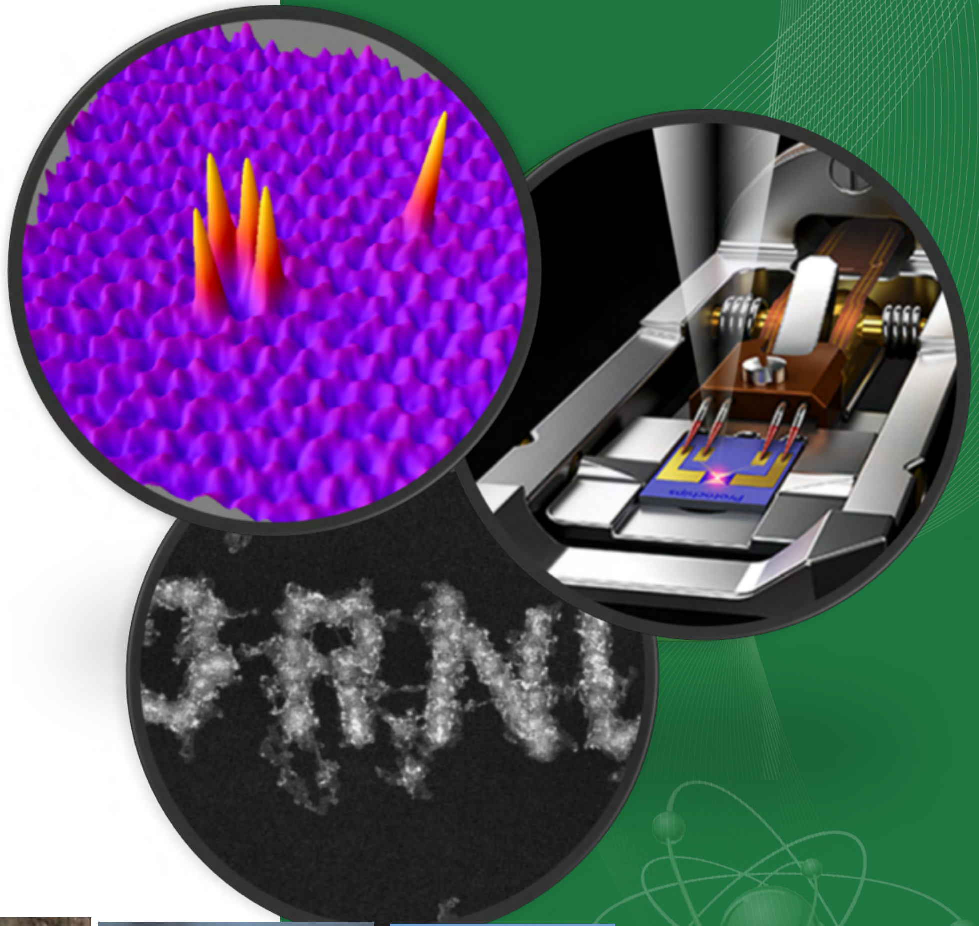
Machine Learning for Scanning Probe and Electron Microscopy for Materials Discovery

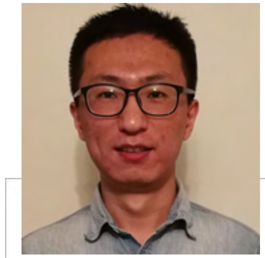


Sergei V. Kalinin

The Center for Nanophase Materials Sciences (CNMS),
Oak Ridge National Laboratory

January 24, 2022

Mahshid Ahmadi



Yongtao Liu

Rama Vasudevan

Maxim Ziatdinov

Kyle Kelley

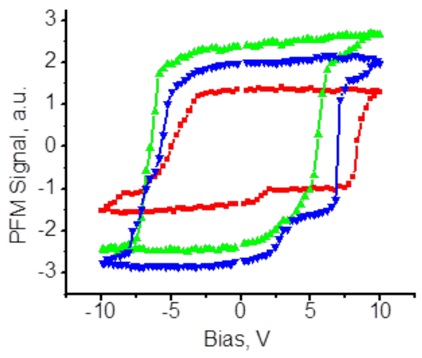
Ondrej Dyck

Kevin Roccapriore

OAK RIDGE National Laboratory


SCANNING PROBE MICROSCOPY
Electrical and Electromechanical Phenomena at the Nanoscale
Sergei Kalinin
Alexei Gruverman
EDITORS
VOLUME 1
Springer
Sergei V. Kalinin
Alexei Gruverman
Editors
Scanning Probe Microscopy of Functional Materials
Springer
7
World Scientific Series in Nanoscience and Nanotechnology
Scanning Probe Microscopy for Energy Research
Dawn A Bonnell • Sergei V Kalinin
World Scientific
Volume 1
World Scientific Series on Emerging Technologies – Volume 1
HANDBOOK ON BIG DATA AND MACHINE LEARNING IN THE PHYSICAL SCIENCES
In 2 Volumes
Volume 1
Big Data Methods in Experimental Materials Discovery
Editors
Surya R. Kalidindi • Sergei V. Kalinin
Turab Lookman
Editors-in-Chief
Sergei V. Kalinin • Ian Foster
World Scientific
HANDBOOK ON BIG DATA AND MACHINE LEARNING IN THE PHYSICAL SCIENCES
Vol. 1
Kalidindi
Kalinin
Lookman

Multidimensional SPM modes



We realized we are doing big data

- What do we learn from the data?
- How much can we believe in results?
- Can we perform experiments better?

