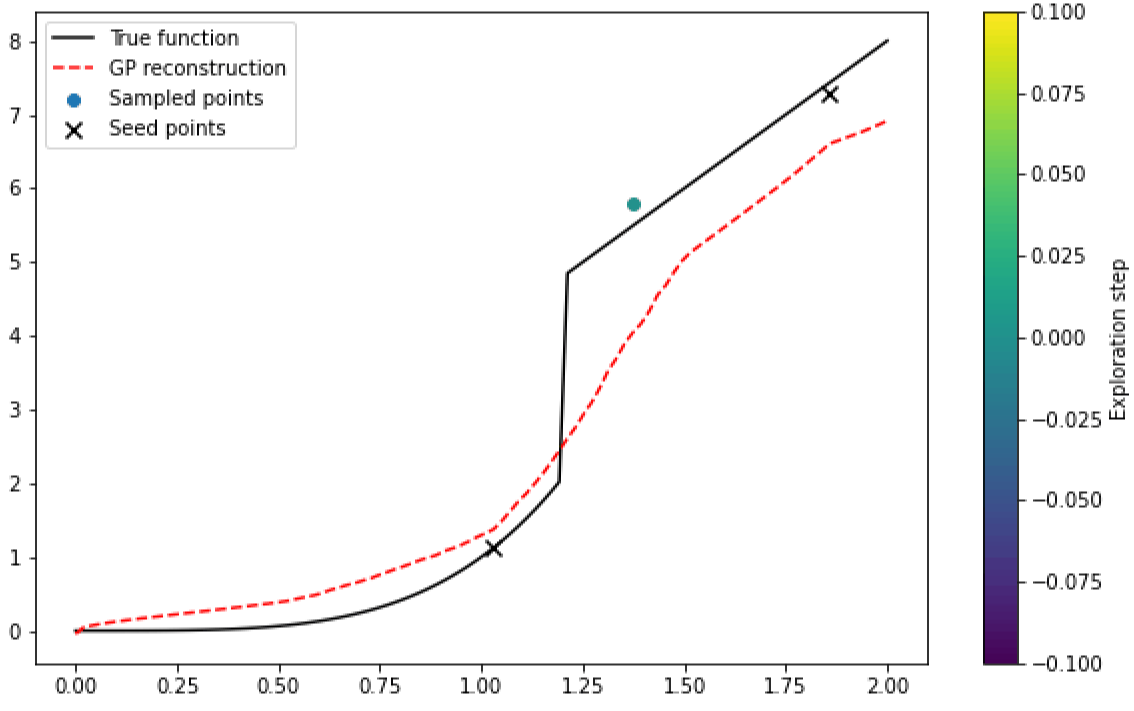
Simple GP search





Structured GP search



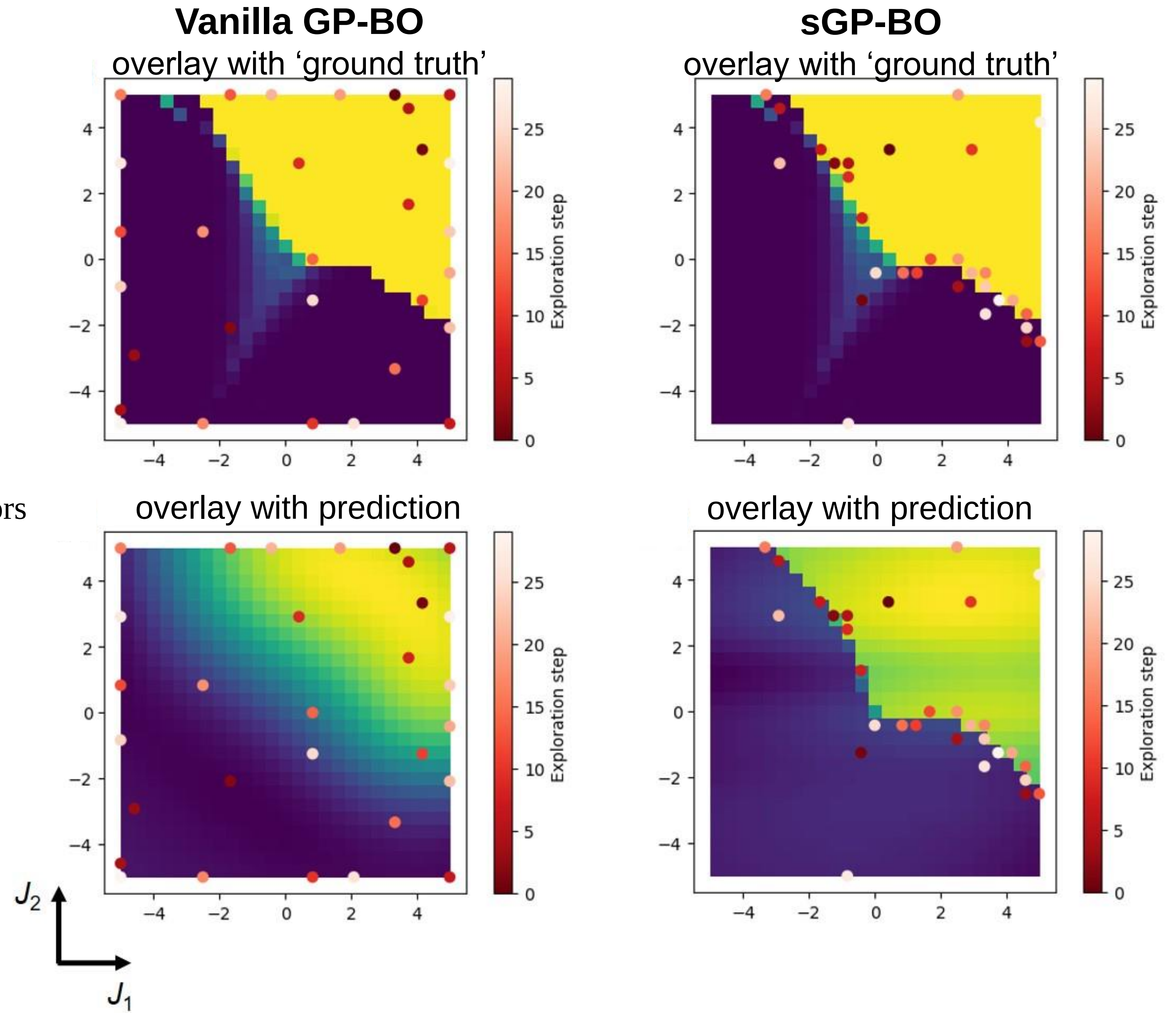


Application to Ising model

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

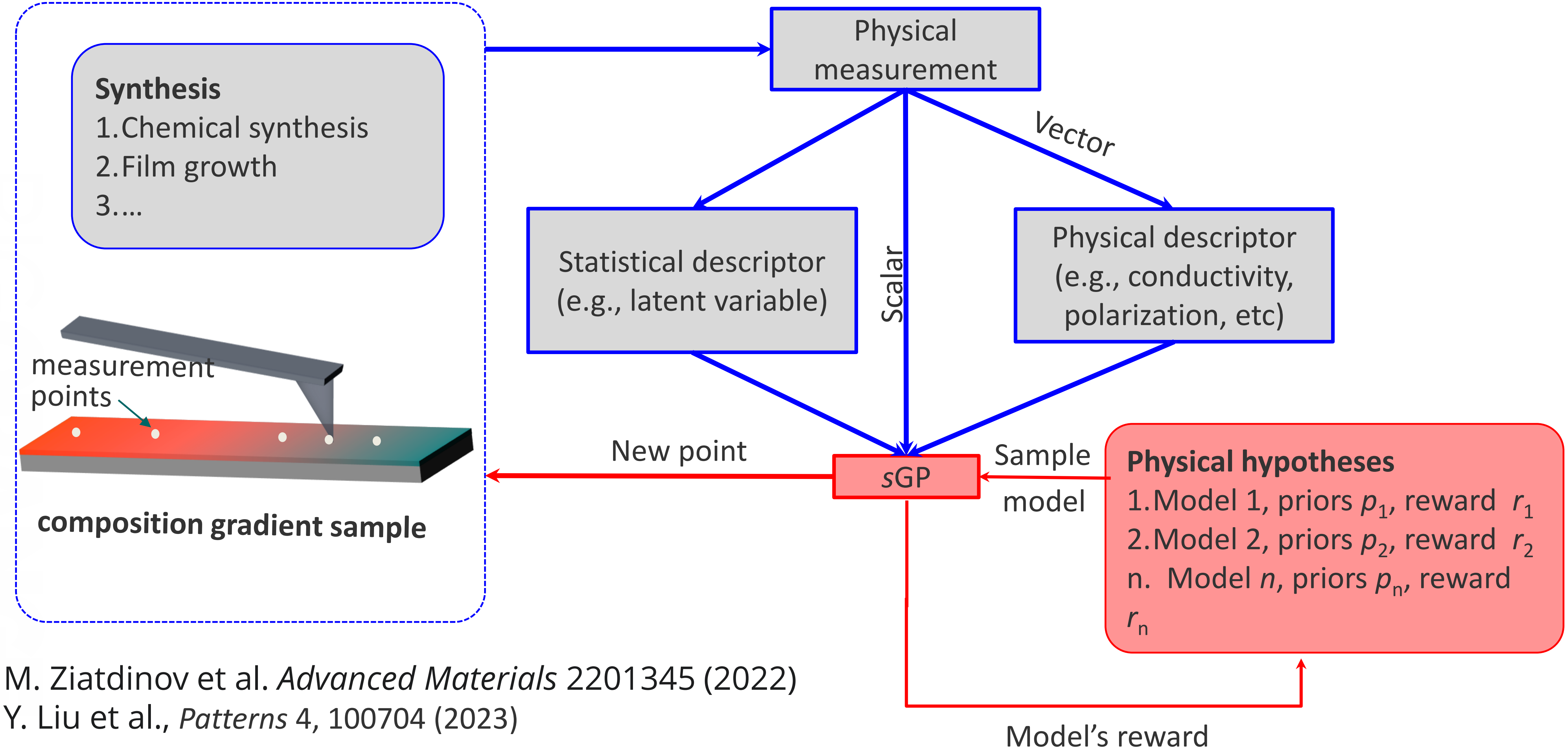


