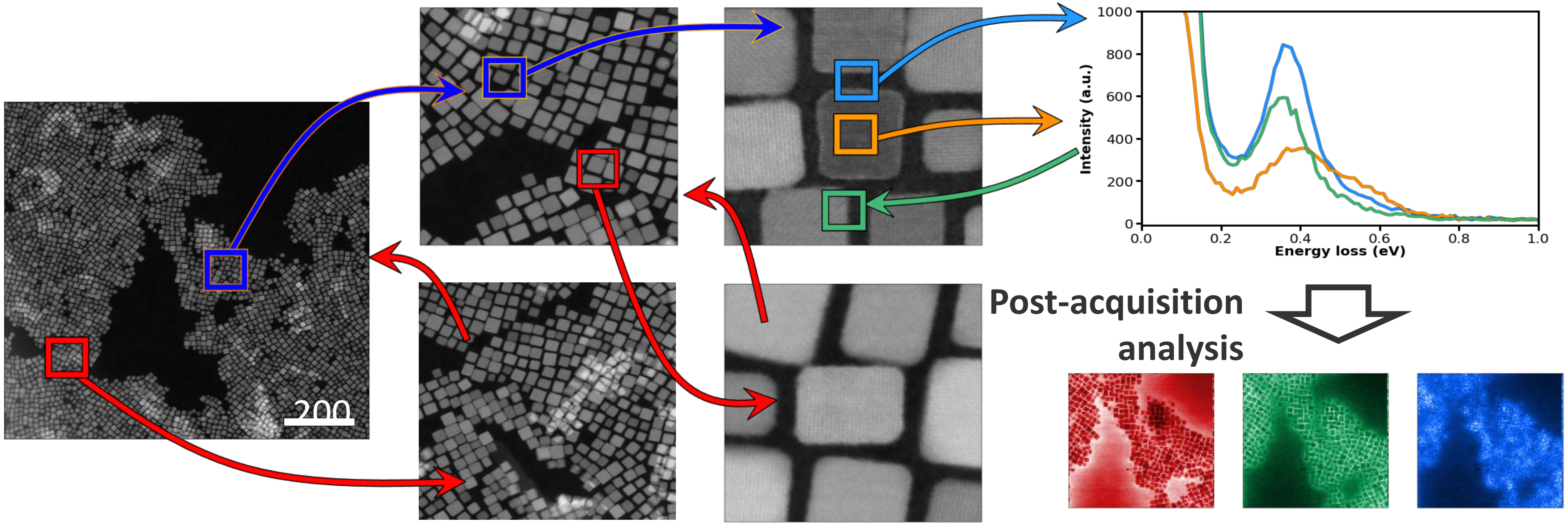
THE UNIVERSITY of TENNESSEE KNOXVILLE

Gaussian Processes and Bayesian Optimization for Automated Experiment

Sergei V. Kalinin

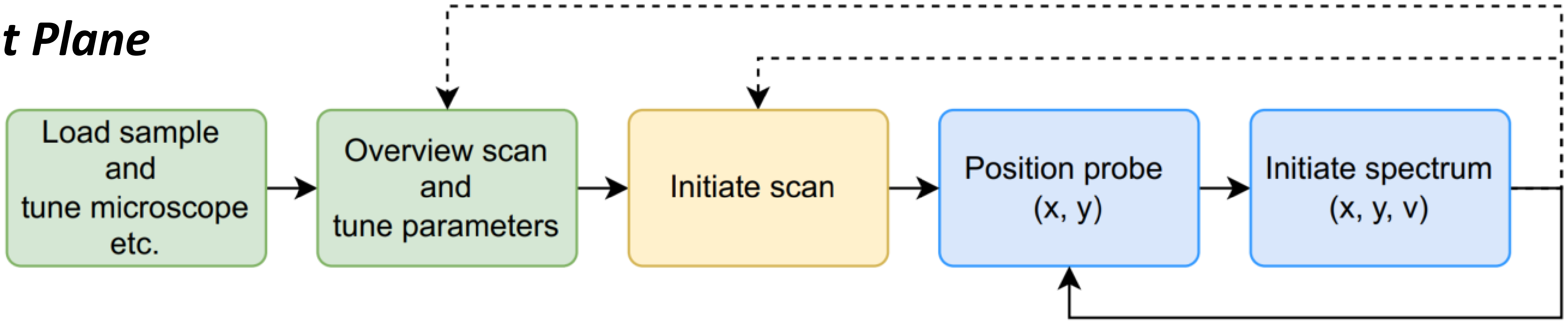
Workflows in STEM



Post-acquisition analysis

Minimal instruction set control language

Instrument Plane



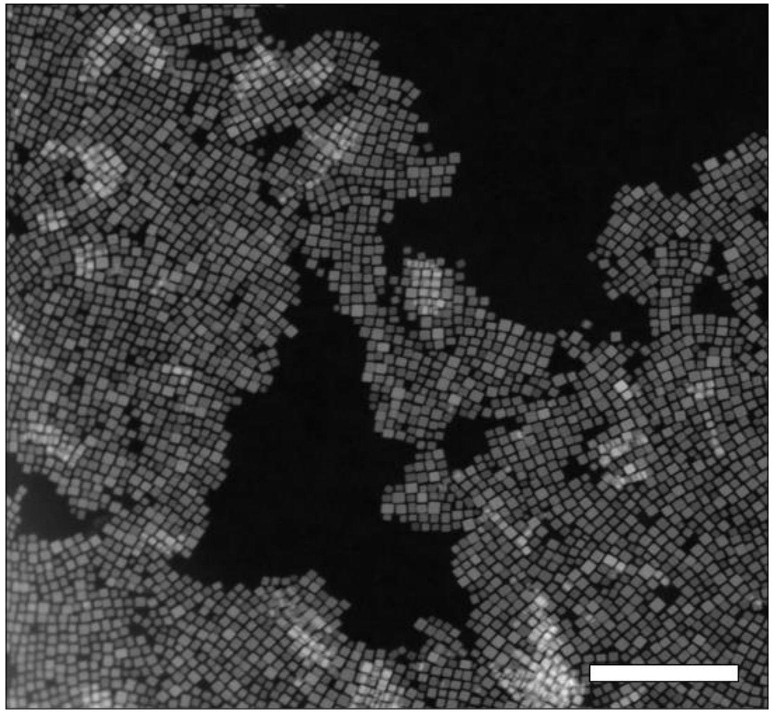
Load sample and tune microscope etc.

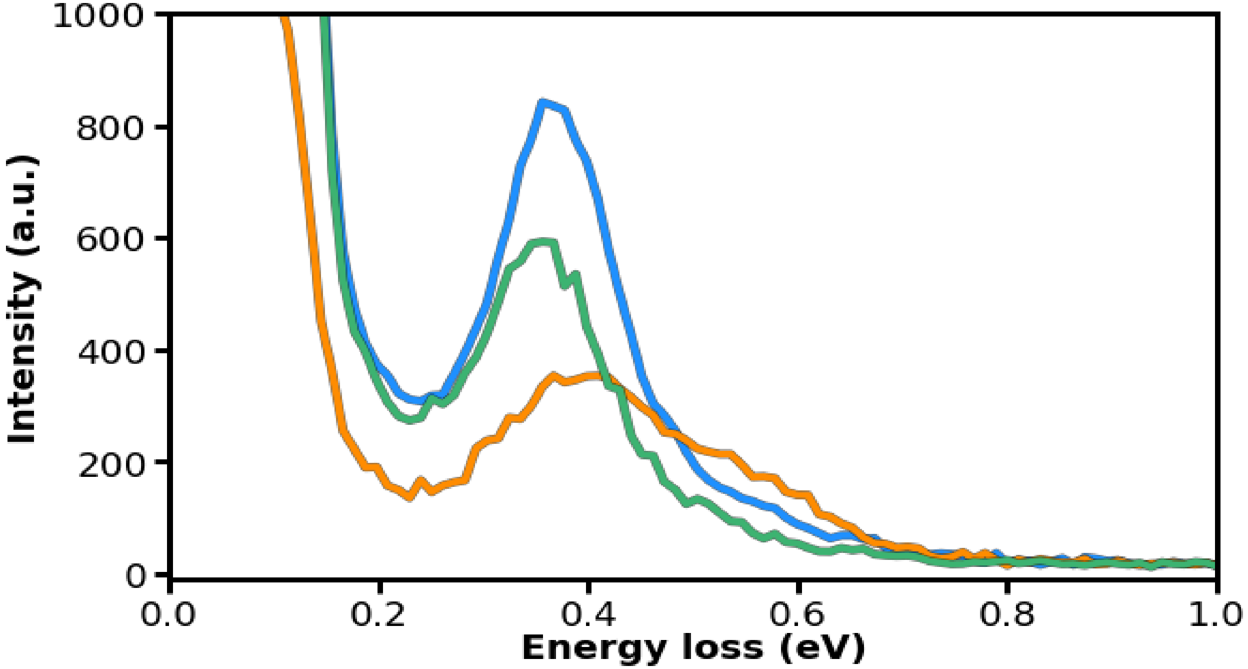
Overview scan and tune parameters

Initiate scan

Position probe (x, y)