Physics perspective:
1. Local stimulus: spectroscopic (3D) SPM modes
2. Phase transitions are hysteretic: First Order Reversal Curves
3. Phase transitions can be rate-controlled

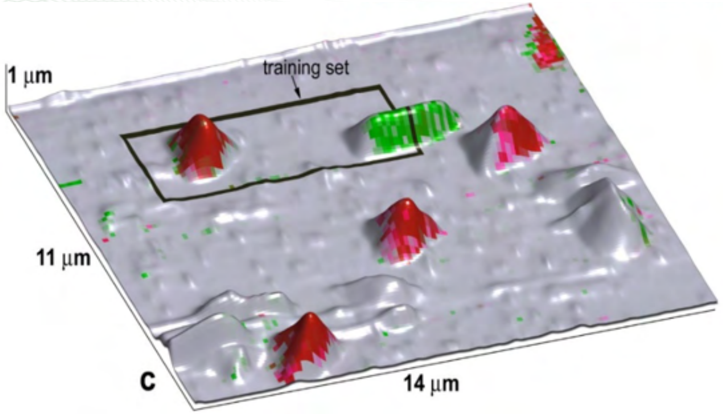
SPM Perspective:
4. SPM requires resonance enhancement

Instrumental limit: photodetector bandwidth (~10 MHz) x DAQ performance (32 Bit)
- **Single frequency/heterodyne:** compression to ~ 1 kHz
- **Band excitation:** 10^2 bins at ~ 1 kHz = 100 kH
- **G-mode:** full streaming at ~10 MHz

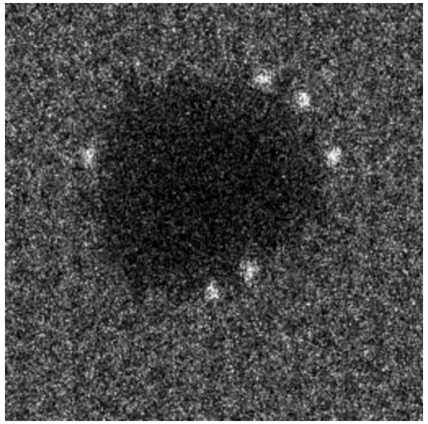
OAK RIDGE National Laboratory

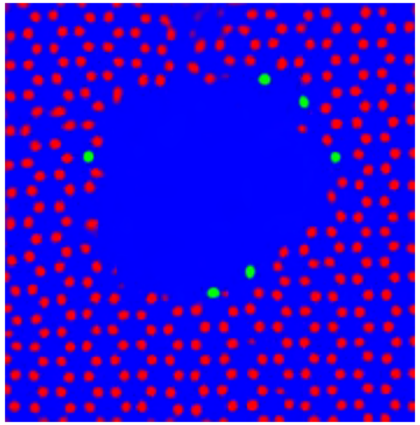
History at ORNL: Machine learning

Shallow networks for data analysis

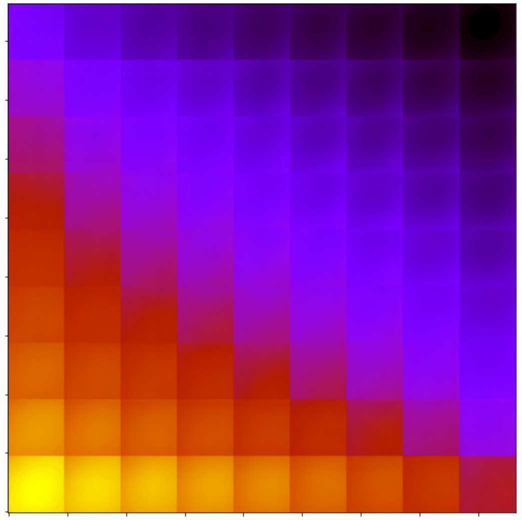


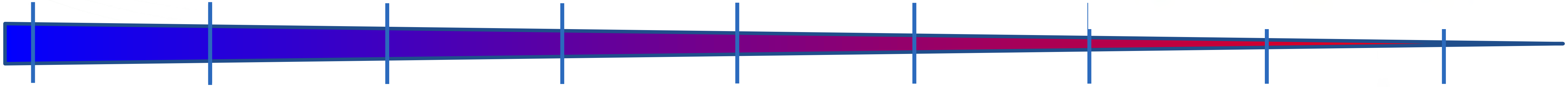
Deep networks for data analysis





Deep networks for learning physics





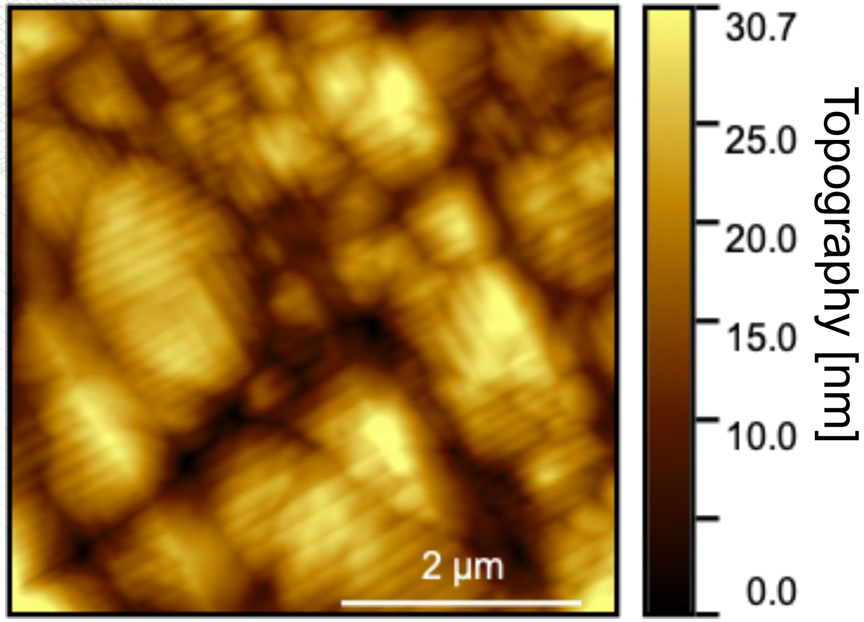
2006	2008	2010	2012	2014	2016	2018	2020	2022

