THE UNIVERSITY of TENNESSEE KNOXVILLE

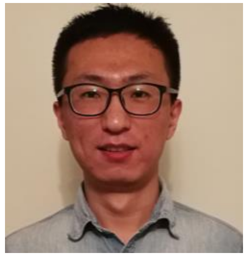
Gaussian Processes and Bayesian Optimization

Sergei V. Kalinin

University of Tennessee, Knoxville and
Pacific Northwest National Laboratory

Maxim Ziatdinov

Mahshid Ahmadi



Yongtao Liu

Rama Vasudevan

Kevin Roccapriore

The dance of policies and rewards

Rewards and objectives:

- What is our (hierarchical) objective?
- Can we define reward(s)?

Inferential biases:

- What do we know before the experiment?
- What do we (hope to) learn after the experiment?

Experiment planning – policies and values

- How do we plan experiment in advance (policies or values based on rewards)?
- Can we ascribe value to certain steps?
- Do we change our policies during experiment?

Bayesian paradigm in science

- Bayes' theorem can be usefully re-written for science as:

Posterior: probability of the model given the data

Likelihood: probability of the data given the model

Prior: probability of the model

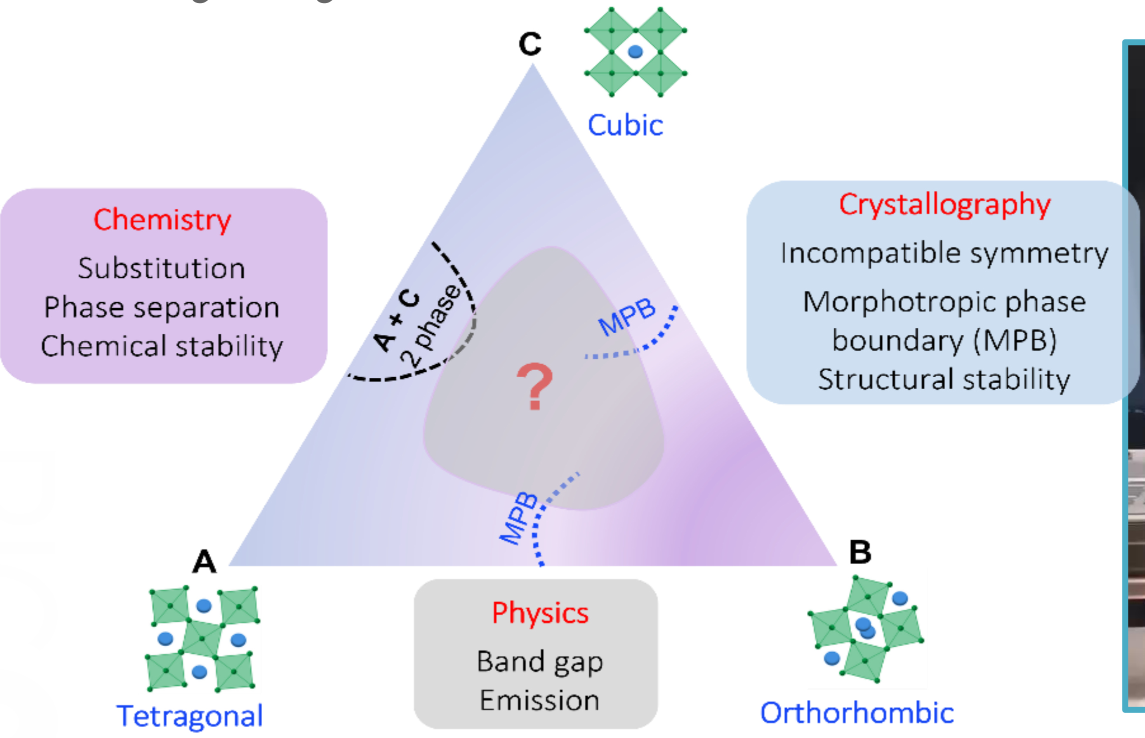
$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})\, P(\text{model})}{P(\text{data})}$$

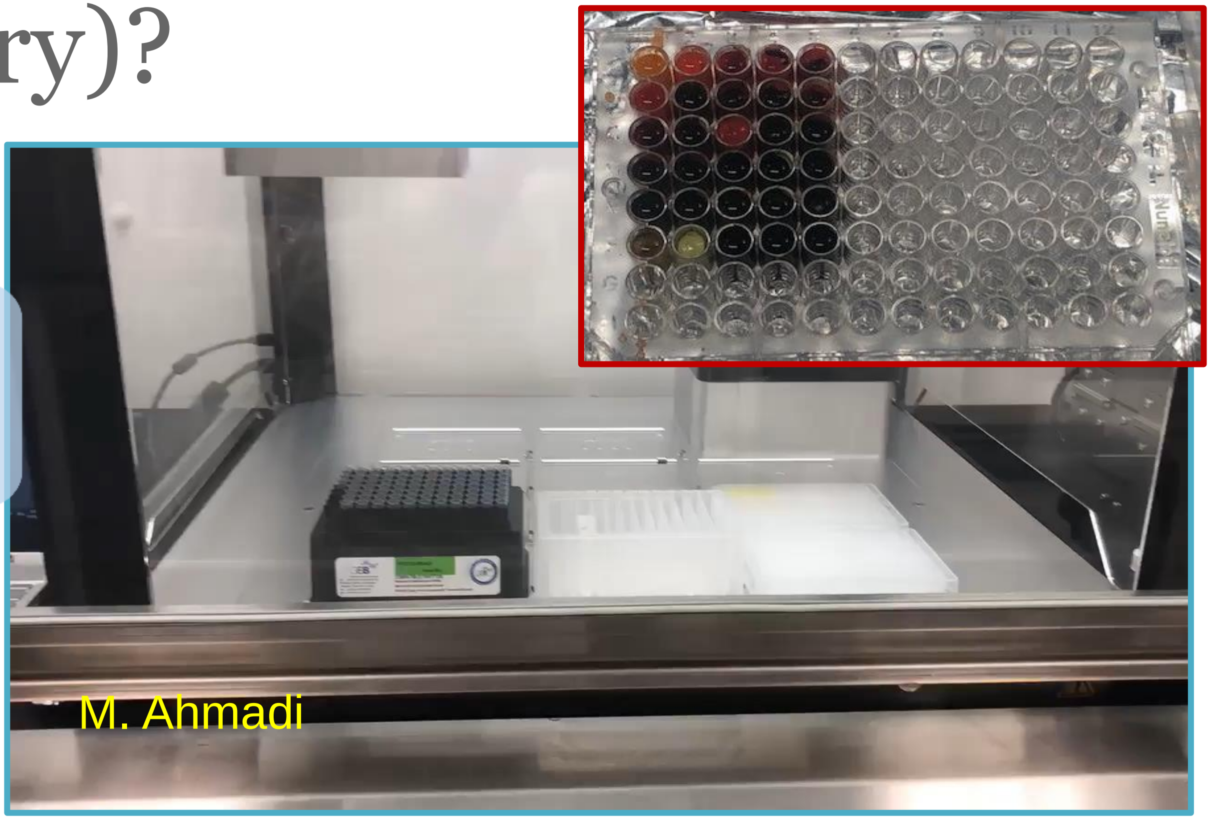
Evidence [can typically be absorbed into the normalization of the posterior]

- Prior
- Posterior
- Belief

Gaussian Processes

Why synthesis (or theory)?