Initiate spectrum (x, y, v)

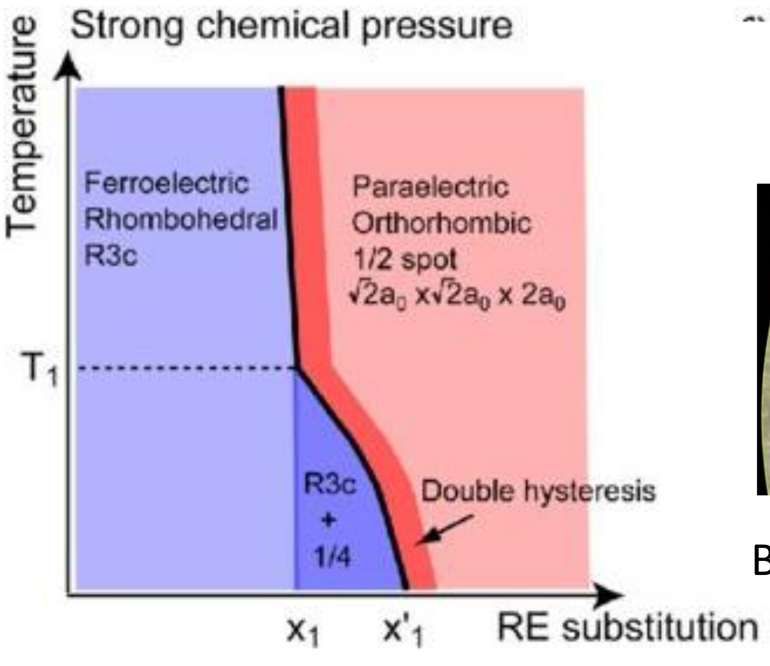




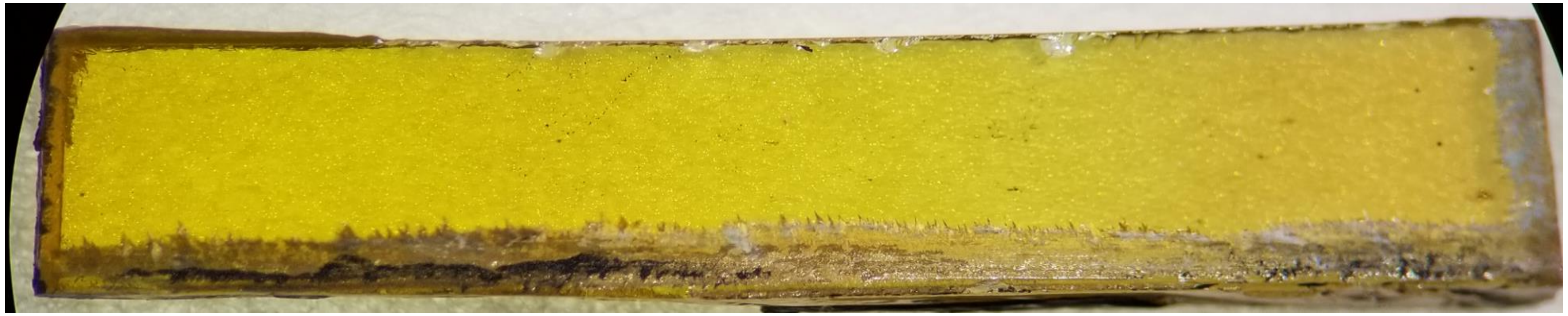
- What can we say about structure?
- Interesting functionalities are expected at the certain structural elements
- We can guess some; we have to discover others
- Multiple goals while running experiment

- **Objective: purpose of the experiment (outside context)**
- **Policy: what do we do depending on observation**
- **Reward: what do we hope to achieve**
- **Value: anticipated reward**

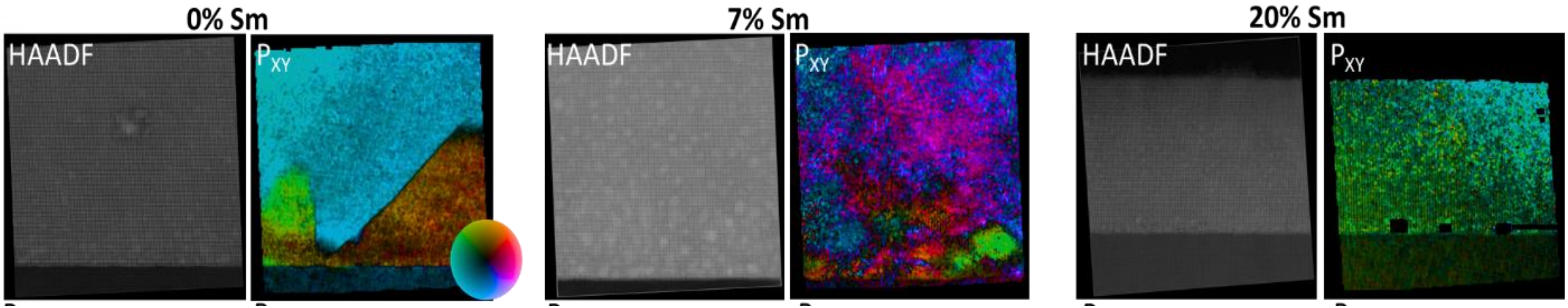
Combinatorial library



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



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



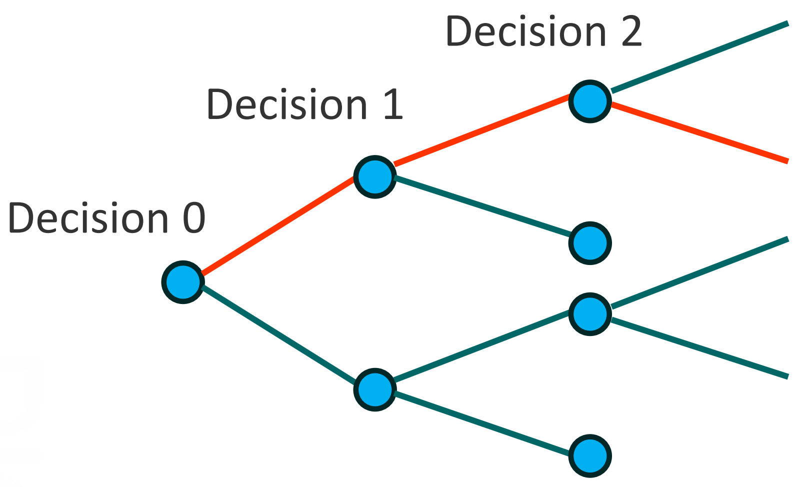
0% Sm

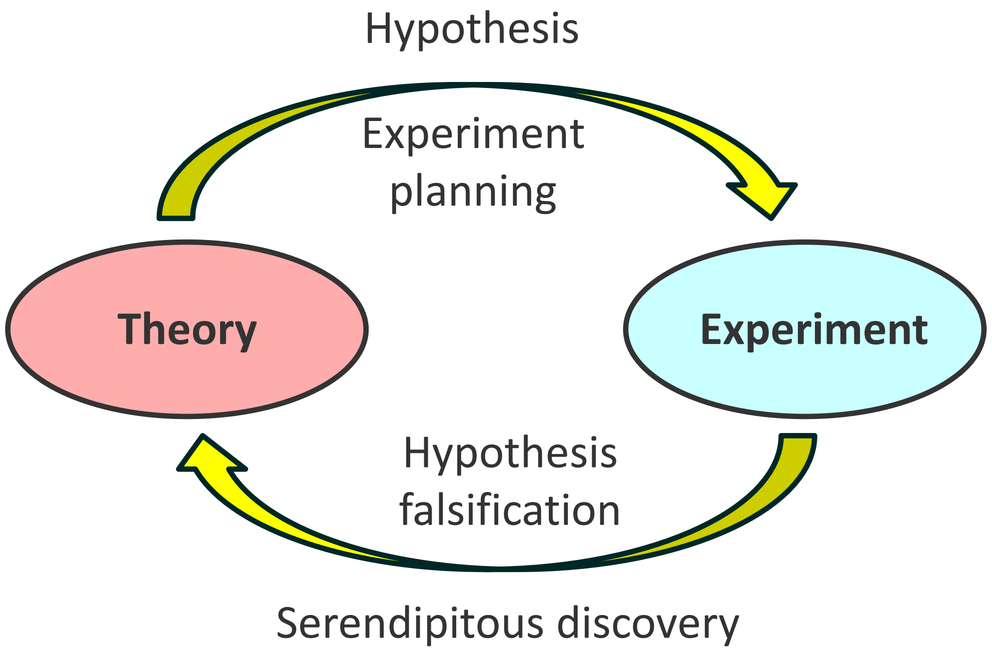
7% Sm

20% Sm

arXiv:2004.11817

What do we hope to achieve?