Colab

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

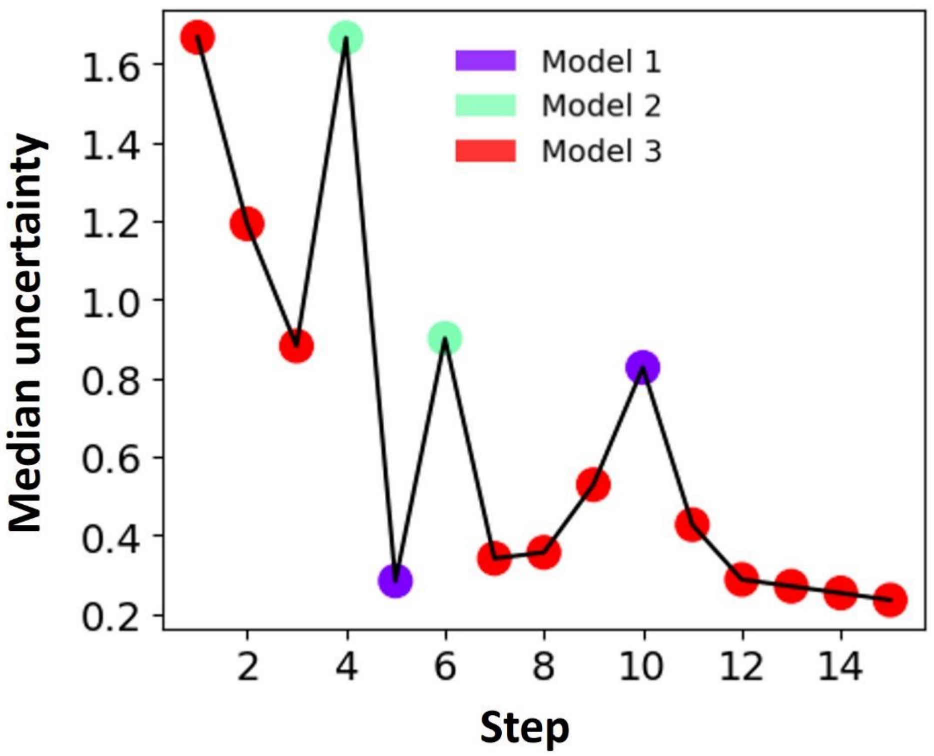


M. Ziatdinov et al. *Advanced Materials* 2201345 (2022)
Y. Liu et al., *Patterns* 4, 100704 (2023)

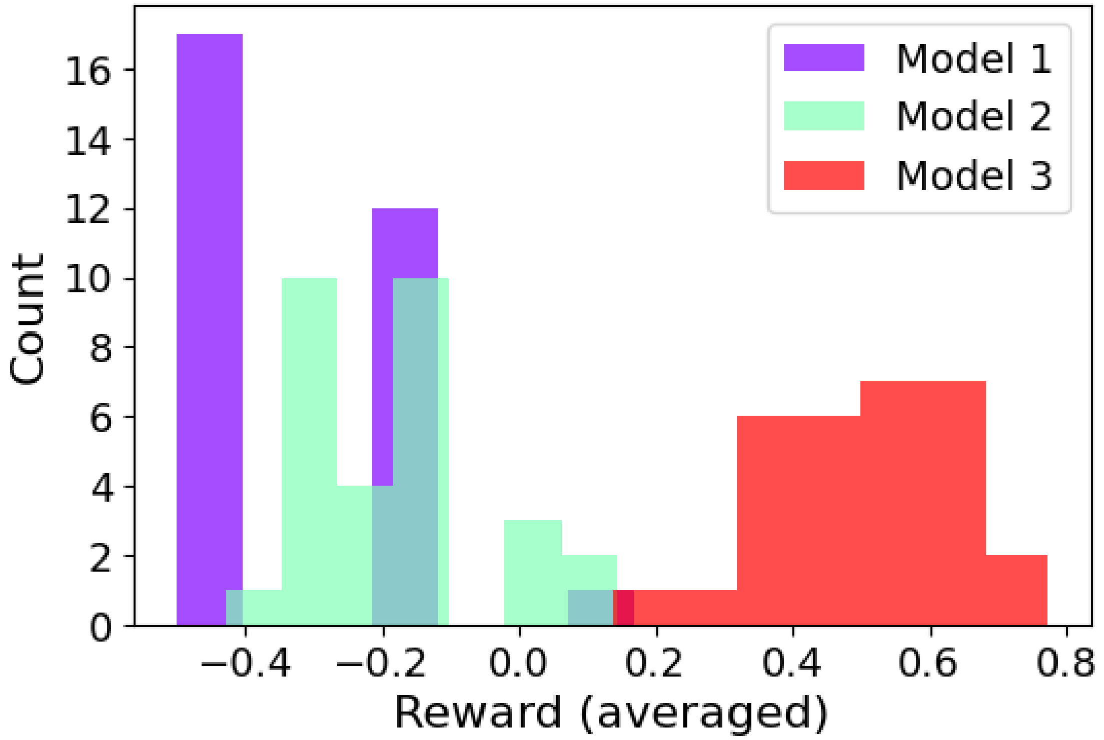