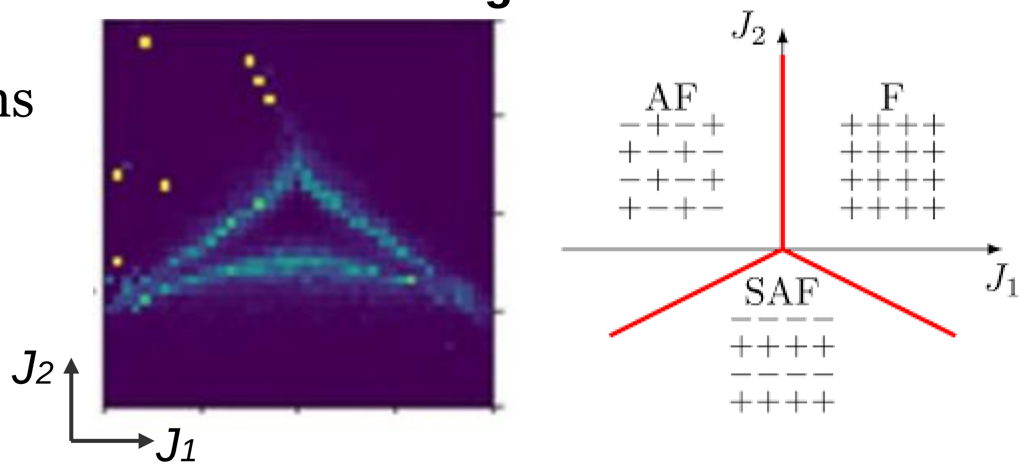




- Automated synthesis in its simplest form requires some way to navigate phase diagrams
- In more complex form, processing space.
- Ideally, incorporate physical knowledge