- **Experiment is a combinatorial space of opportunities:**

  → Investing only in scaling of throughput is only a linear improvement
- **Science is a cycle between theory-driven hypothesis generation and experiment**:

  → We need to accelerate all elements of the cycle
- **Experimental and computational tool development:**

  → Currently constrained by human paradigm

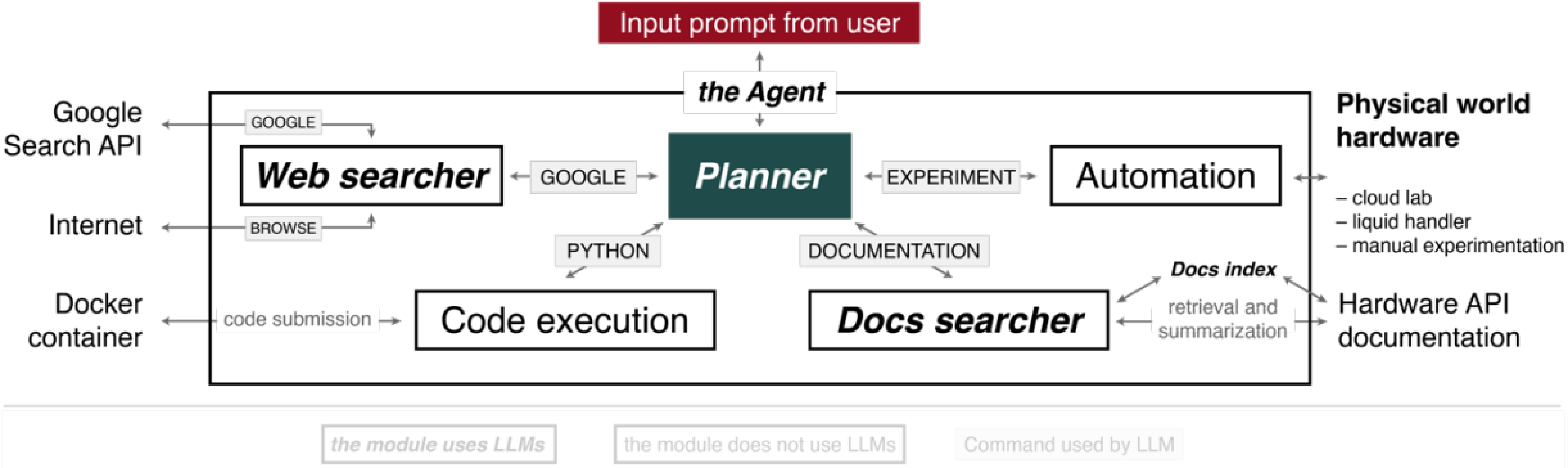
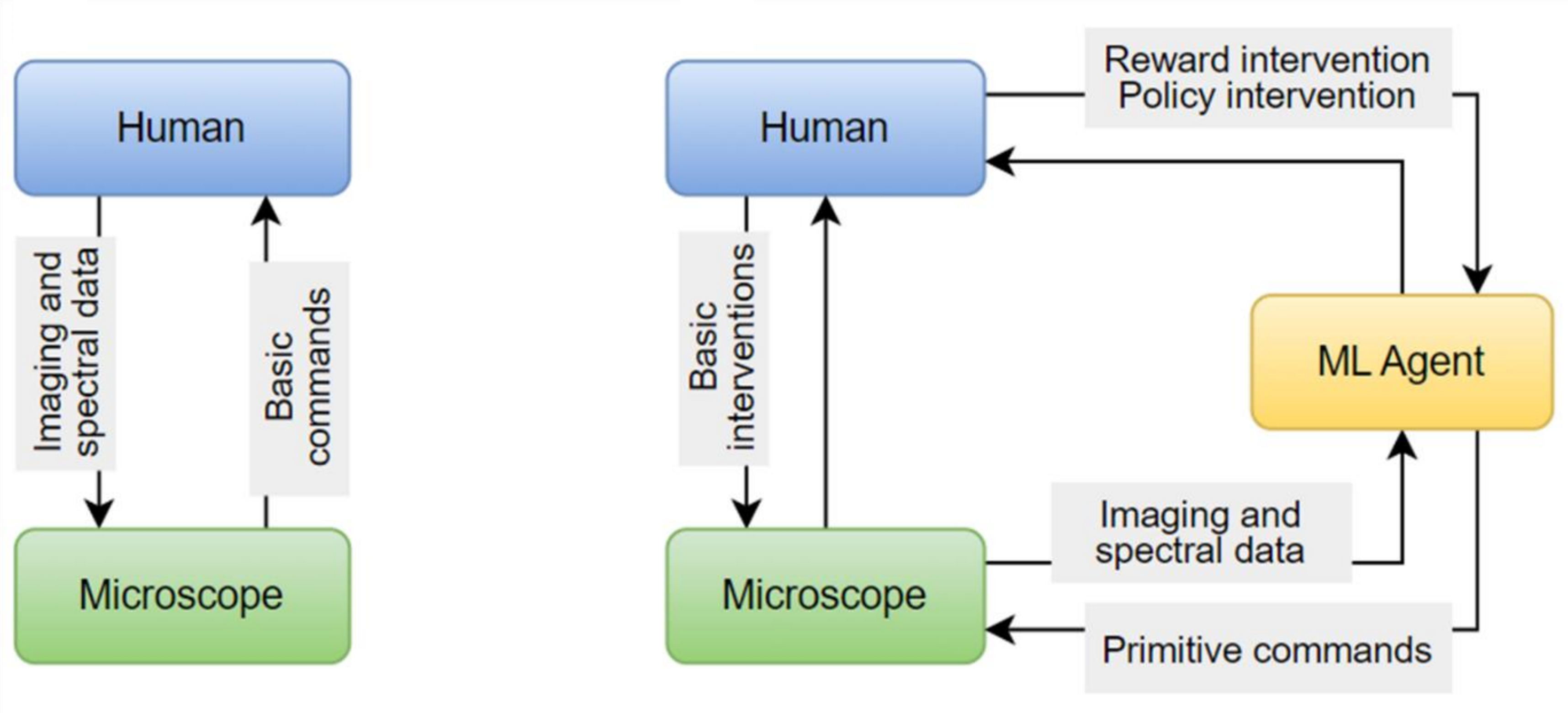


Figure 1. Overview of the system architecture. The **Agent** is composed of multiple modules that exchange messages. Some of them have access to APIs, the Internet, and Python interpreter.

Daniil A. Boiko, Robert MacKnight, Gabe Gomes, Emergent autonomous scientific research capabilities of large language models, arXiv:2304.05332

- **Foundational models have access only to past knowledge:**
  → Integration of past knowledge, but only part of the answer

There is no single solution



- **Humans are (collectively) very good at human-level tasks:**
  → Transition to ML in the discovery loop is non-trivial
- **Myopic optimization is a necessary start, but cannot be expected to yield large gains**:
  → We need beyond zero horizon optimization
- **ML driven tools have very different controls then human-driven:**
  → Transition from horse to the car

Machine Learning for Automated Experiments

1. Machine learning is now everywhere…. But it is difficult to see it in the lab!

2. Applying ML in experimental sciences can be a very daunting proposition:
   a. Understand the problem
   b. Know (some) ML
   c. Develop code: from prototype to operationalization
   d. Implement on the working instrument
   e. Connect to the cloud
   f. Understand the results