Hypothesis Learning: Synthetic data

Synthetic data represents a 1D discontinuous phase transition

Evolution of uncertainty for a single seed



Results for 30 random seeds

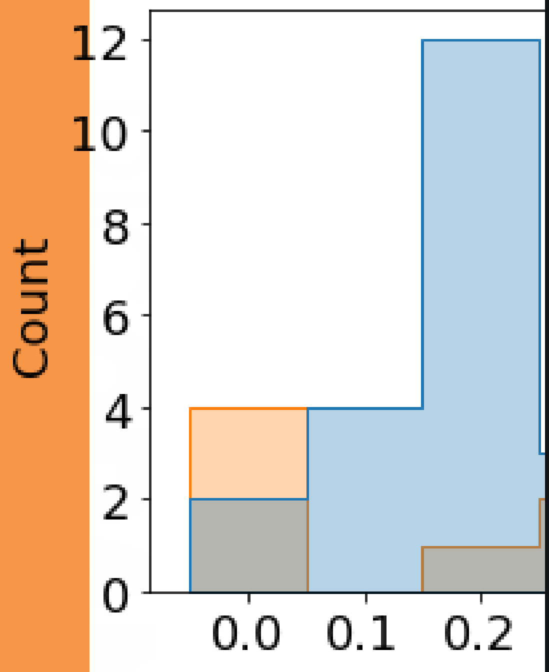
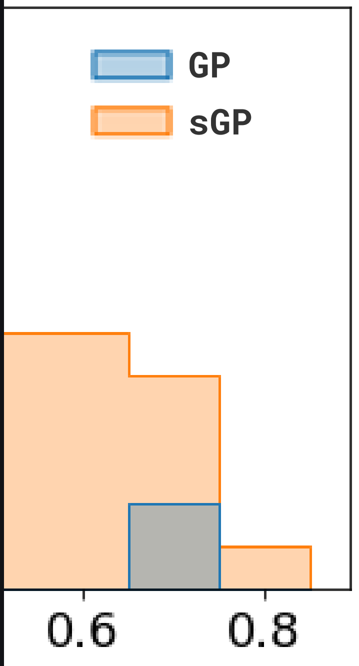