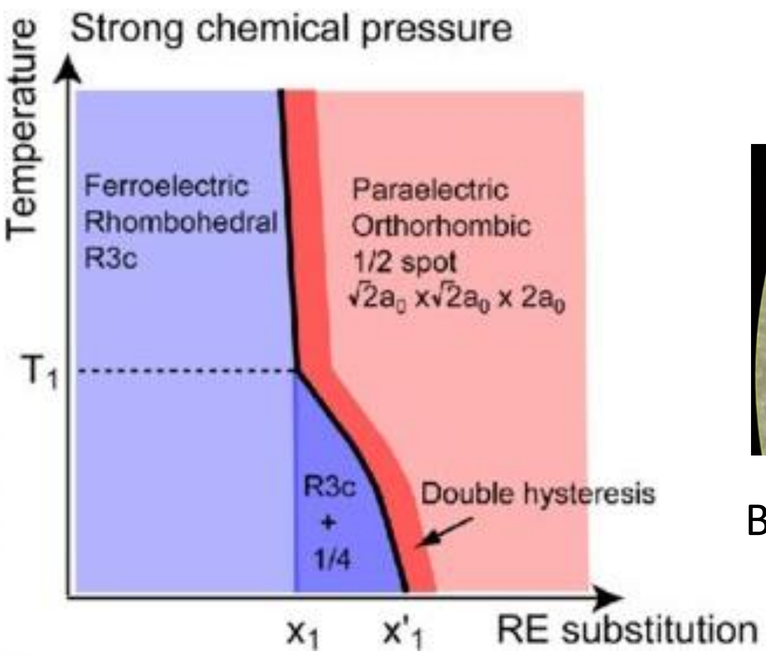
- Similar problem - theory

Ising model

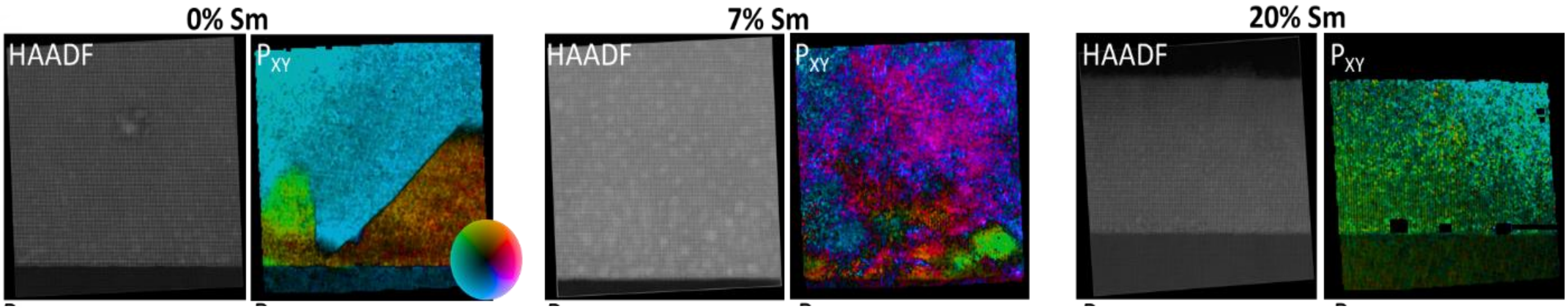


Combinatorial library

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

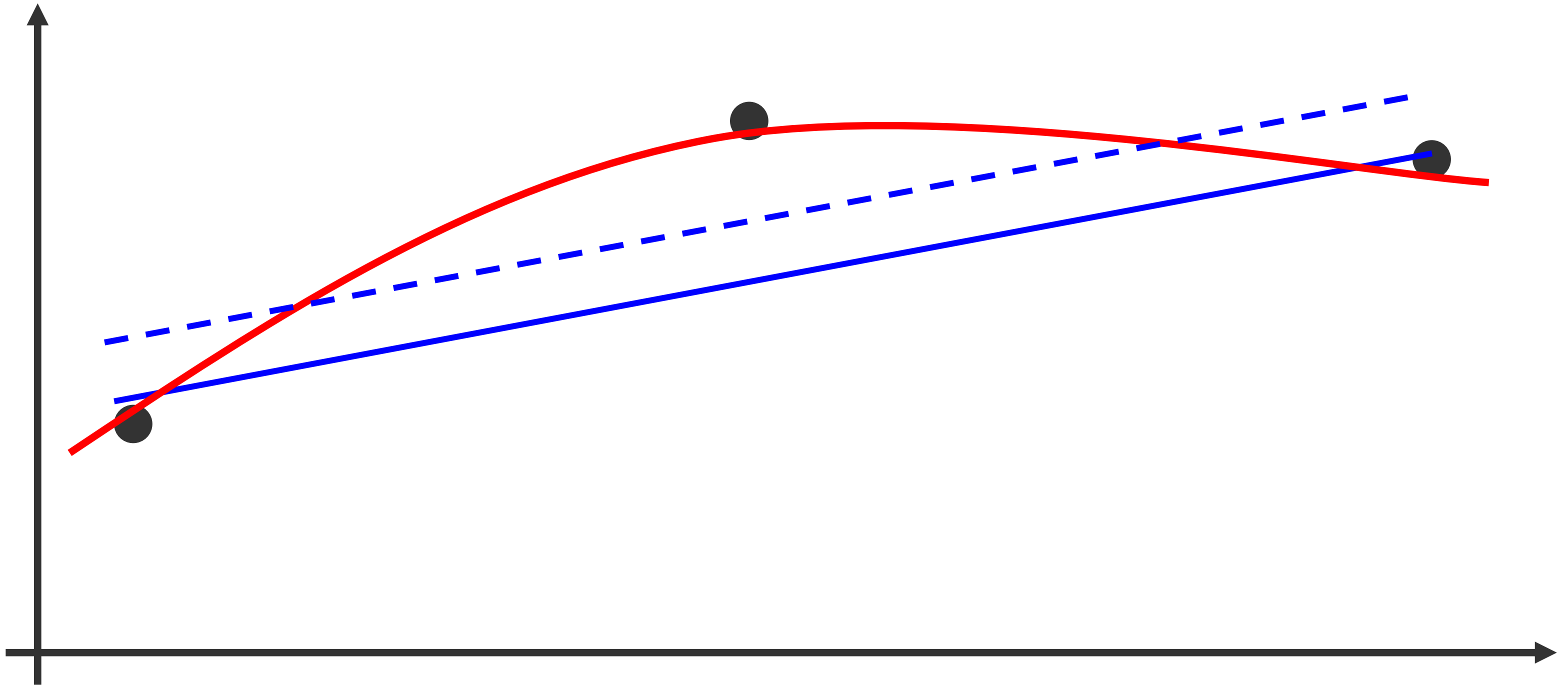


$BiFeO_3$ Linear est. 7%Sm $BiFeO_3$ 20%Sm $BiFeO_3$

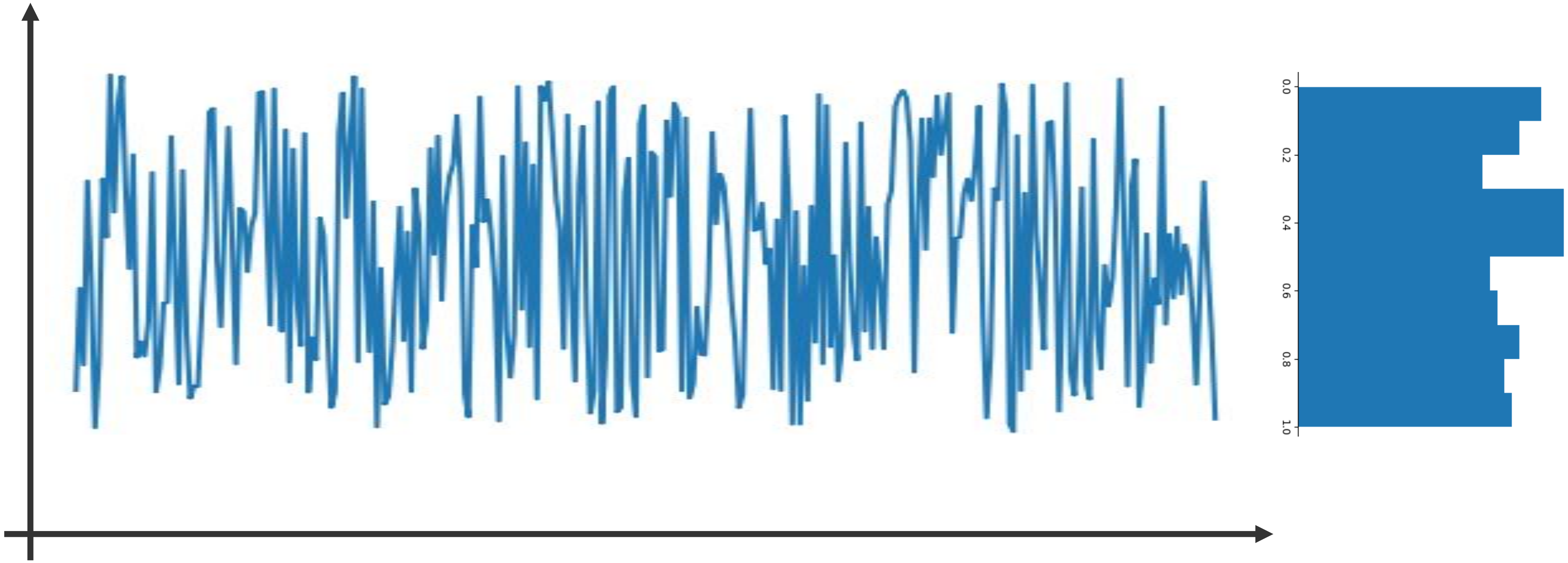


arXiv:2004.11817

What do we know if we do not know anything?



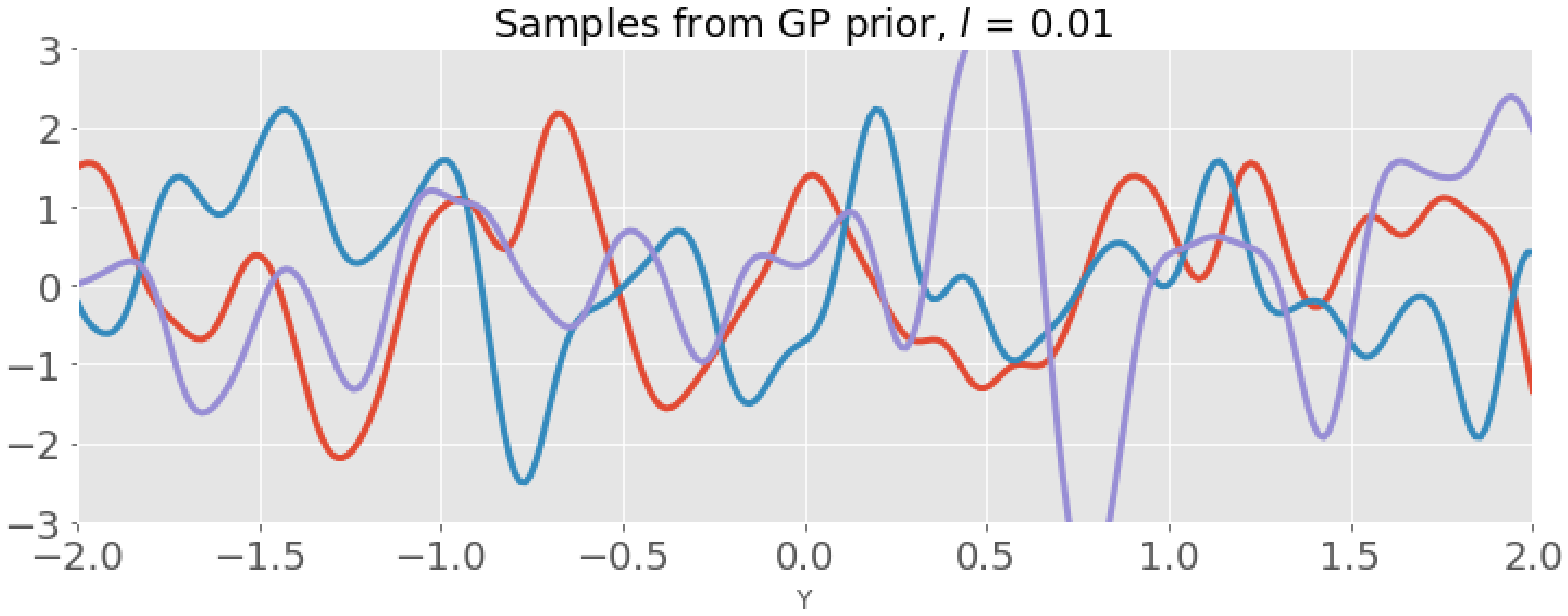
What do we know if we do not know anything?