3. The code base and libraries can change (TF/Keras -> PyTorch -> Jax -> ?)

4. Basic concepts are often incomprehensible/unfamiliar (policy, reward, acquisition function, myopic policy)

5. For ML community, experimental sciences are complex as well (small data, causal, physical laws, active learning)

Machine Learning for Automated Experiments

1. We are not going to try to learn all aspects of it. ML in domain/experiment is always a teamwork.

2. The first step is to define your problem – what is that you want to accomplish?

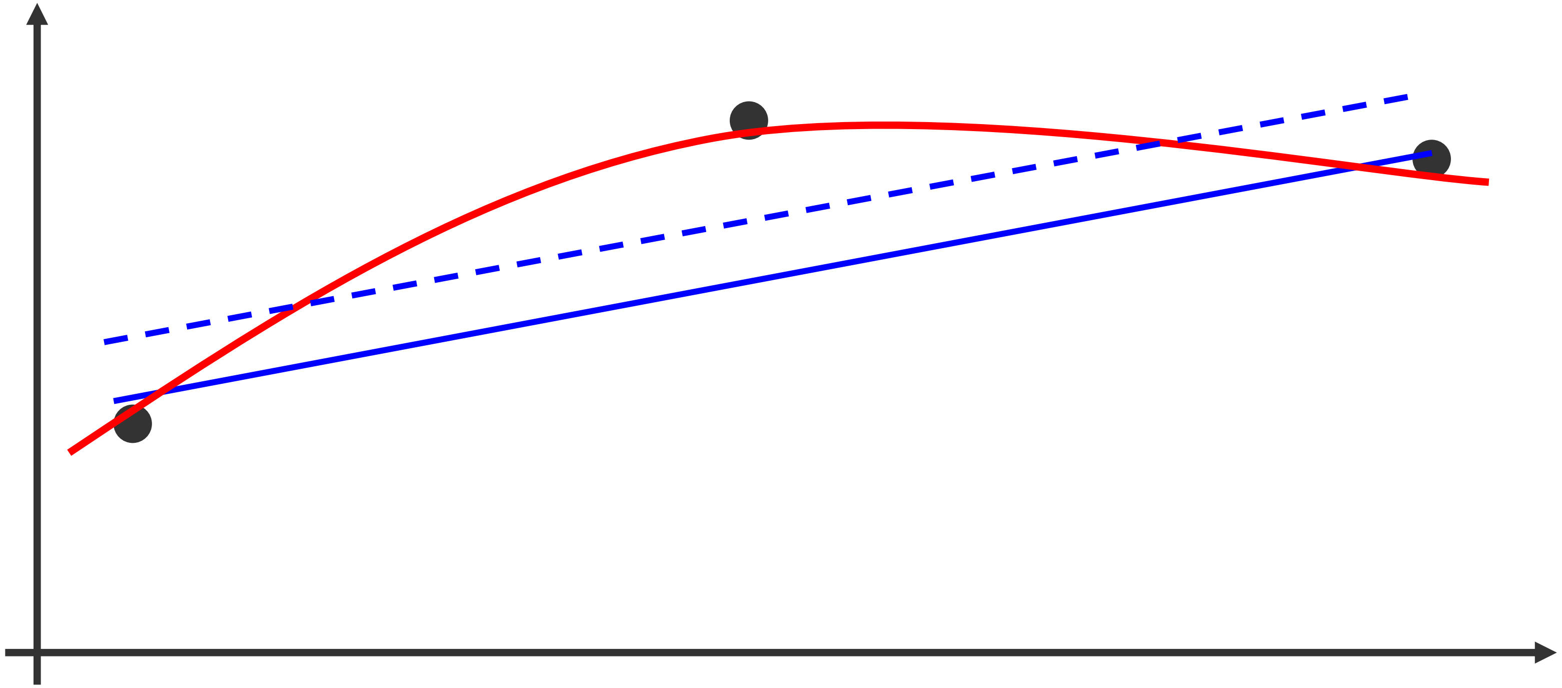
3. The second step is to define your hyper language – what are the things that you know how to do (or can learn to do)?

4. The third step is to work backward from your problem and define the workflow in terms of your hyper language. Do you even need machine learning to solve it?

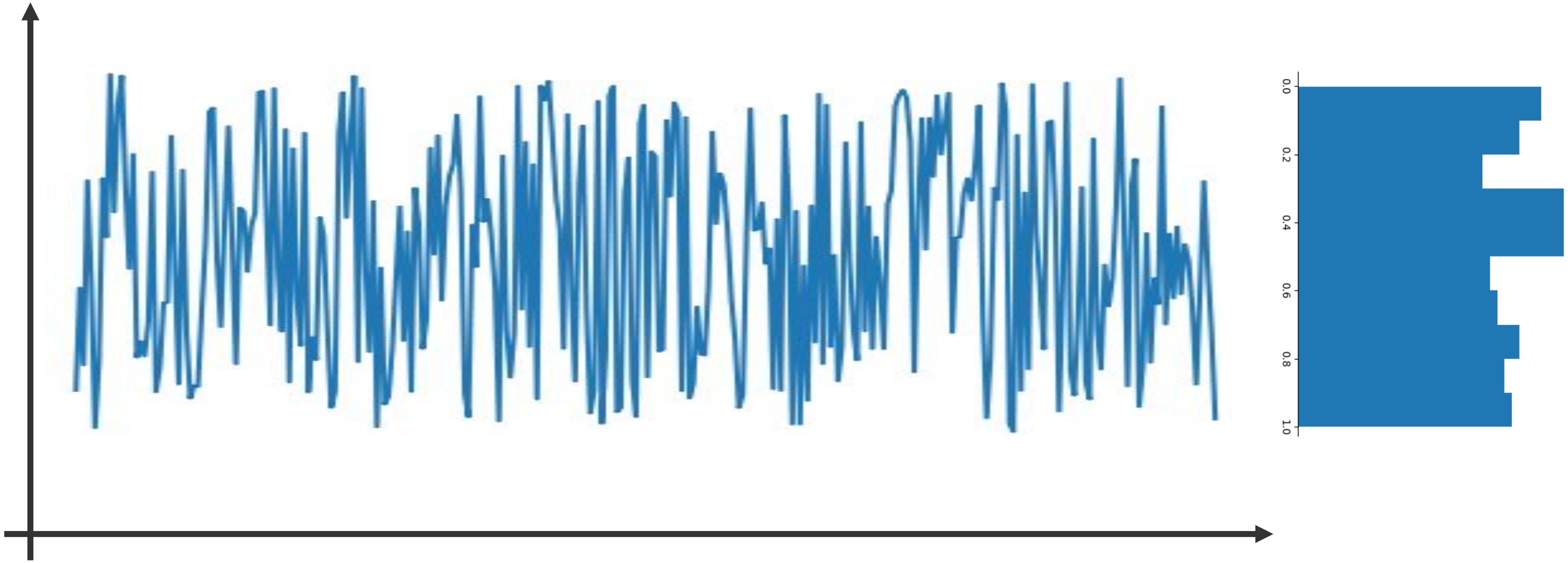
5. For ML, knowing code is (in some sense) secondary. However, it is critical to understand HOW it works.

Automated Experiment:
… with John Snow priors…

What do we know if we do not know anything?



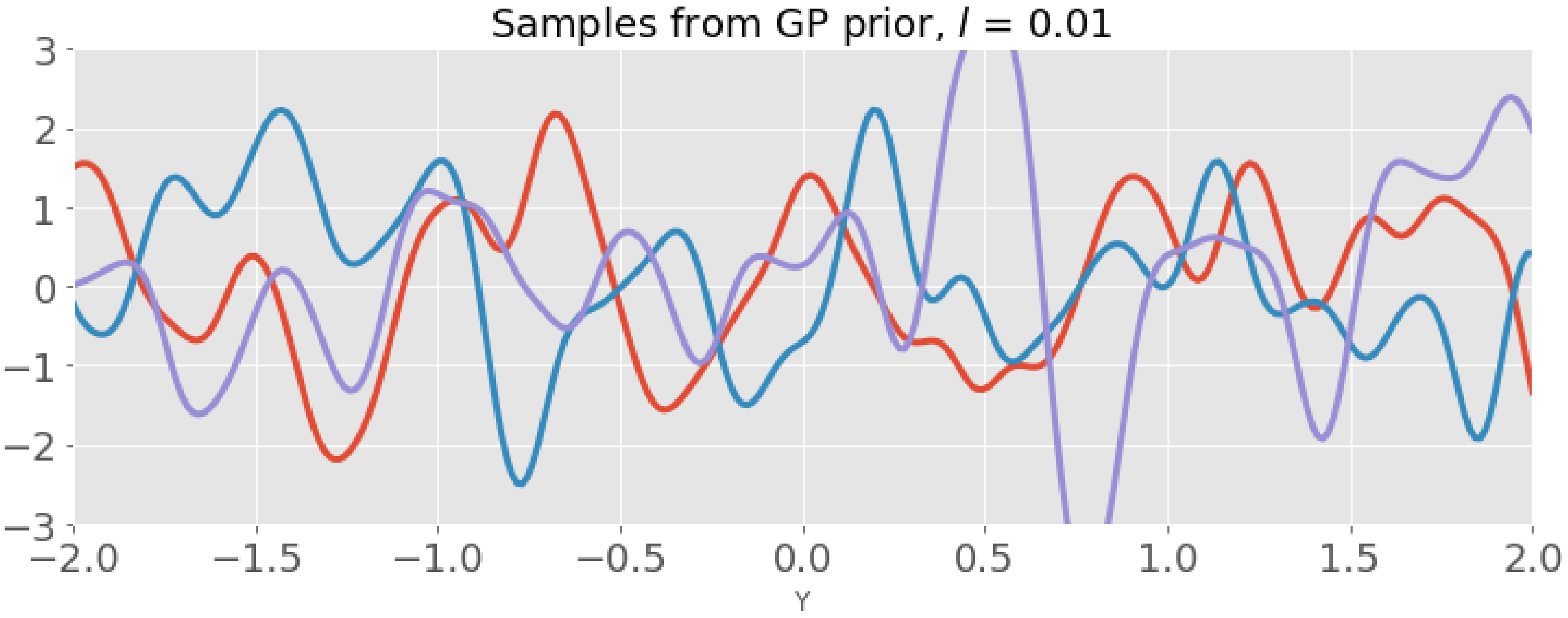
What do we know if we do not know anything?