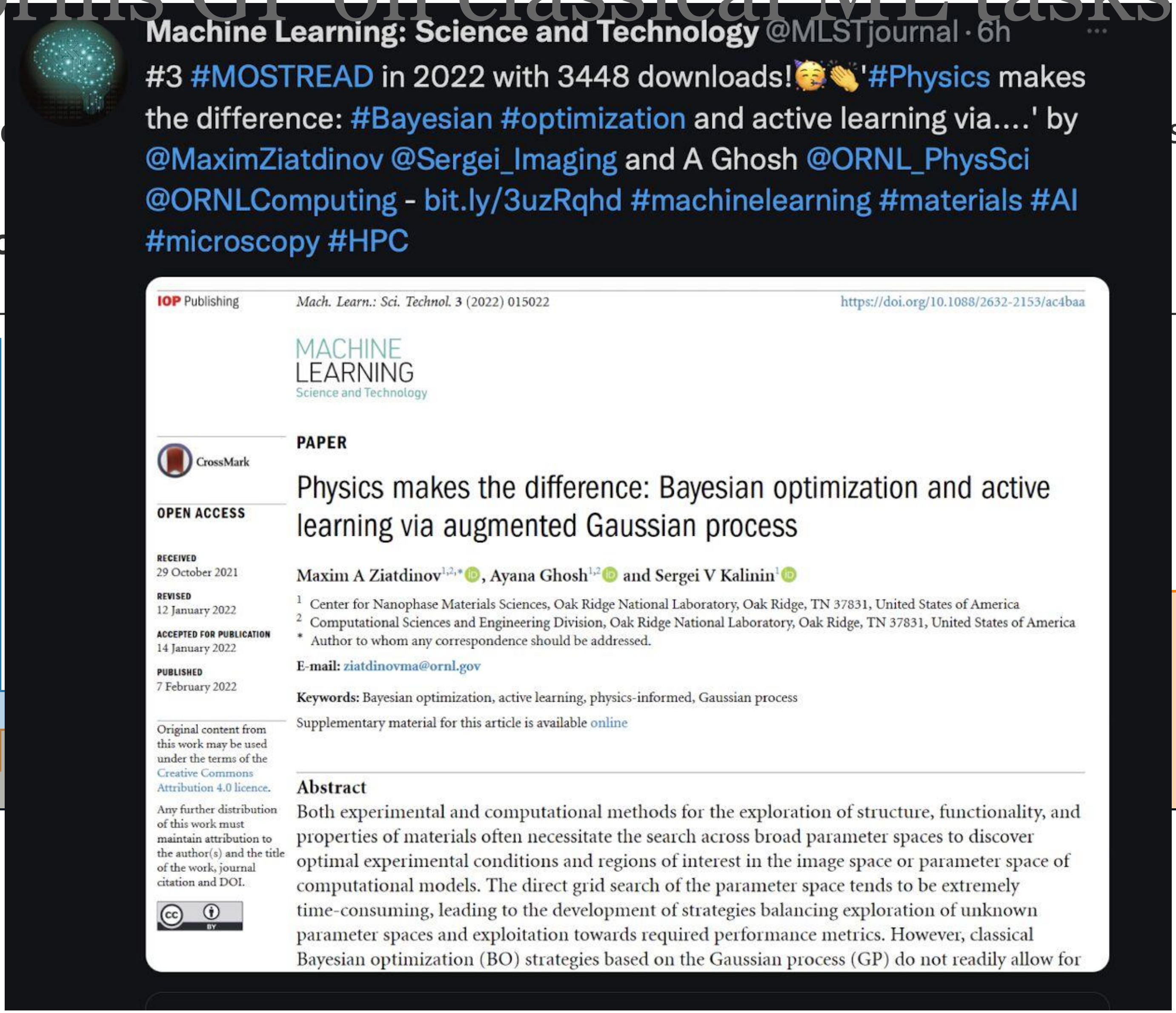
The hypothesis learning learns a correct data distribution with a small number of sparse measurements while also identifying a correct model that describes the system's behavior

Outperforms GP on classical ML tasks

Outp… …s

Tho…



Machine Learning: Science and Technology @MLSTjournal · 6h

#3 #MOSTREAD in 2022 with 3448 downloads! '#Physics makes the difference: #Bayesian #optimization and active learning via....' by @MaximZiatdinov @Sergei_Imaging and A Ghosh @ORNL_PhysSci @ORNLComputing - bit.ly/3uzRqhd #machinelearning #materials #AI #microscopy #HPC

IOP Publishing Mach. Learn.: Sci. Technol. 3 (2022) 015022 https://doi.org/10.1088/2632-2153/ac4baa

MACHINE LEARNING Science and Technology

CrossMark

OPEN ACCESS

RECEIVED 29 October 2021

REVISED 12 January 2022

ACCEPTED FOR PUBLICATION 14 January 2022

PUBLISHED 7 February 2022

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

Physics makes the difference: Bayesian optimization and active learning via augmented Gaussian process

Maxim A Ziatdinov[1,2,*], Ayana Ghosh[1,2] and Sergei V Kalinin[1]

1 Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

2 Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

* Author to whom any correspondence should be addressed.

E-mail: ziatdinovma@ornl.gov

Keywords: Bayesian optimization, active learning, physics-informed, Gaussian process