Gaussian Process Regression

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

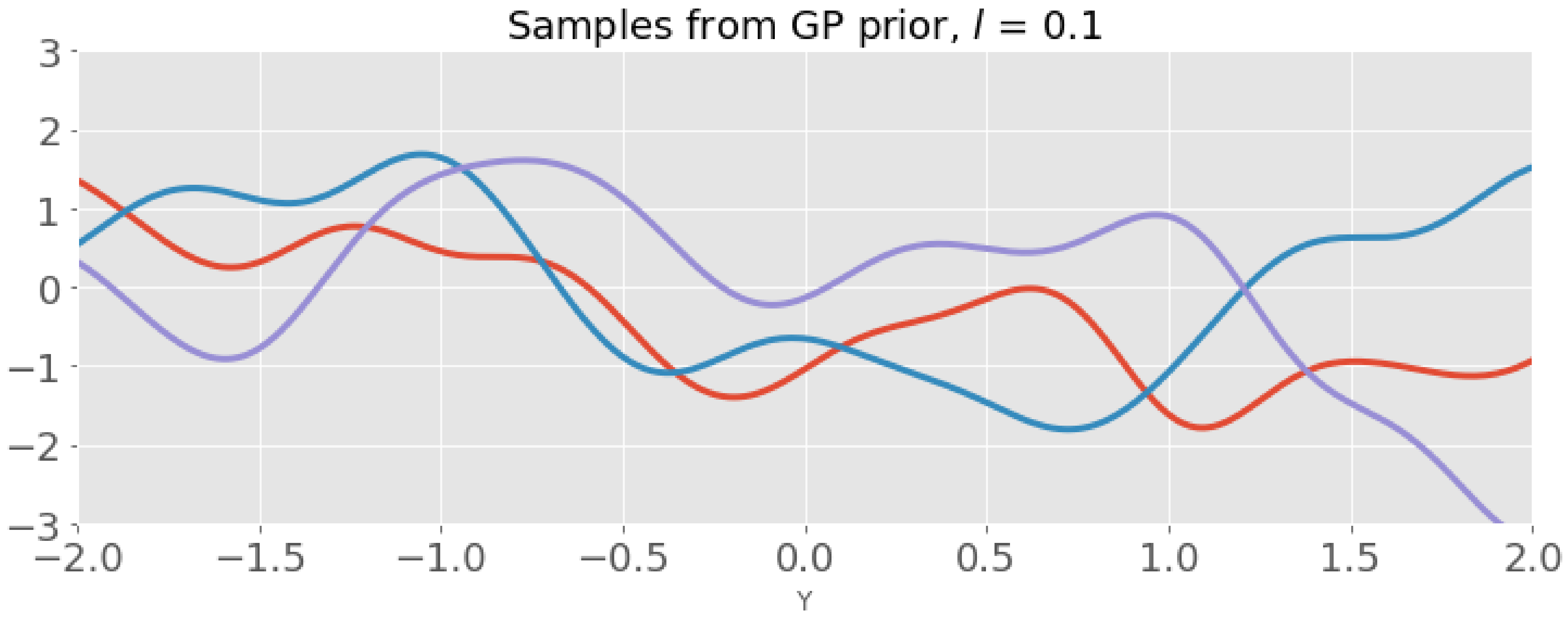
$$k(x, x') = \exp\left(-\frac{1}{2l}(x - x')^2\right)$$



Gaussian Process Regression

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

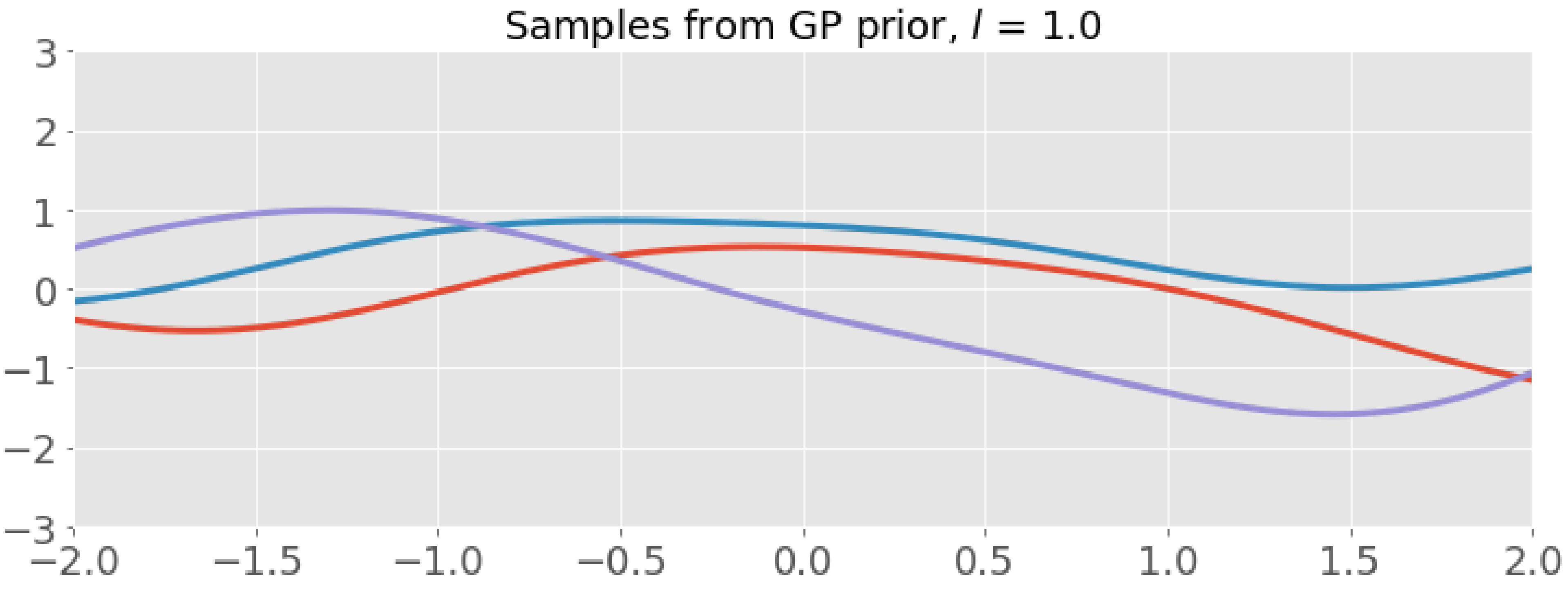
$$k(x, x') = \exp\left(-\frac{1}{2l}(x - x')^2\right)$$



Gaussian Process Regression

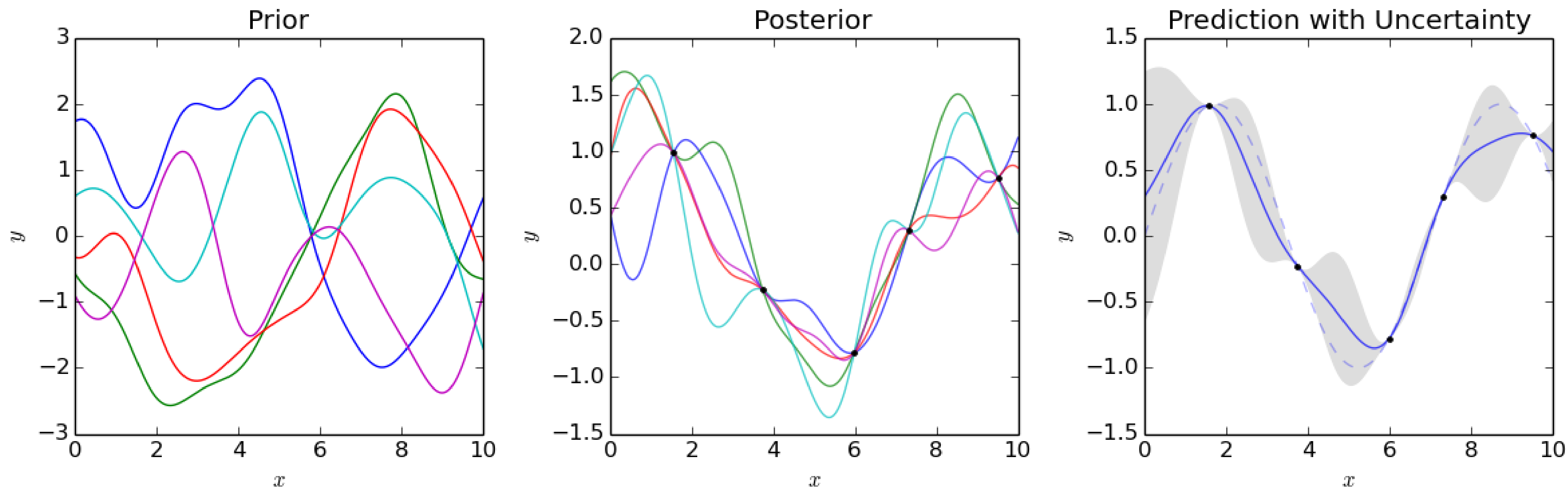
- Covariance matrix (kernel) determines what type of functions we will allow.