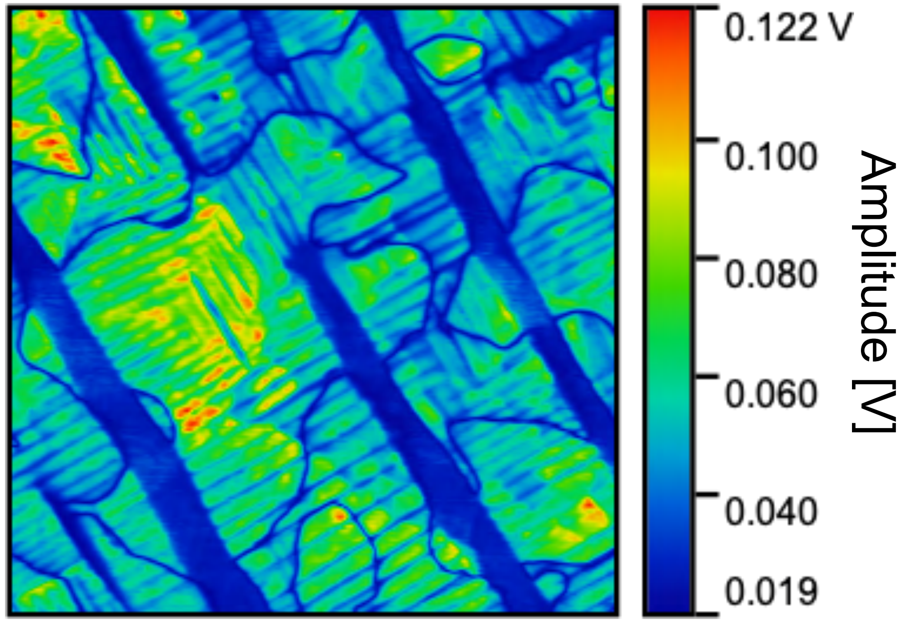
- Principal component analysis
- Neural network recognition
- NN-based theory-experiment marching
- Physics-based ML
- Deep learning
- • ELIT ML
  • Invariant autoencoders

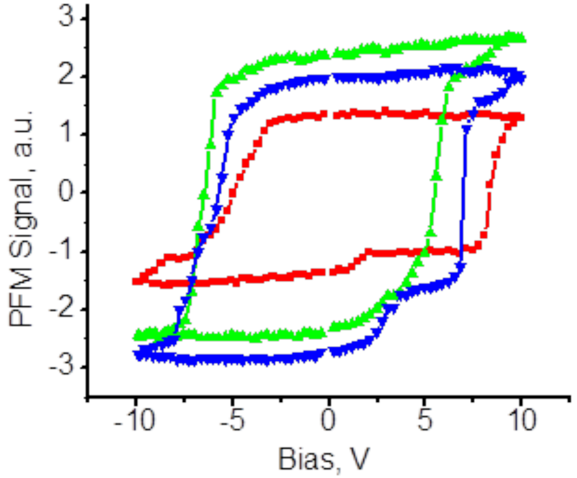
M.P. Nikiforov, A.A. Vertegel, V.V. Reukov, G.L. Thompson, S.V. Kalinin, and S. Jesse, *Functional recognition imaging using artificial neural networks: Applications to rapid cellular identification by broadband electromechanical response*, Nanotechnology **20**, 405708 (2009).

O.S. Ovchinnikov, S. Jesse, P. Bintacchit, S. TrolierMcKinstry, and S.V. Kalinin, *Disorder identification in hysteresis data: recognition analysis of random-bond random-field Ising model*, Phys. Rev. Lett. **103**, 157203 (2009).

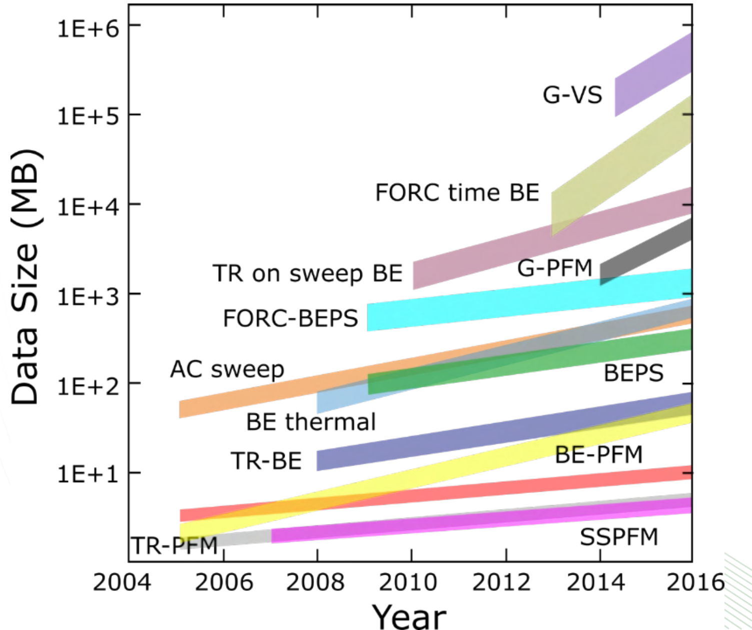
Not all regions are equally interesting!