Supplementary material for this article is available online

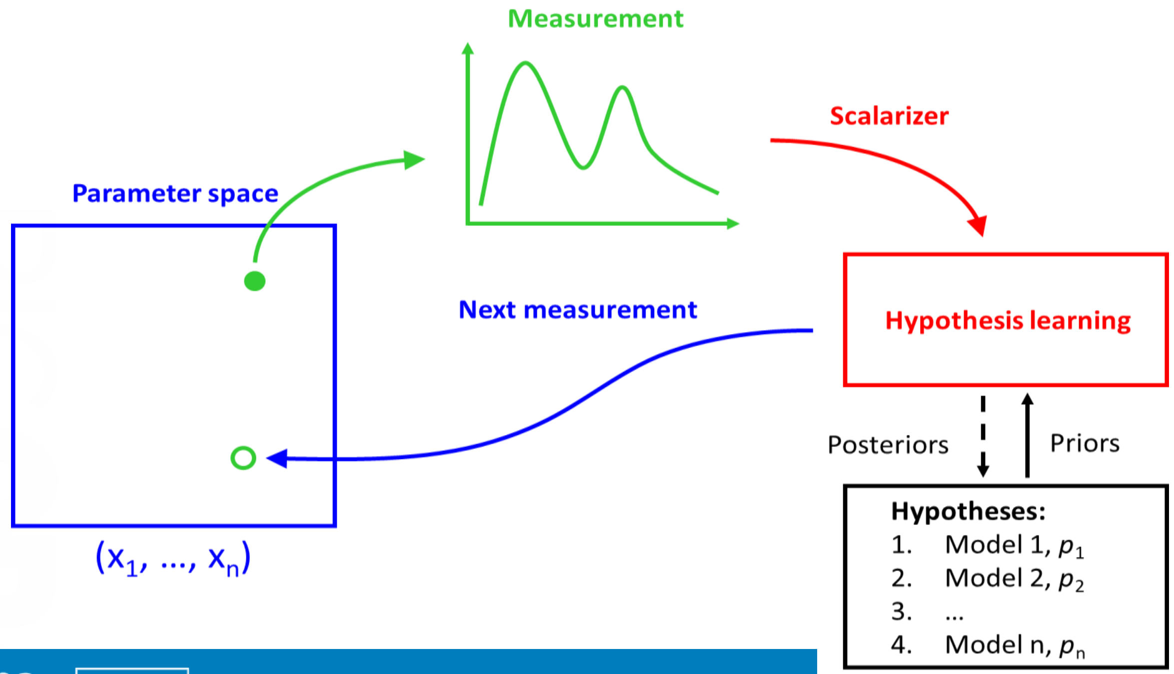
Abstract

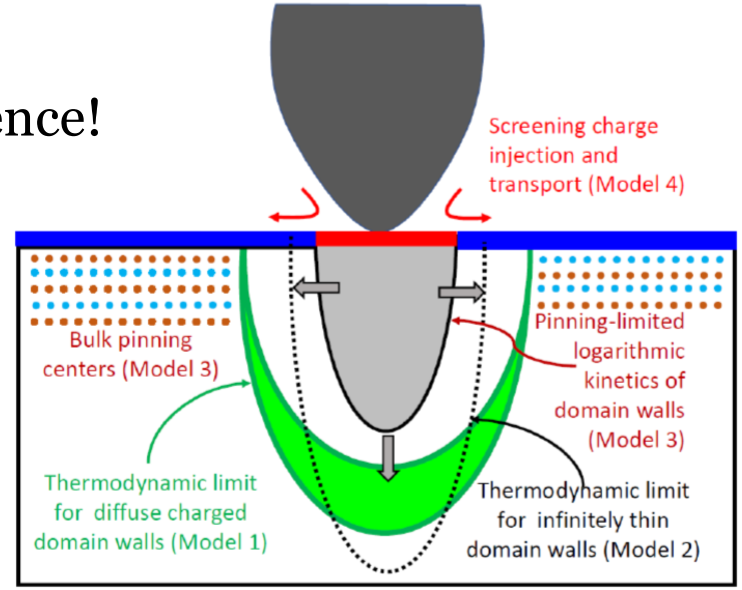
Both experimental and computational methods for the exploration of structure, functionality, and properties of materials often necessitate the search across broad parameter spaces to discover optimal experimental conditions and regions of interest in the image space or parameter space of computational models. The direct grid search of the parameter space tends to be extremely time-consuming, leading to the development of strategies balancing exploration of unknown parameter spaces and exploitation towards required performance metrics. However, classical Bayesian optimization (BO) strategies based on the Gaussian process (GP) do not readily allow for