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



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

Gaussian Process Regression

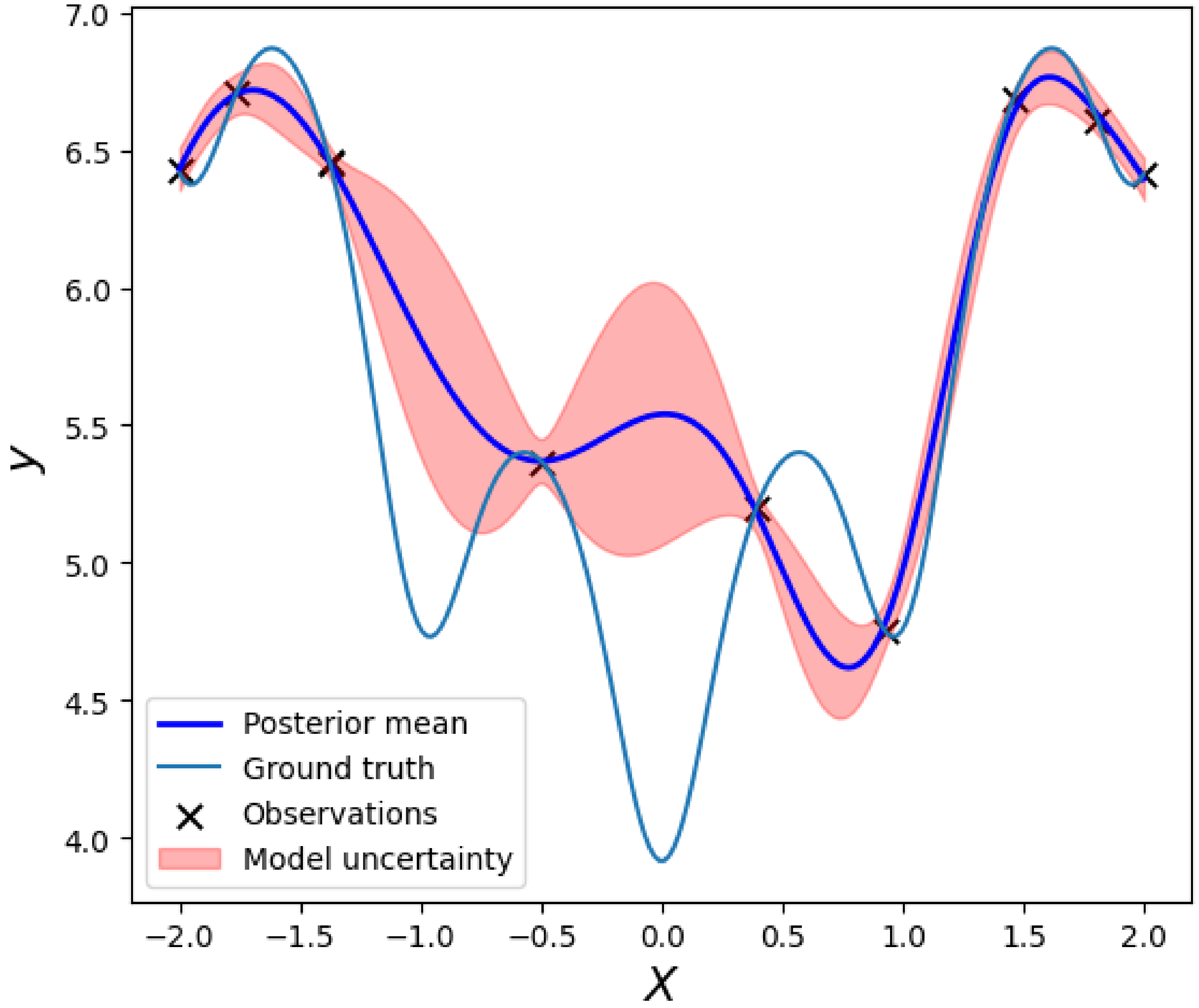


Prior:	What can the function be before the measurement
Data:	Measurement
Posterior:	What can the function be after measurement
Policy:	How do we balance exploration and exploitation (acquisition function)

GP Vocabulary

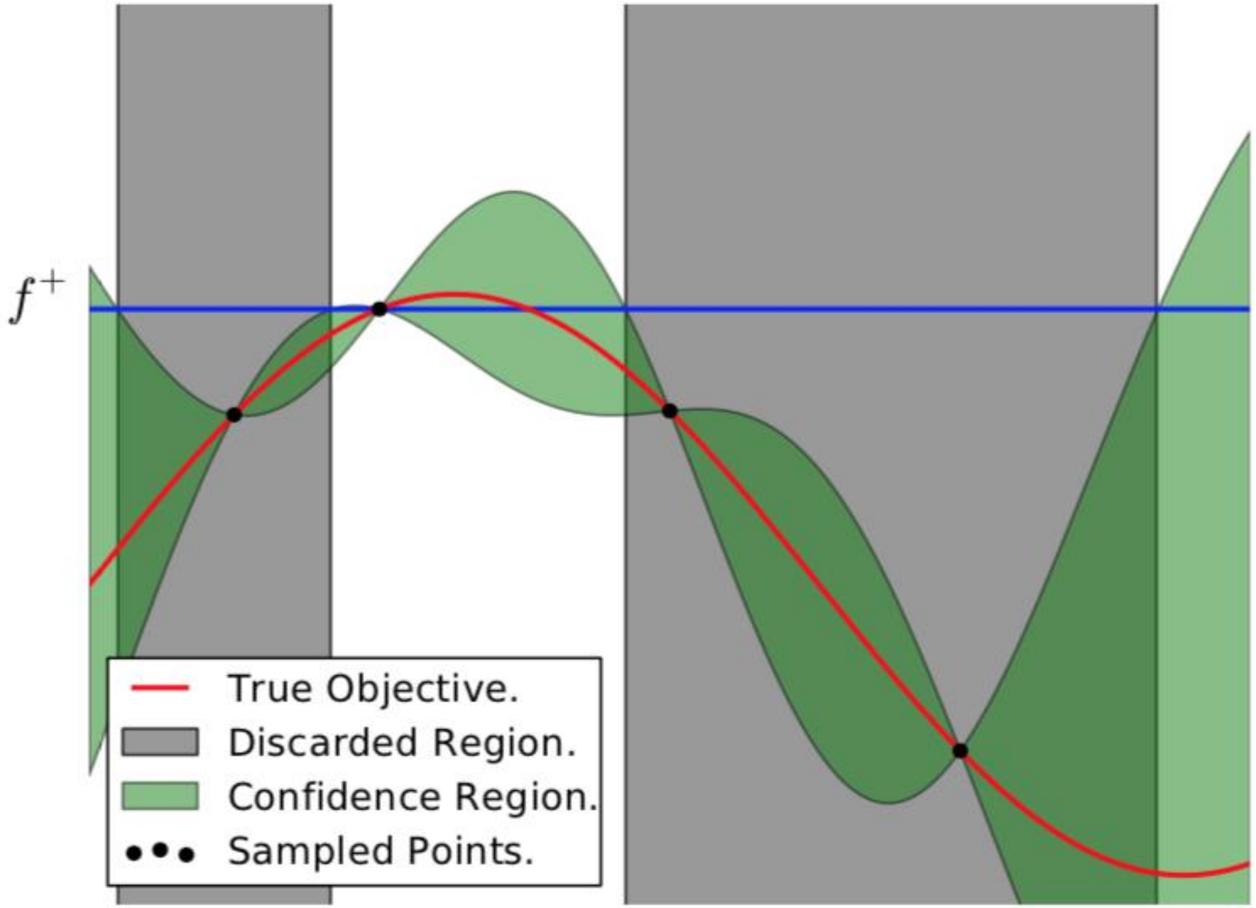
- Gaussian Process
- Kernel and kernel parameters
- Kernel Priors
- Noise Priors
- Posteriors