Gaussian Process Regression

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

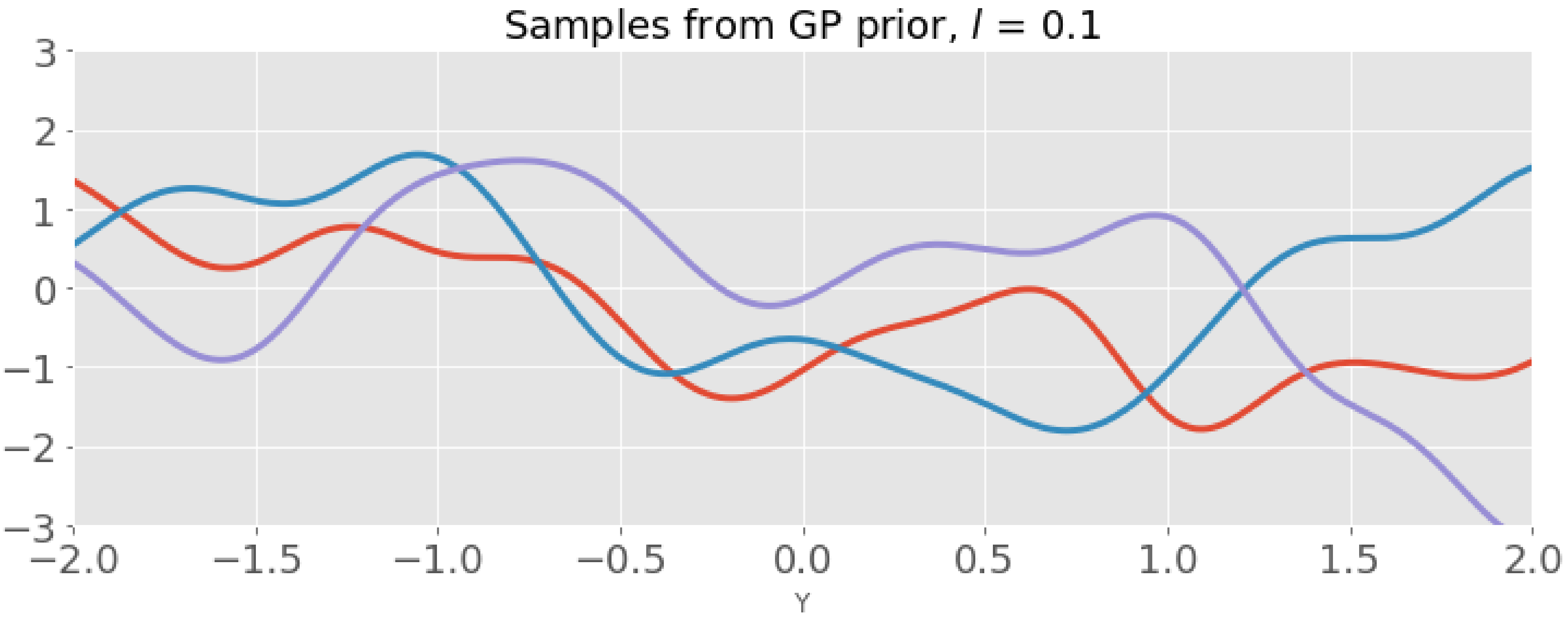
$$\mathrm{k(x,x')} = \exp\left(-\frac{1}{2l}(x-x')^2\right)$$



Gaussian Process Regression

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

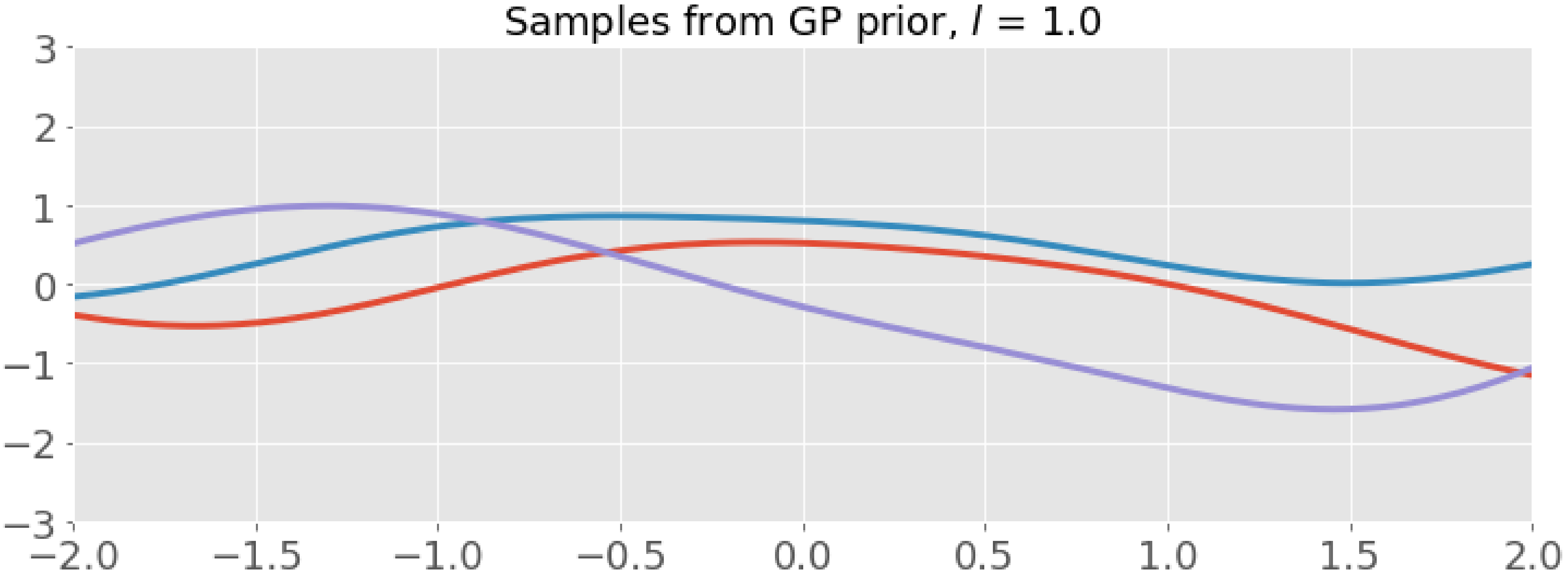
$$\mathrm{k}(\mathrm{x},\mathrm{x}') = \exp\left(-\frac{1}{2l}(x-x')^2\right)$$