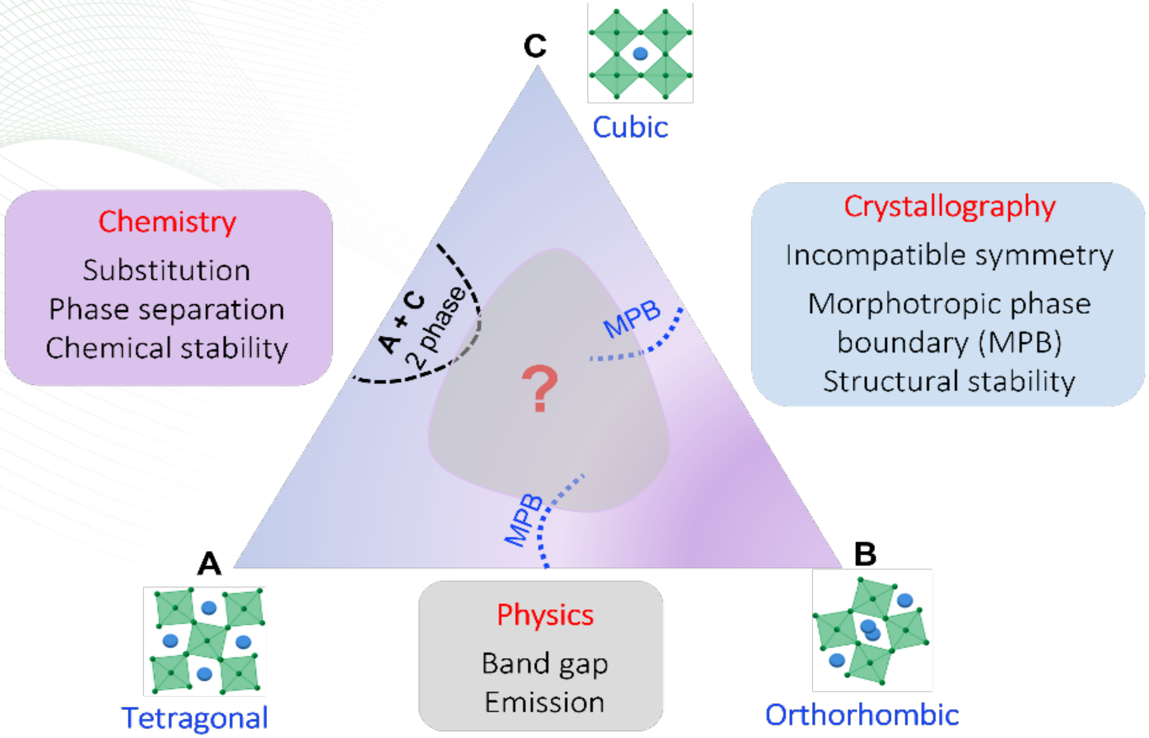
- Interesting functionalities are expected at the certain elements of domain structure
- We can guess some; we have to discover others
- Can we run PFM so that we either explore only selected regions, or discover new functionalities?

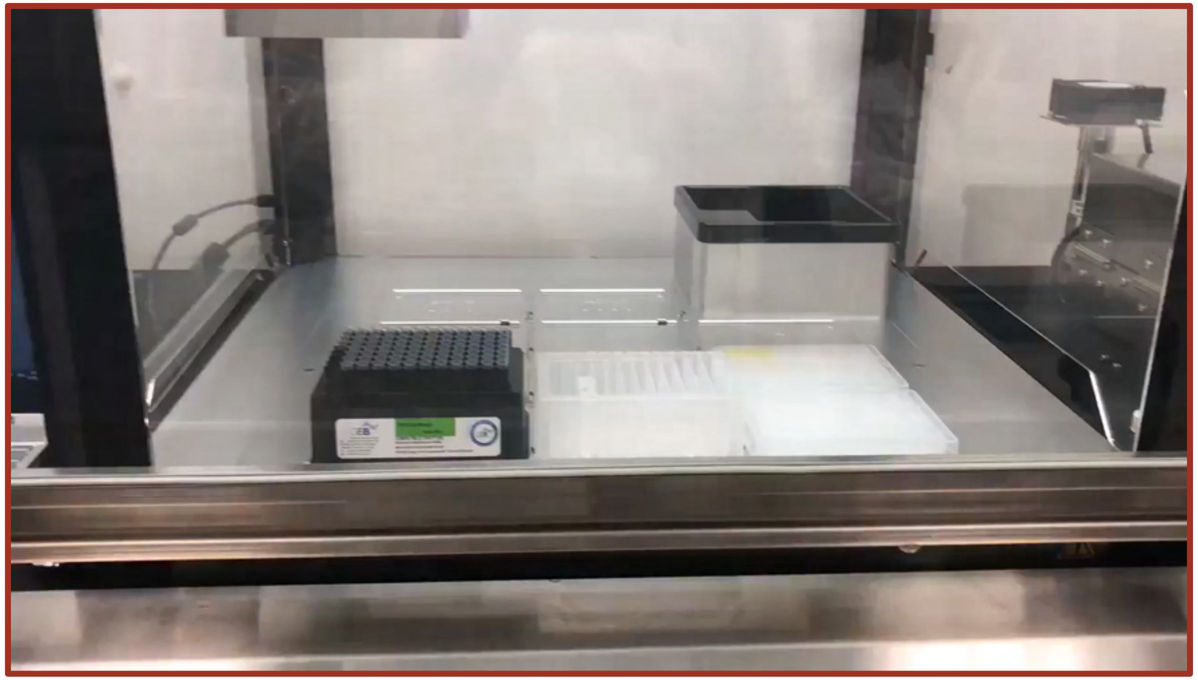


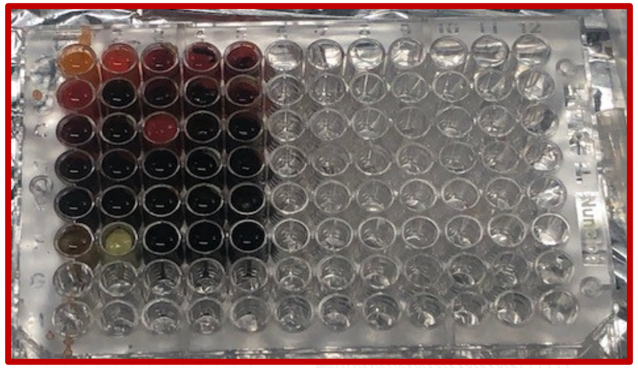
Automated experiment in PFM

- Non-rectangular scans
- Image-reconstruction based on non-rectangular scans
- Decision making

Why synthesis?