Colab



7.0
6.5
6.0
5.5
5.0
4.5
4.0
y
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
X
Posterior mean
Ground truth
Observations
Model uncertainty

Bayesian Optimization

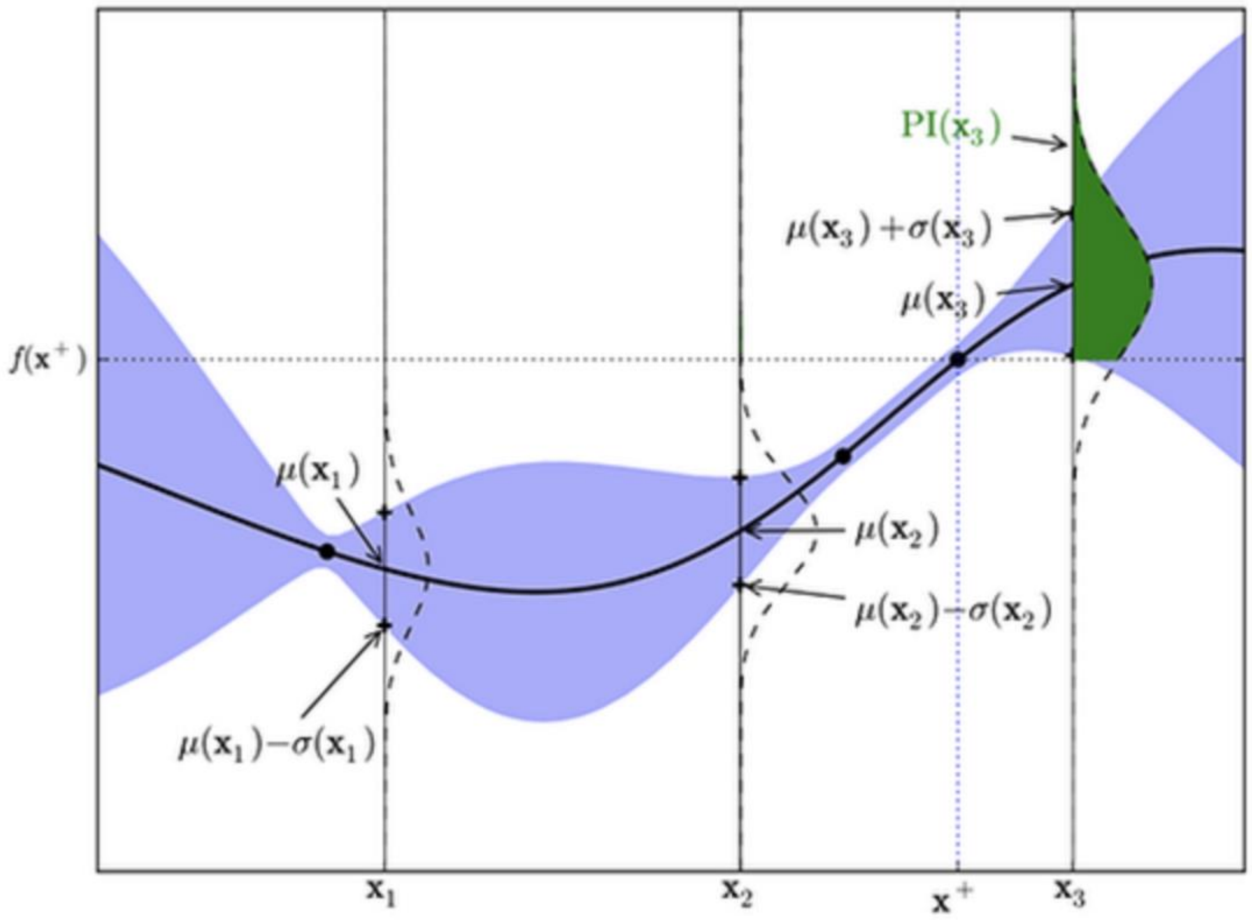


- We have some measurements in space X, and we want to maximize some property f(X).
- How can we decide what point to measure next to best maximize f?
- We need to balance the exploration of the space with exploitation of regions near we have already know

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



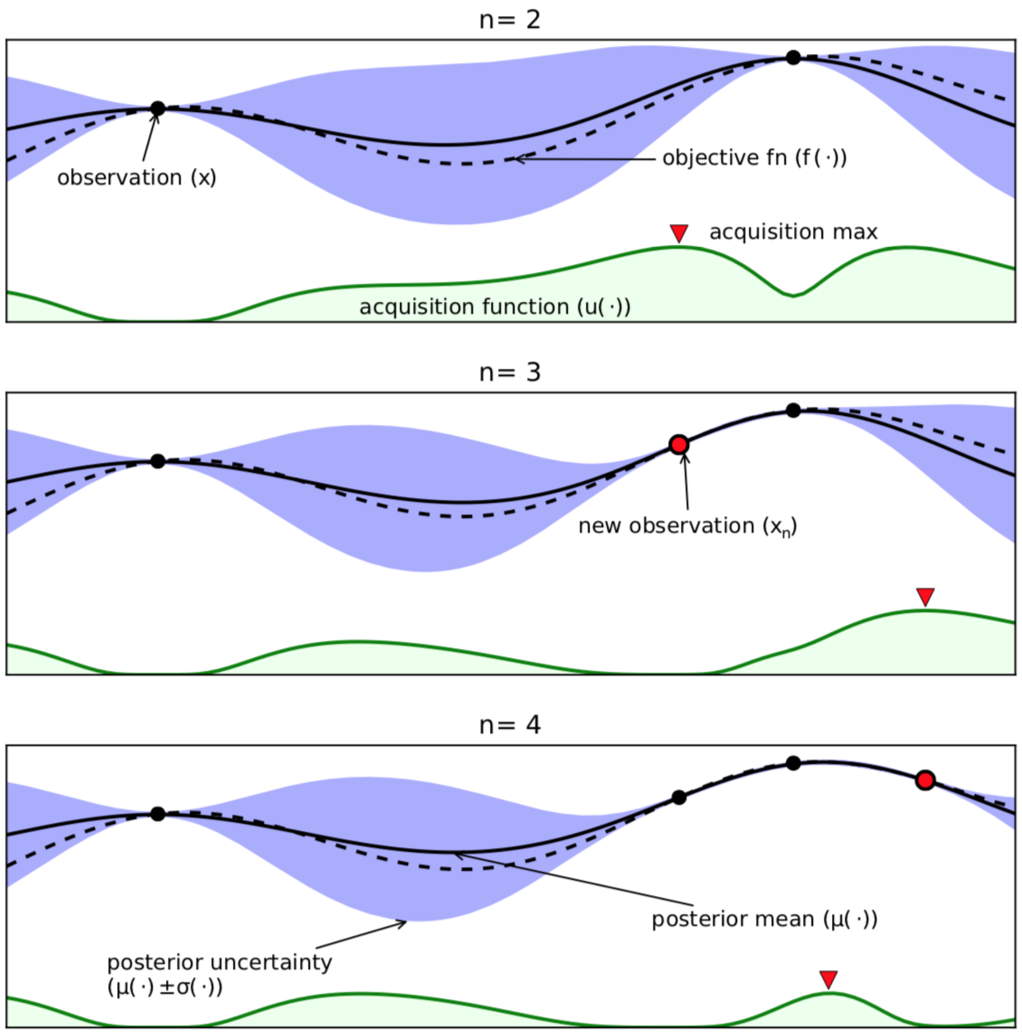
1. **Upper confidence bound:** simplest possible - just take the upper confidence bound from the prediction