Gaussian Process Regression

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

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



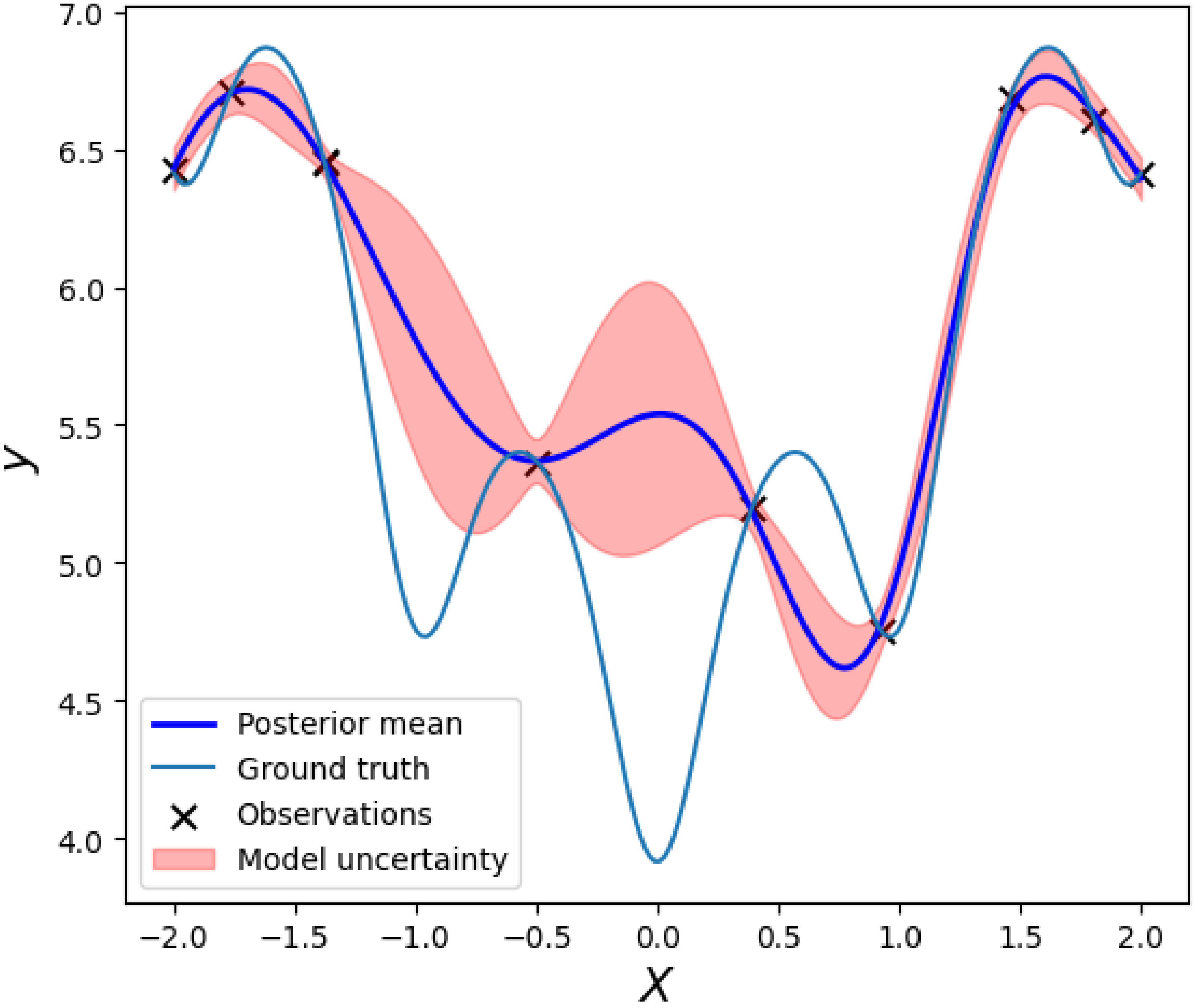
L controls the length scale – sort of how far points should be to make them independent of each other.

What have we learned from Part 1:

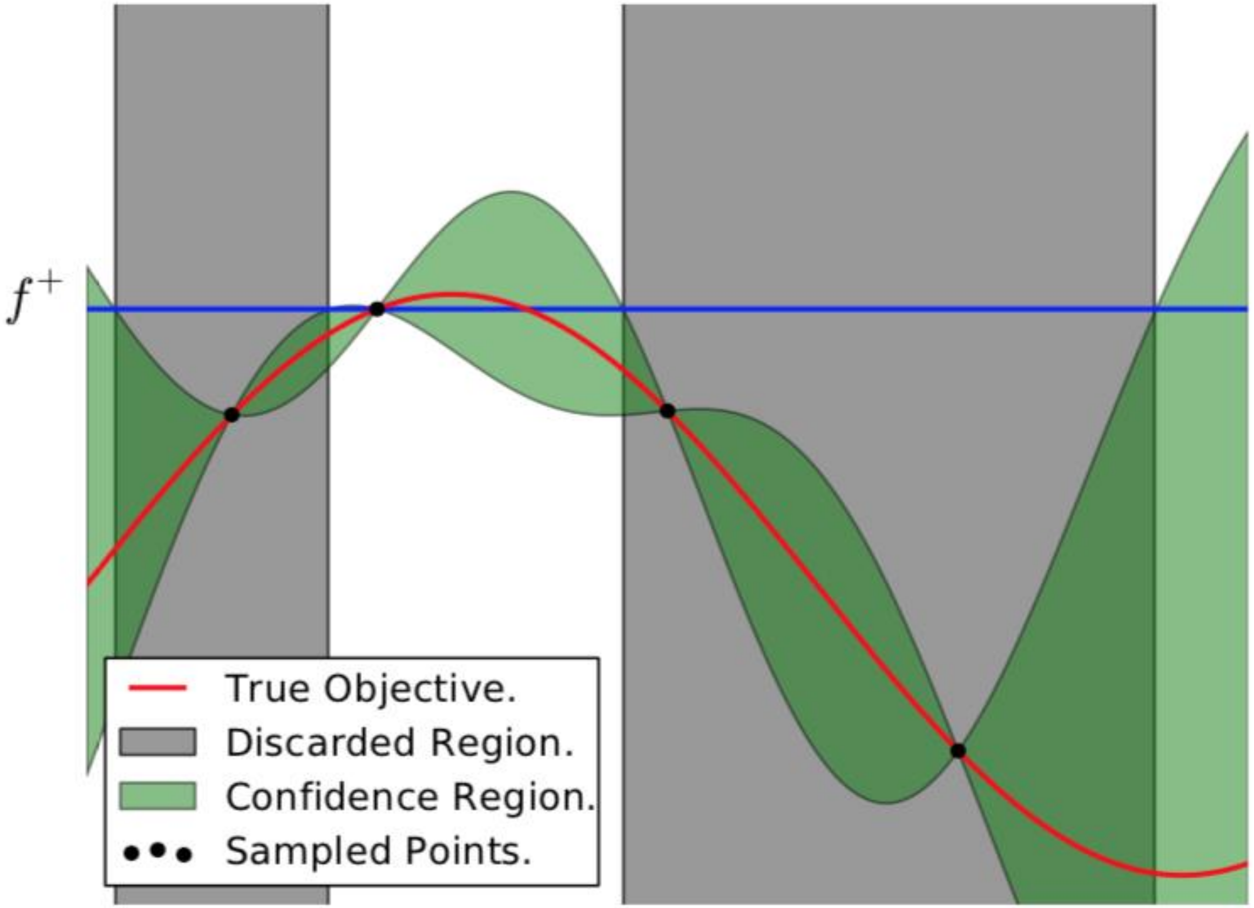
- Gaussian Process
- Kernel and kernel parameters
- Kernel Priors
- Noise Priors
- Posteriors

What will we learn in part 2:

- Bayesian Optimization
- Bayesian Optimization based on Gaussian Process
- Acquisition Functions