M. Ahmadi

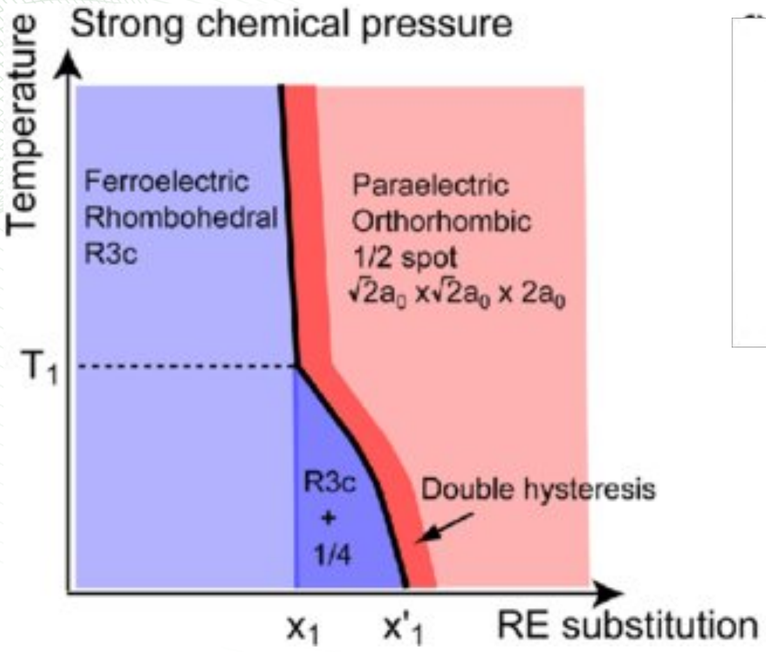
- Automated synthesis in its simplest form requires some way to navigate phase diagrams
- … in more complex form, processing space.

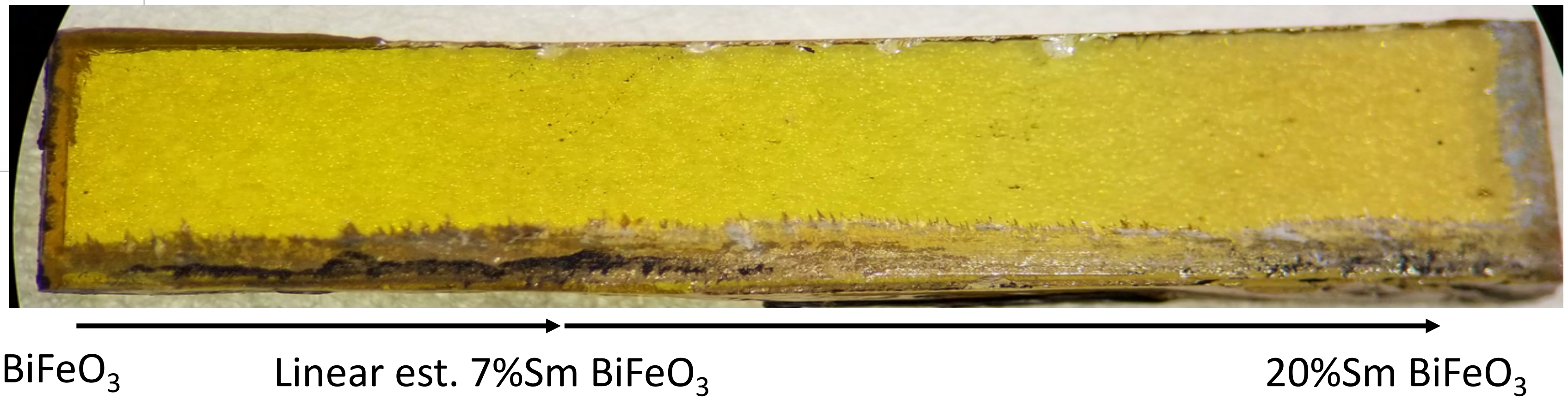
- Ideally, incorporate physical knowledge and modelling