M. Ziatdinov et al., *Machine Learning: Science and Technology* 3, 015022 (2022)

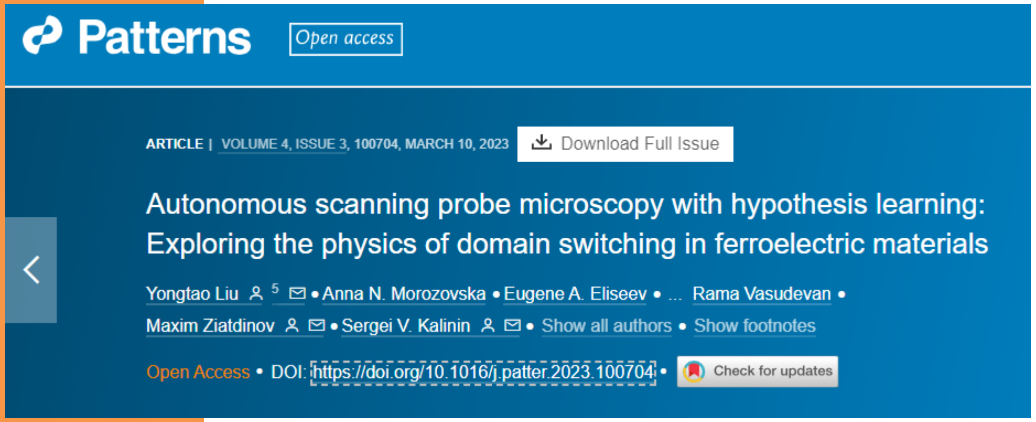
Hypothesis Learning

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

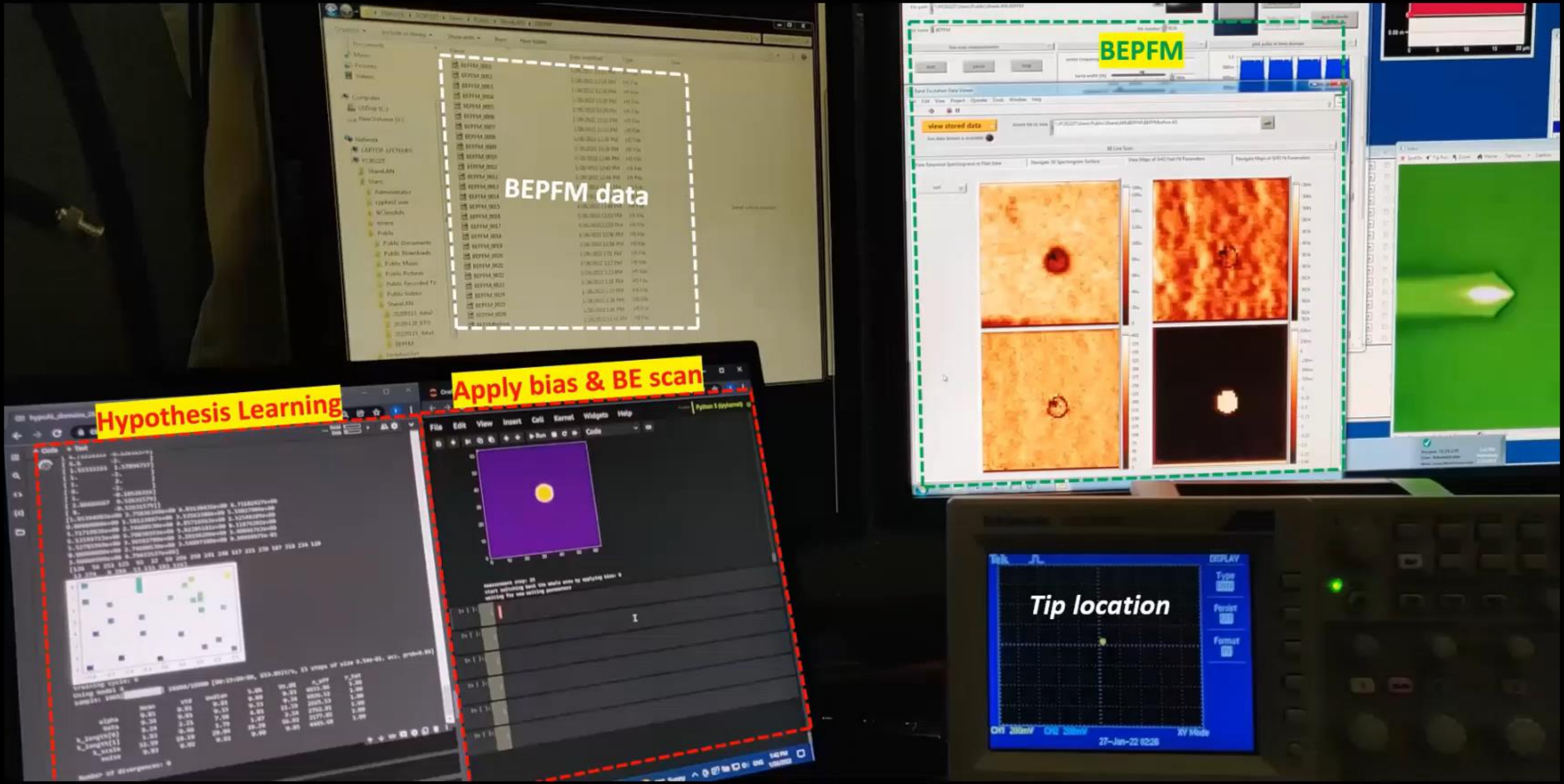

Patterns Open access

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

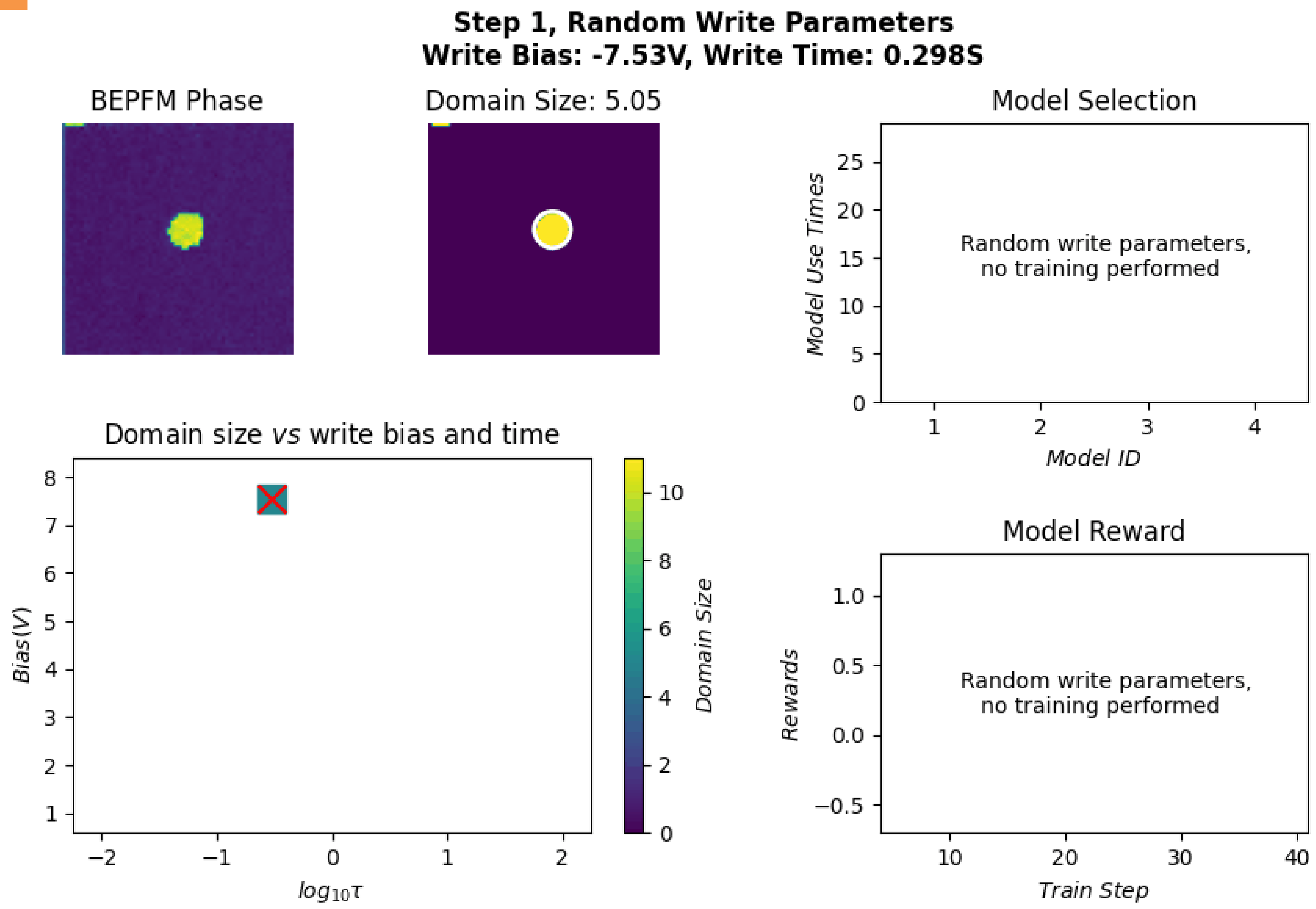
Autonomous scanning probe microscopy with hypothesis learning: Exploring the physics of domain switching in ferroelectric materials