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

Obtain/update training data *D*

↓

Train surrogate model $X \rightarrow y$

↓

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

↓

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

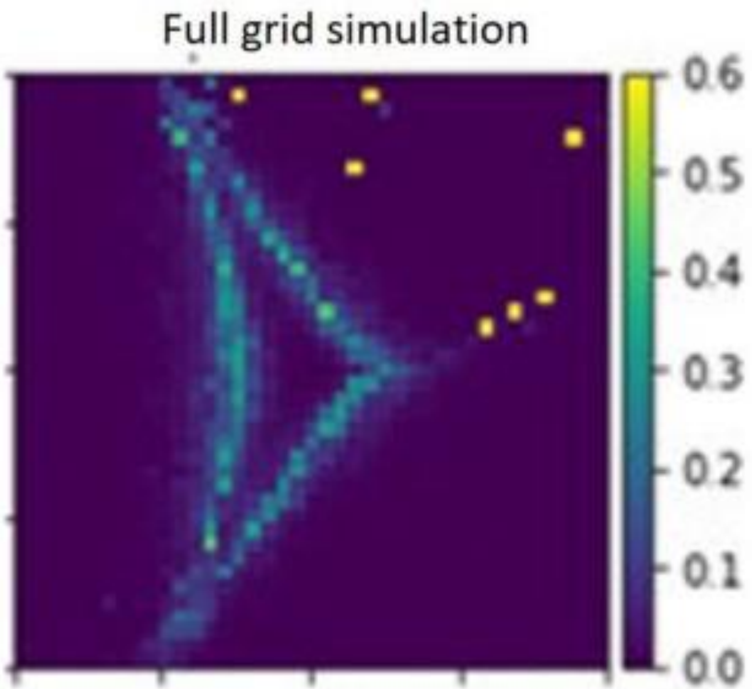
Evaluate new point

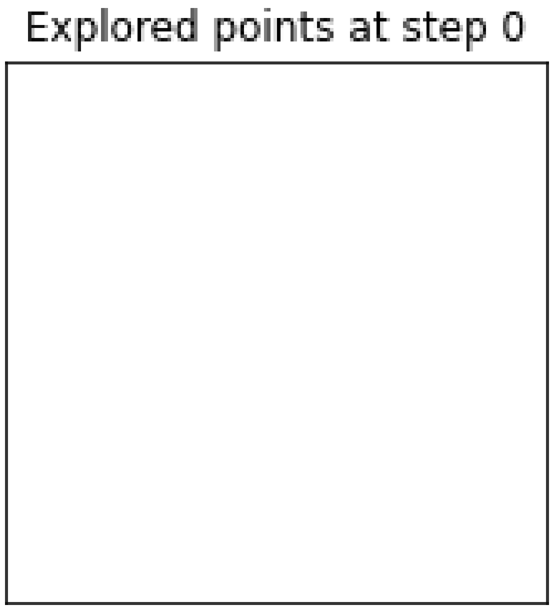
$X_*^i = \mathrm{argmax}(acq)$

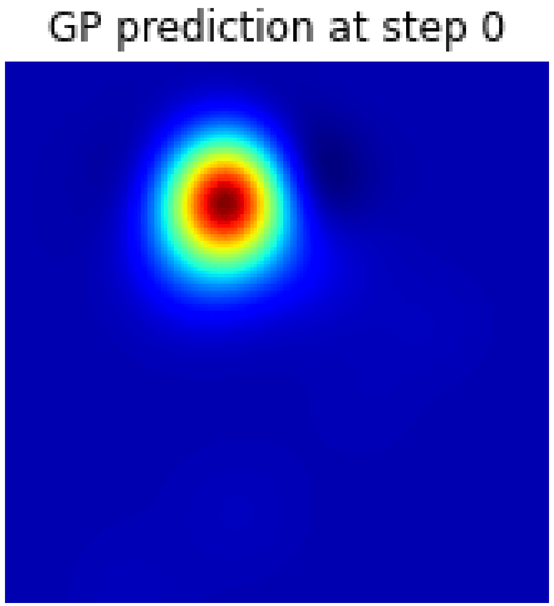
N. de Freitas et al., Taking the Human Out of the Loop: A Review of Bayesian Optimization , *Proceedings of the IEEE* **104**, 148 (2015)

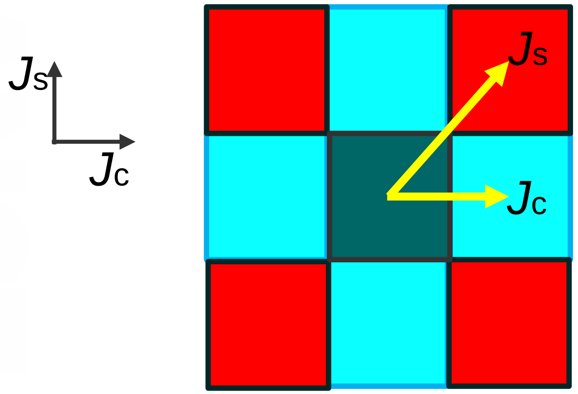
Bayesian Optimization for physical discovery