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

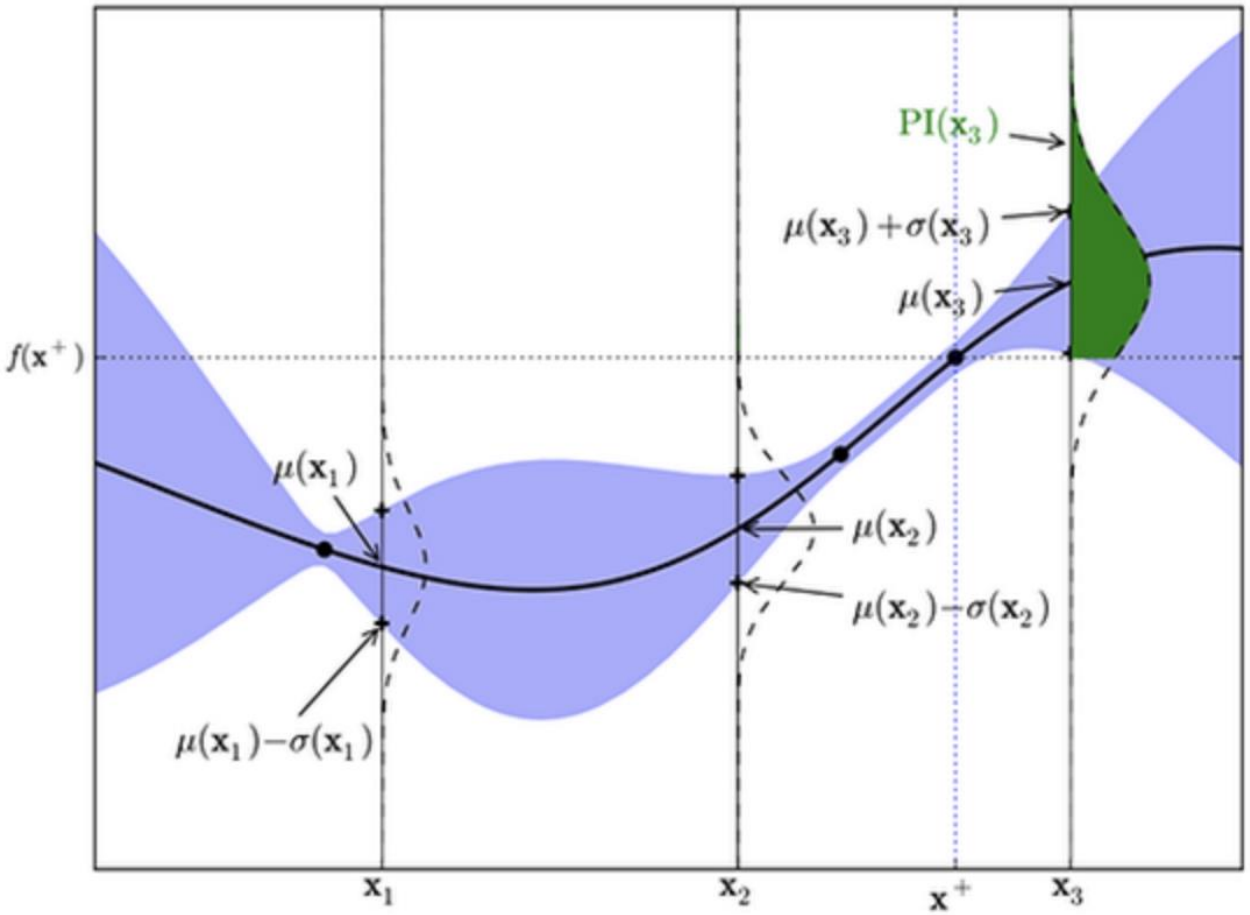


- We have some measurements in space X, and we want to maximize some property f(X).
- How can we decide what point to measure next to best maximize f?
- We need to balance the exploration of the space with exploitation of regions near we have already know

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

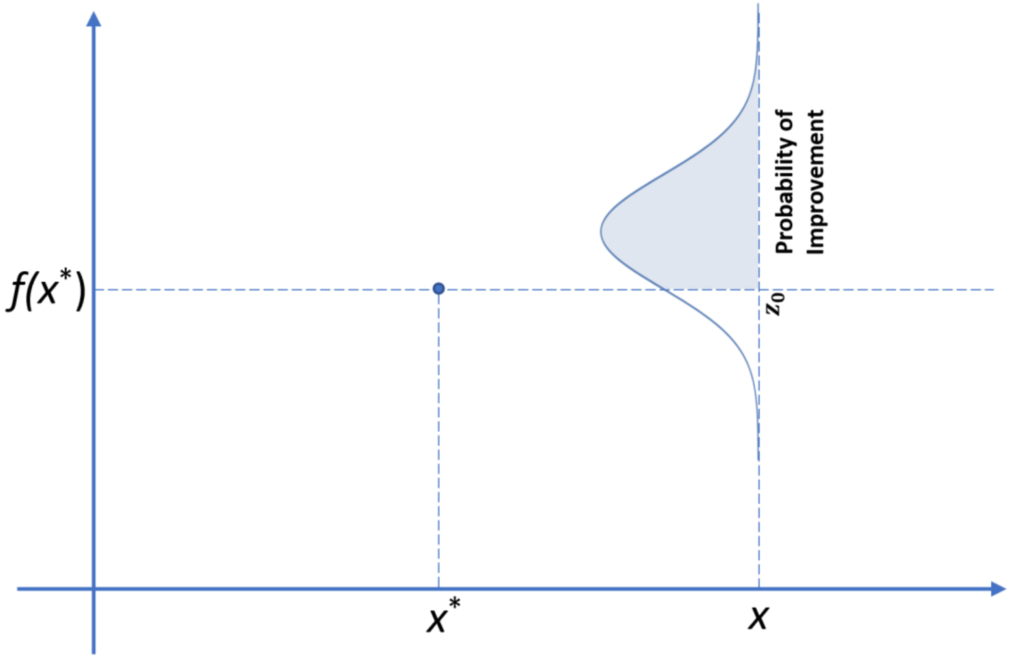
4. **There are (always) more…**

Expected improvement is defined as

$$\mathrm{EI}(\mathbf{x}) = \mathbb{E}\max(f(\mathbf{x}) - f(\mathbf{x}^+), 0)$$

where $f(\mathbf{x}^+)$ is the value of the best sample so far and $\mathbf{x}^+$ is the location of that sample i.e. $\mathbf{x}^+ = \mathrm{argmax}_{\mathbf{x}_i \in \mathbf{x}_{1:t}} f(\mathbf{x}_i)$. The expected improvement car evaluated analytically under the GP model[3]:

$$\mathrm{EI}(\mathbf{x}) = \begin{cases} (\mu(\mathbf{x}) - f(\mathbf{x}^+) - \xi)\Phi(Z) + \sigma(\mathbf{x})\phi(Z) & \text{if } \sigma(\mathbf{x}) > 0 \\ 0 & \text{if } \sigma(\mathbf{x}) = 0 \end{cases}$$



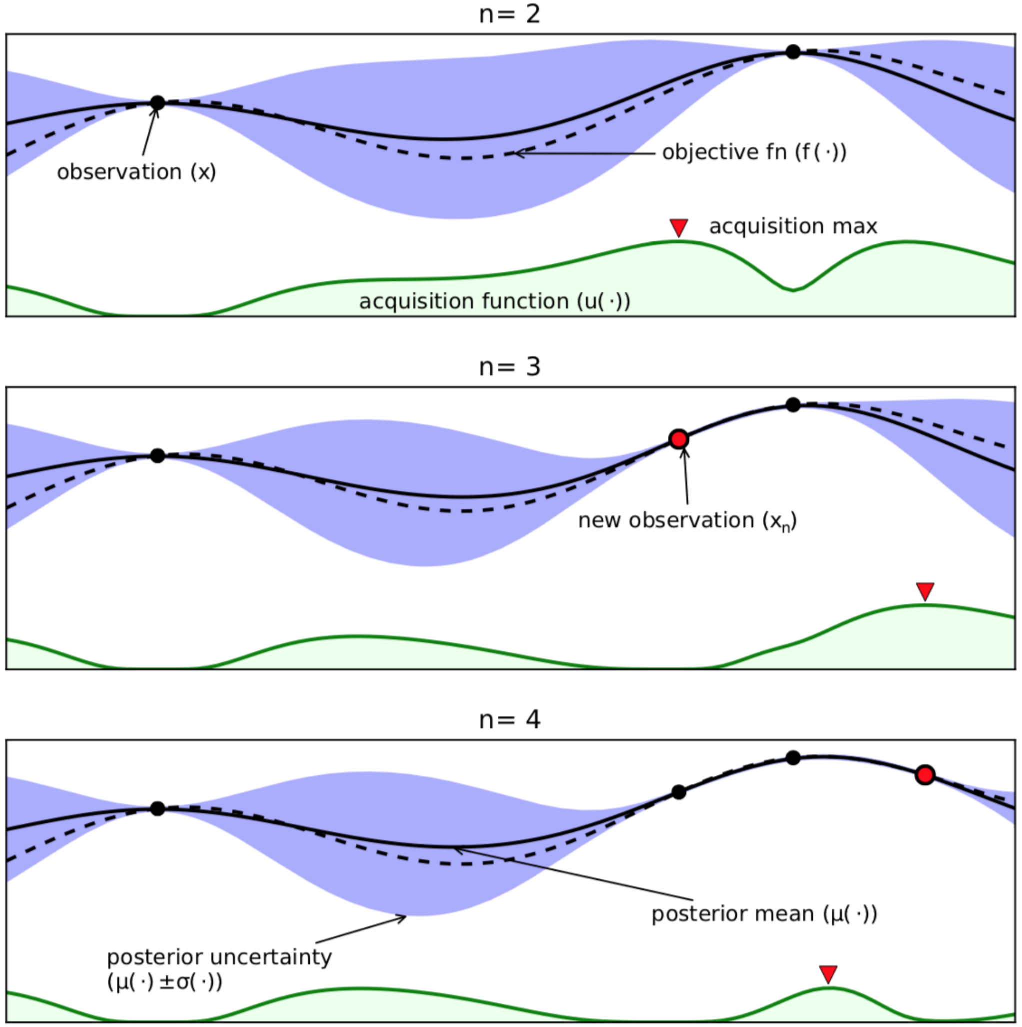
where

$$Z = \begin{cases} \frac{\mu(\mathbf{x}) - f(\mathbf{x}^+) - \xi}{\sigma(\mathbf{x})} & \text{if } \sigma(\mathbf{x}) > 0 \\ 0 & \text{if } \sigma(\mathbf{x}) = 0 \end{cases}$$