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

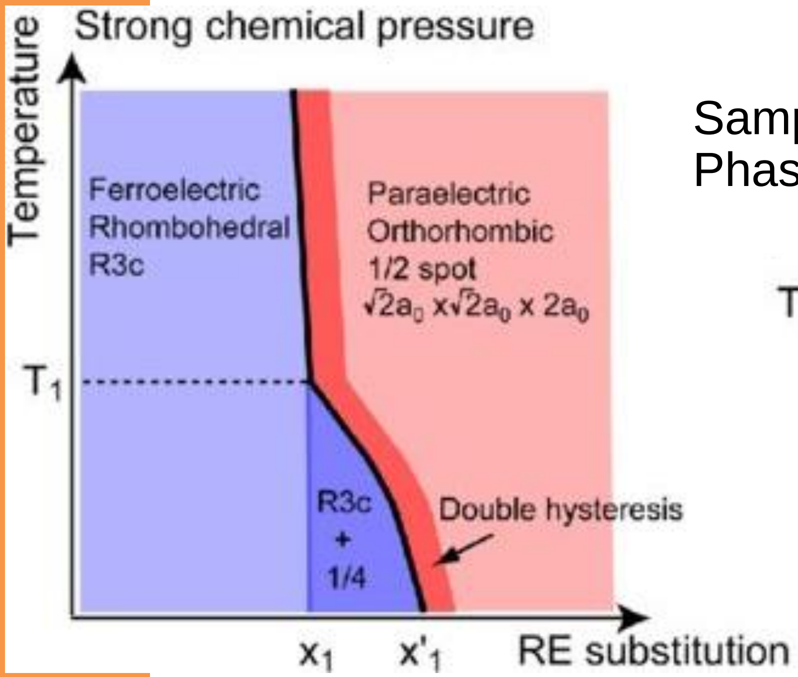
Research Article

Hypothesis Learning in Automated Experiment: Application to Combinatorial Materials Libraries

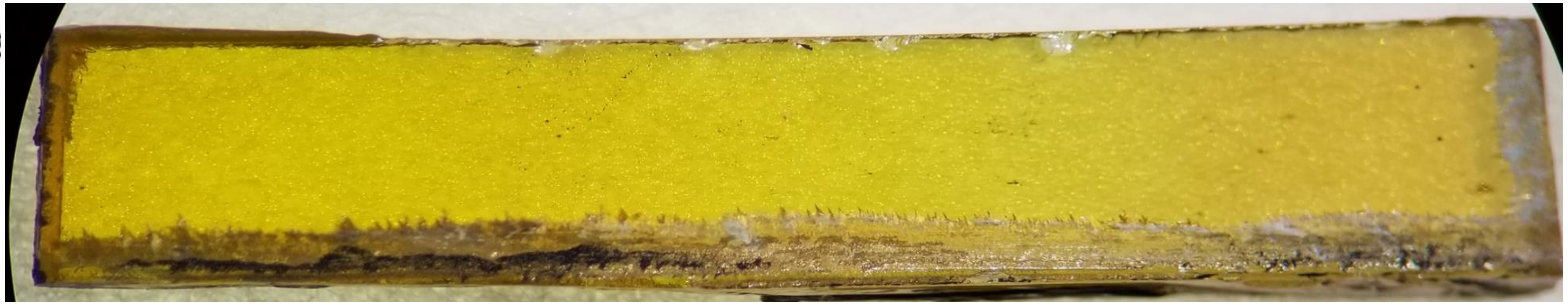
Maxim A. Ziatdinov, Yongtao Liu, Anna N. Morozovska, Eugene A. Eliseev, Xiaohang Zhang, Ichiro Takeuchi, Sergei V. Kalinin

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.