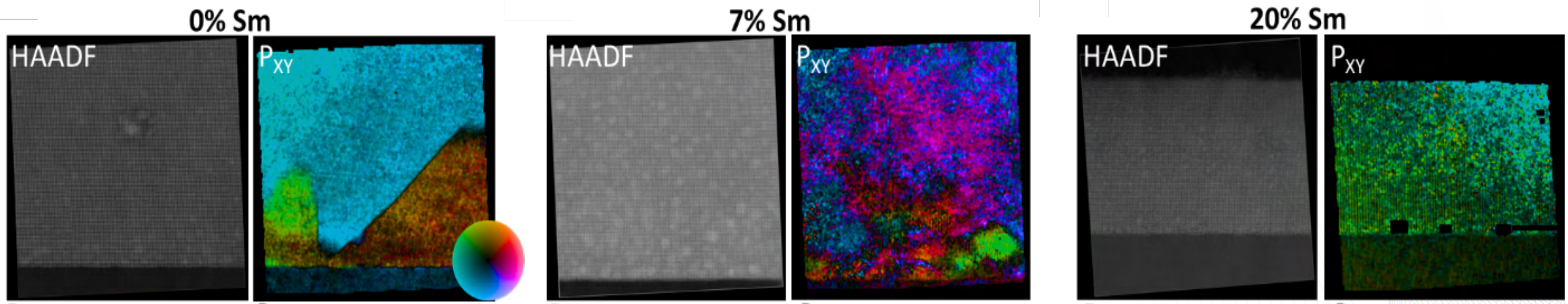
Oak Ridge National Laboratory

Combinatorial library

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







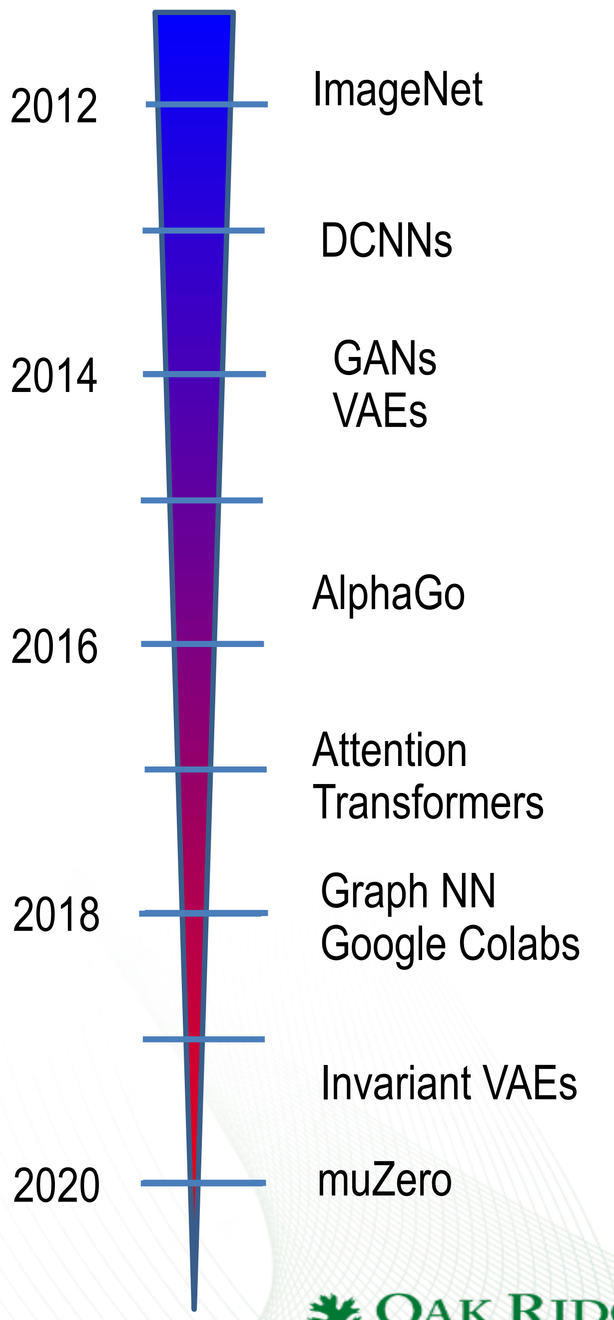
arXiv:2004.11817

Why Machine Learning (in Imaging)?

- Last decade has experienced an explosive growth of machine learning and artificial intelligence applications
- These developments have spanned areas from computer vision to medicine to autonomous systems and games
- However, the progress and impact as applied to experimental physical sciences has been minimal….

Why is it difficult?

- Requires domain expertise and domain-specific goals
- Deeply causal and hypothesis drive nature of domain sciences
- No single answer: culture, not a method
- Infrastructure, open code, open data



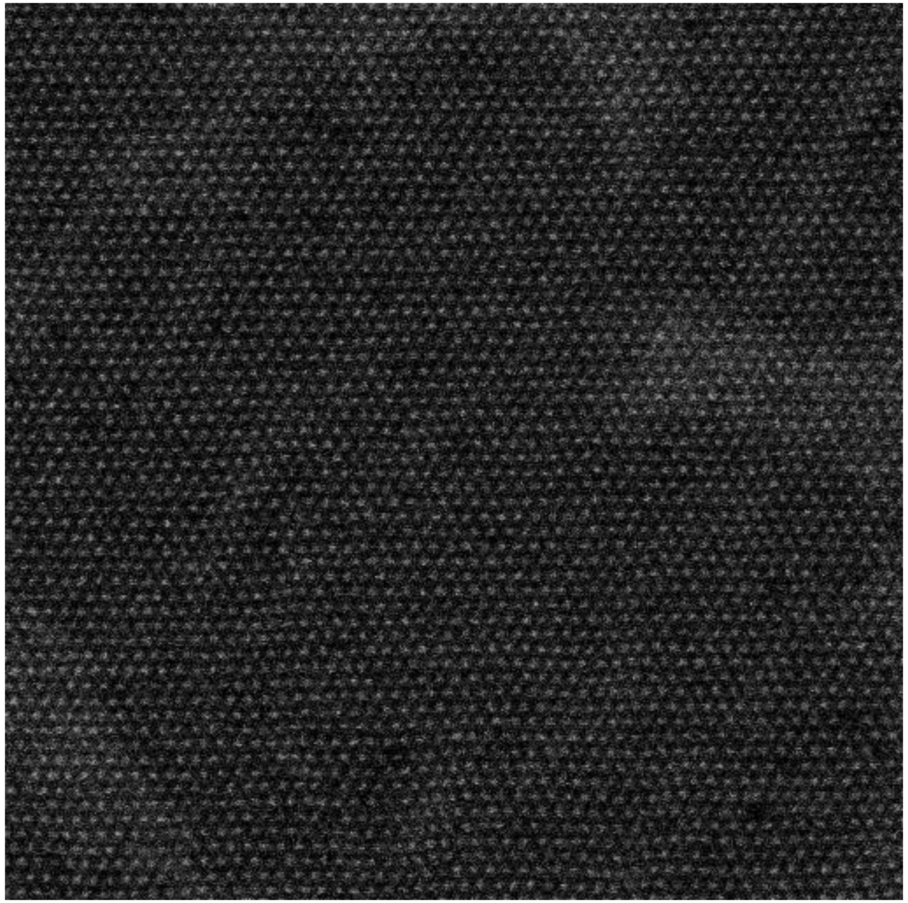
Supervised learning:

know what thy look for….