Discovering regions where heat capacity is maximized in NNN Ising model

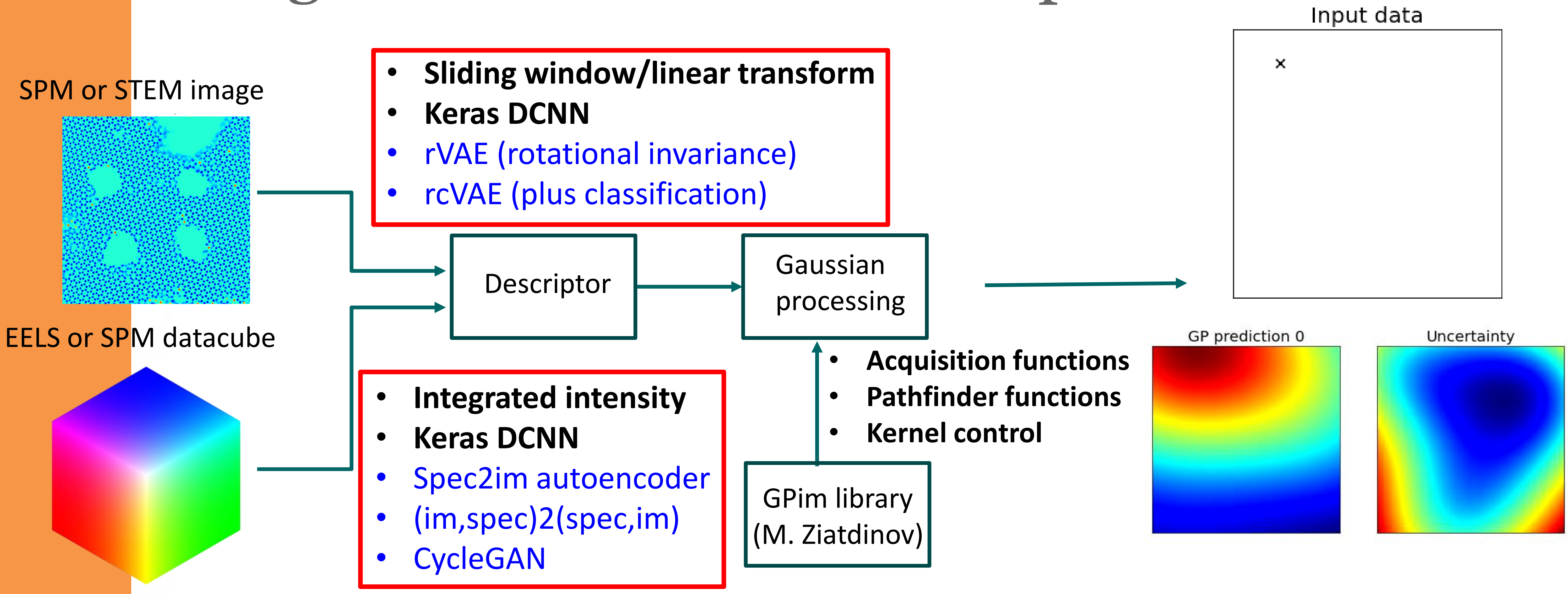






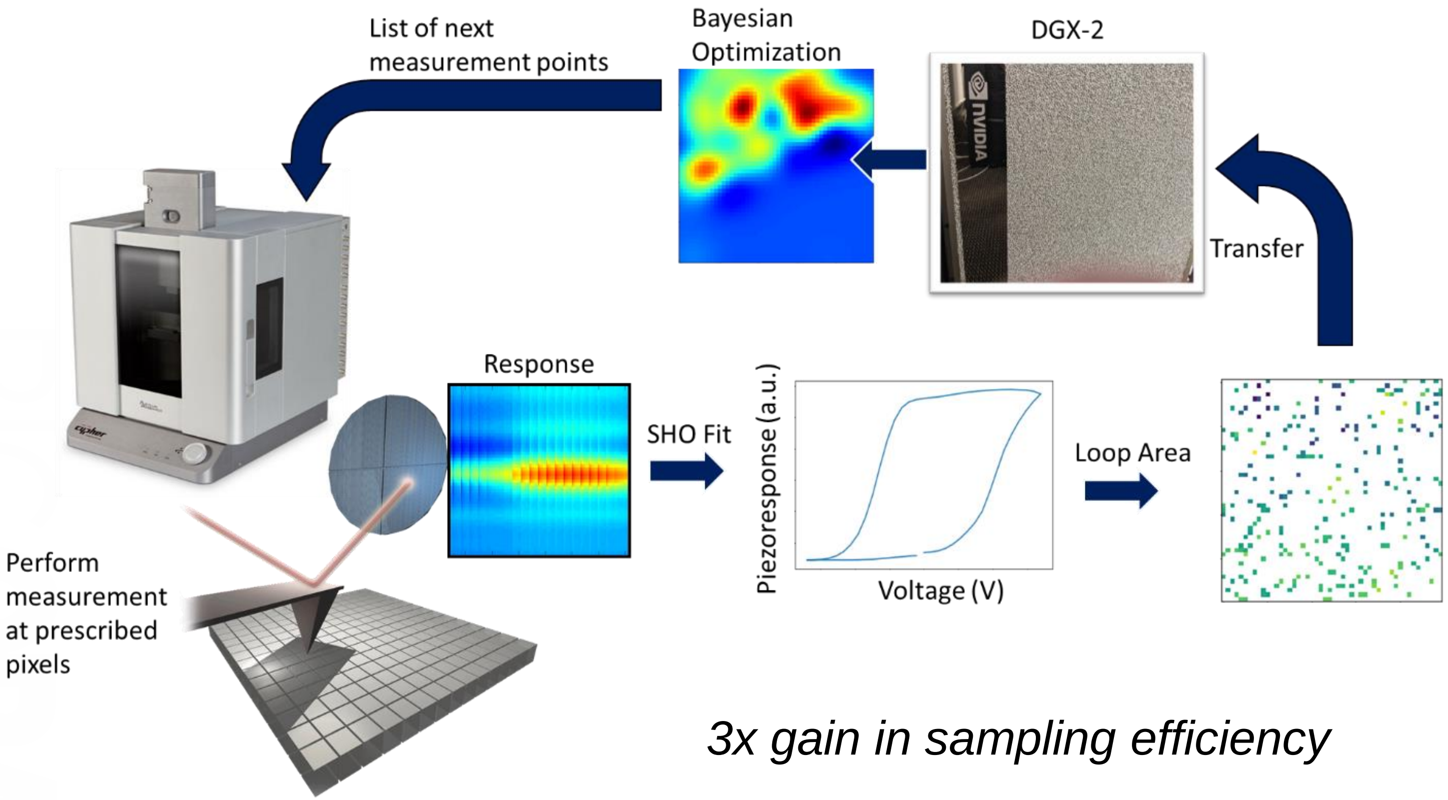


Going real time: automated experiment



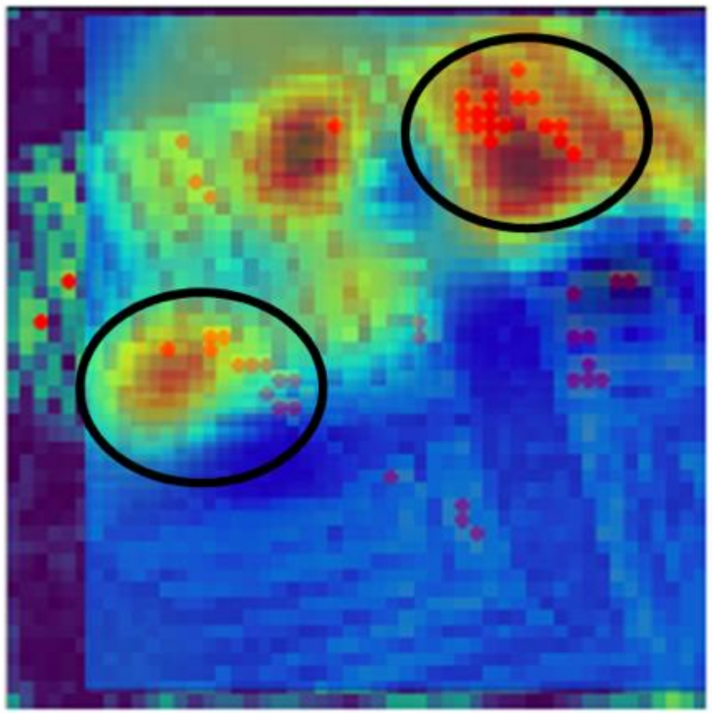
- AE based on structural analysis for STEM data
- AE based on spectral data in PFM
- AE based on DL for EELS data
- Feature of interest finding for mesoscopic images

BO for Self-Driving Microscopy



3x gain in sampling efficiency

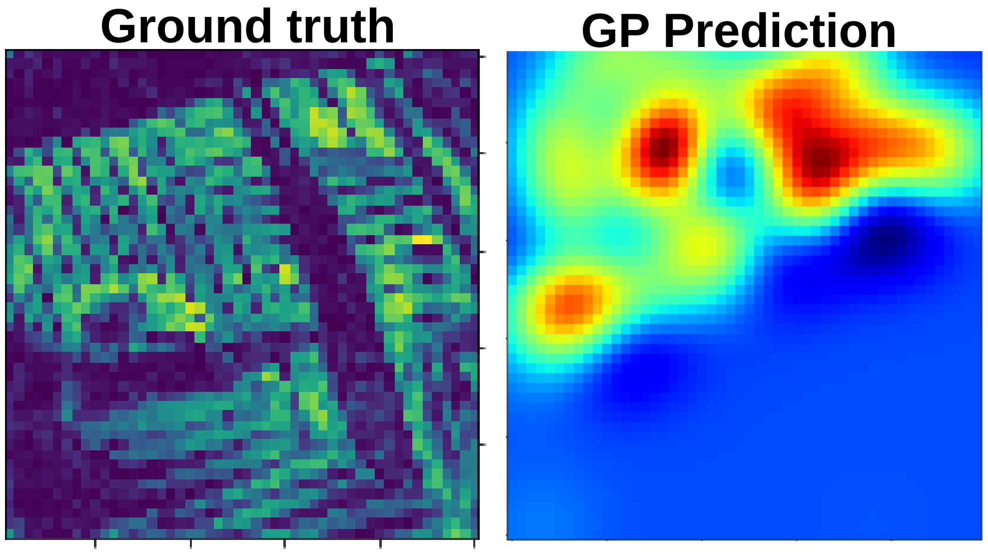
Comparison with "ground truth"



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

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)