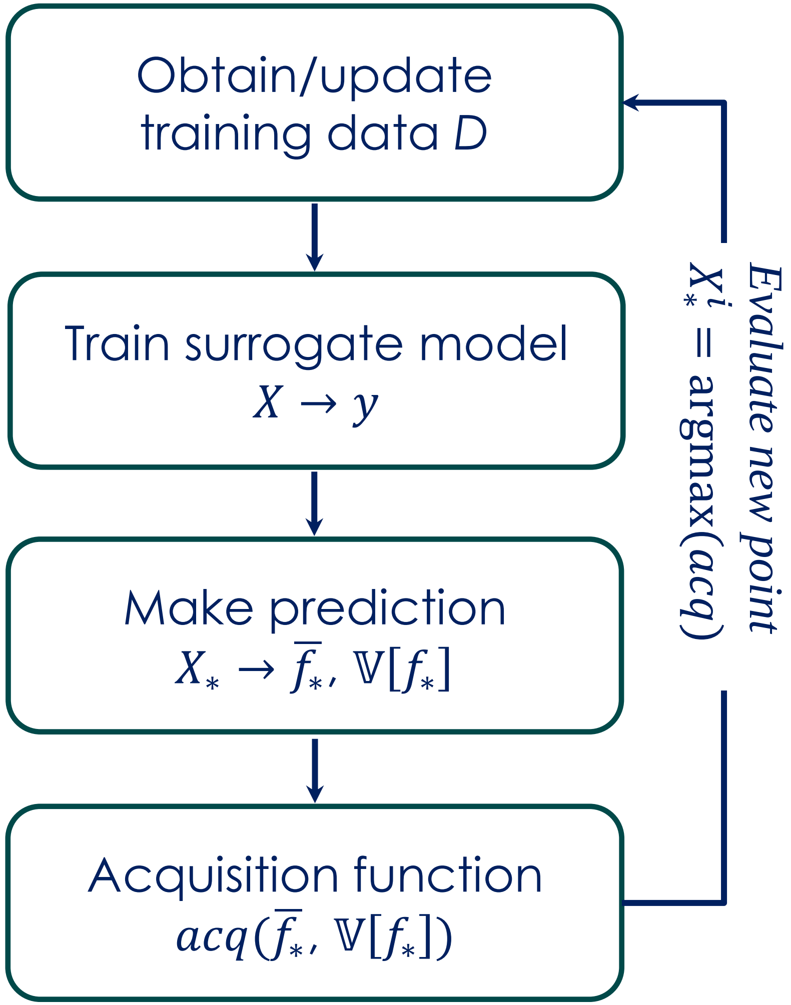
where $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are the mean and the standard deviation of the GP posterior predictive at $\mathbf{x}$, respectively. Φ and ϕ are the CDF and PDF of the standard normal distribution, respectively. The first summation term in Equation (2) is the exploitation term and second summation term is the exploration term.

Parameter ξ in Equation (2) determines the amount of exploration during optimization and higher ξ values lead to more exploration. In other words, with increasing ξ values, the importance of improvements predicted by the GP posterior mean $\mu(\mathbf{x})$ decreases relative to the importance of potential improvements in regions of high prediction uncertainty, represented by large $\sigma(\mathbf{x})$ values. A recommended default value for ξ is 0.01.

The basics: Bayesian Optimization



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



Obtain/update training data *D*

Train surrogate model $X \to y$

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

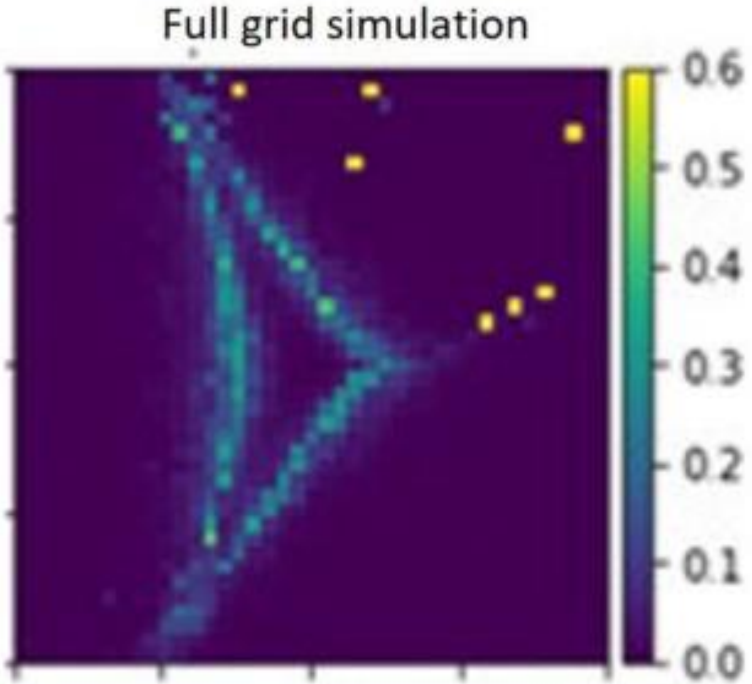
Acquisition function $acq(\bar{f}_*, \mathbb{V}[f_*])$

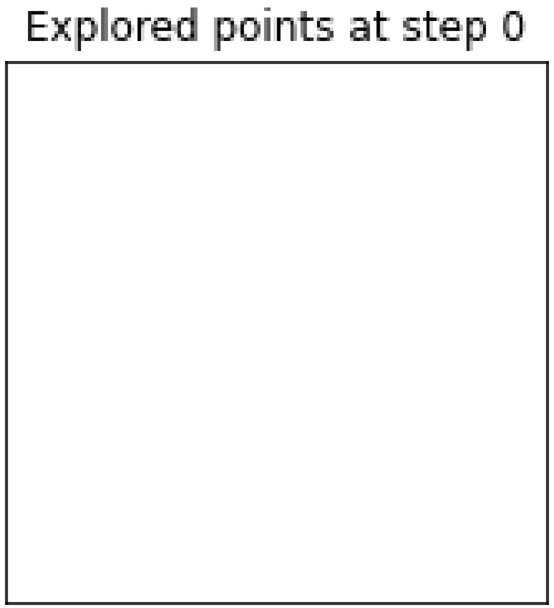
Evaluate new point
$X_*^i = \text{argmax}(acq)$

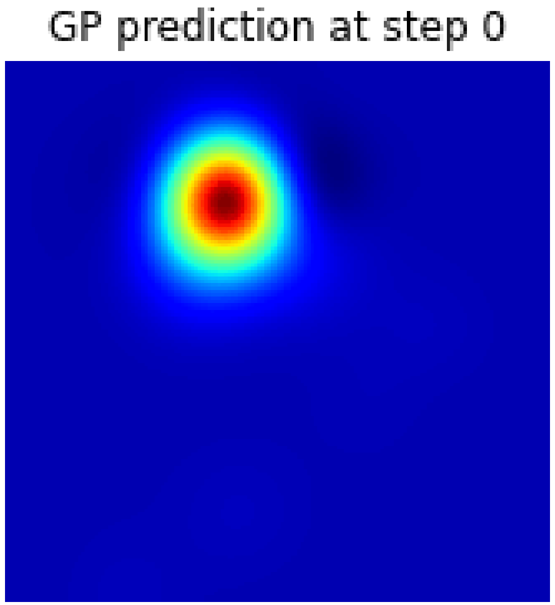
N. de Freitas et al., Taking the Human Out of the Loop: A Review of Bayesian Optimization , *Proceedings of the IEEE* **104**, 148 (2015)

Bayesian Optimization for physical discovery

Discovering regions where heat capacity is maximized in NNN Ising model


Full grid simulation


Explored points at step 0


GP prediction at step 0