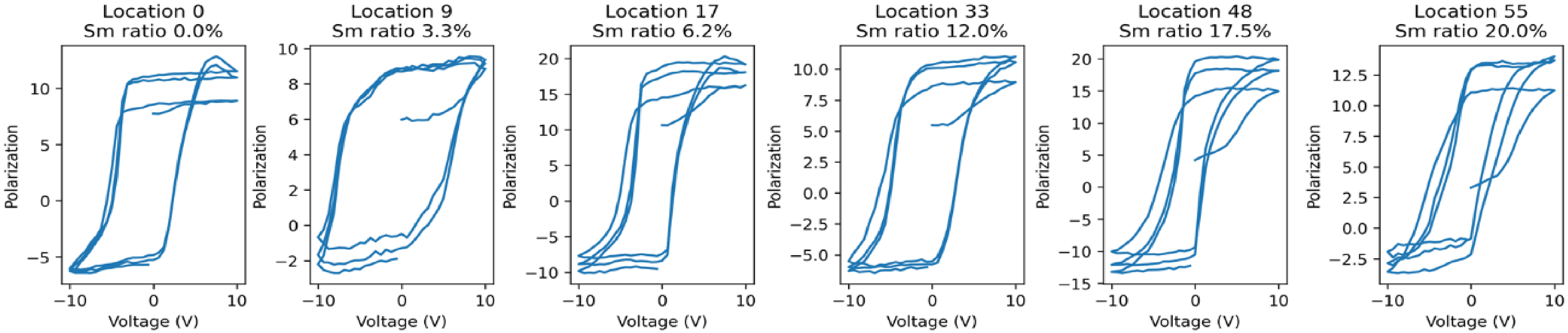
T2



BiFeC

Linear est. 7%Sm $BiFeO_3$

20%Sm $BiFeO_3$



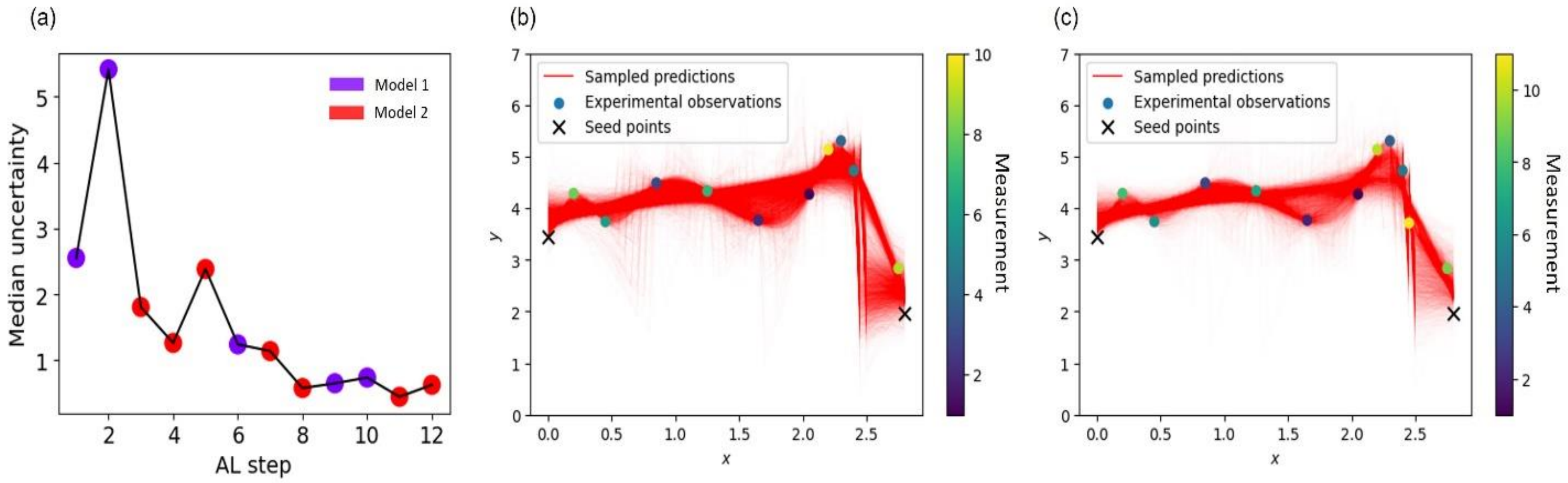
Hypothesis Selection for Ferroelectric

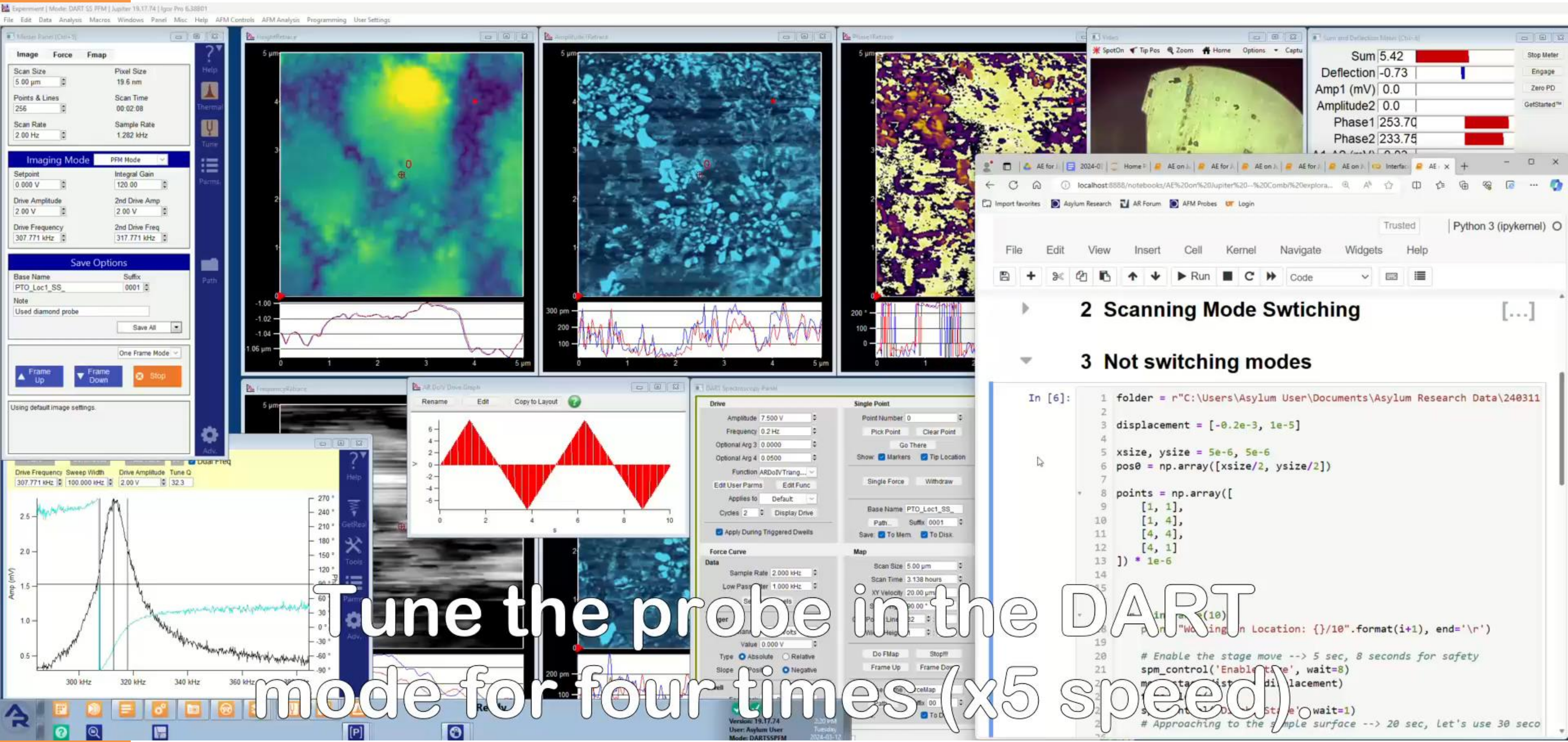
Model 1 (second order phase transition):

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

Model 2 (first order phase transition):

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





Tune the probe in the DART mode for four times (x5 speed)

Colab 2