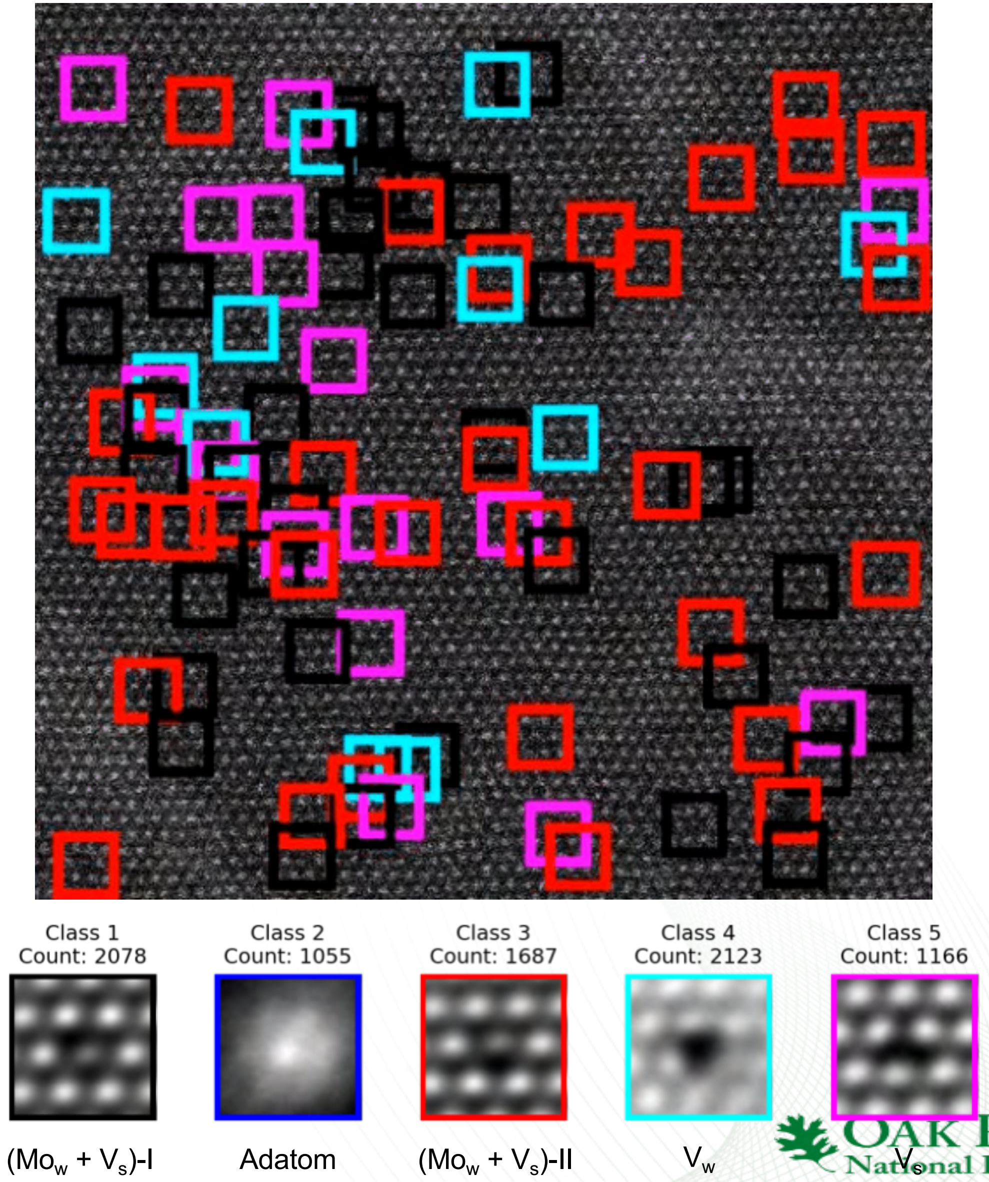
… and simplify it

Learning the defect evolution

Experimental



Decoded

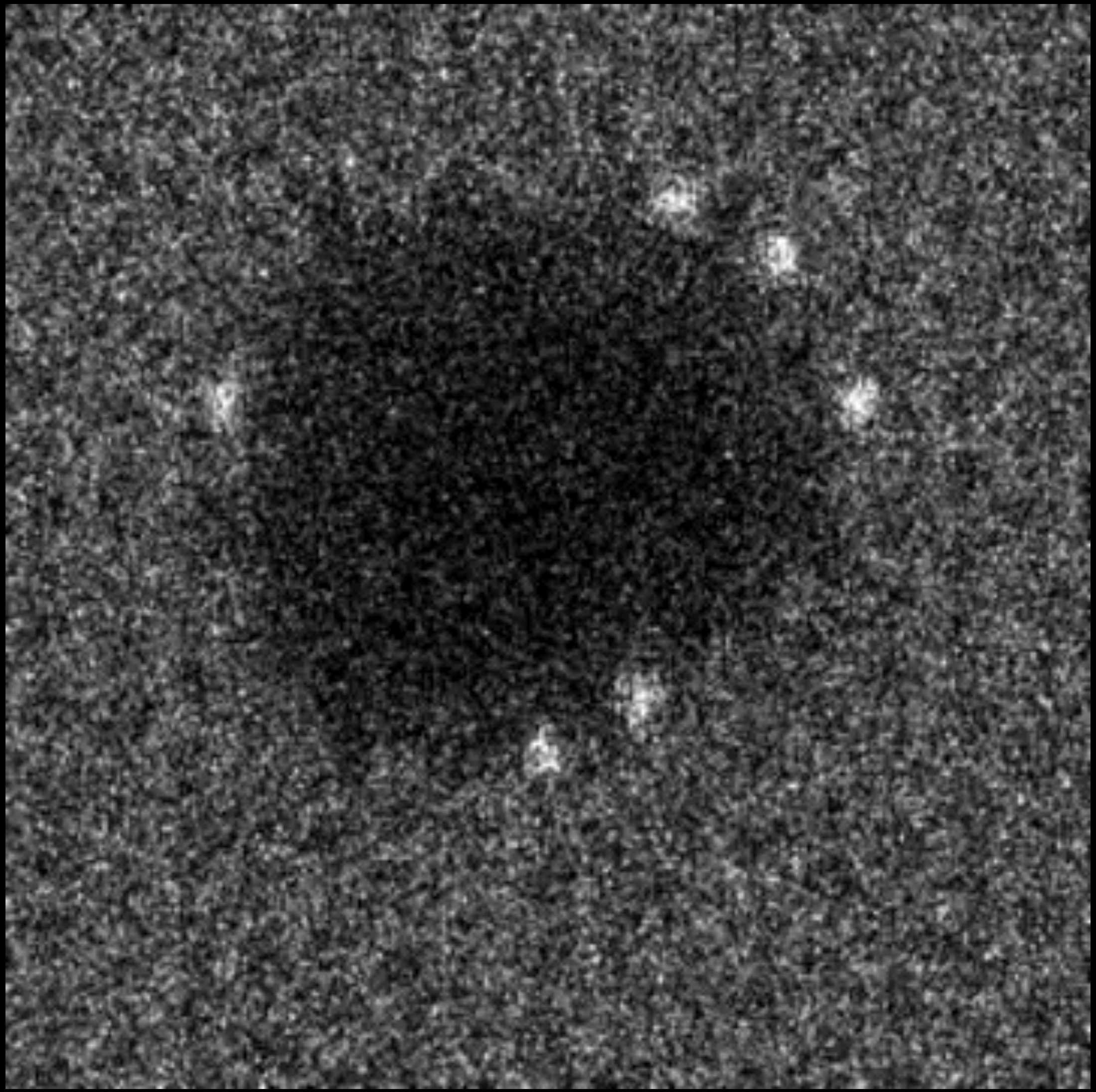


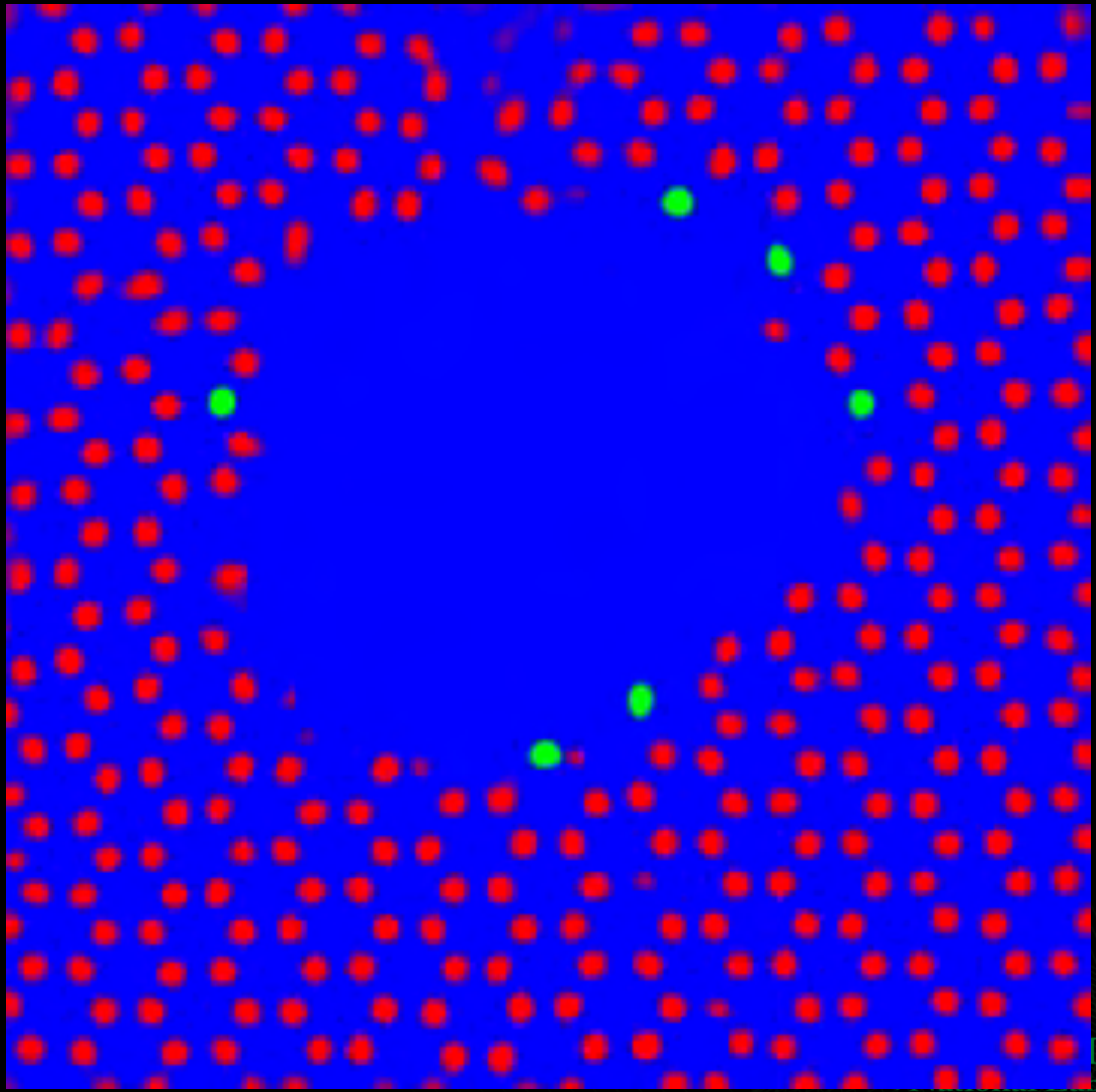
Sample: WS2
E-beam energy: 60 kV

Data collected by Ondrej Dyck (CNMS/ORNL)

Deep learning works like a charm for:

- Drift correction
- Denoising
- Data processing/dimensionality reduction
- Feature finding (physics is in the training set)





Autoencoders: variational, rotational, conditional

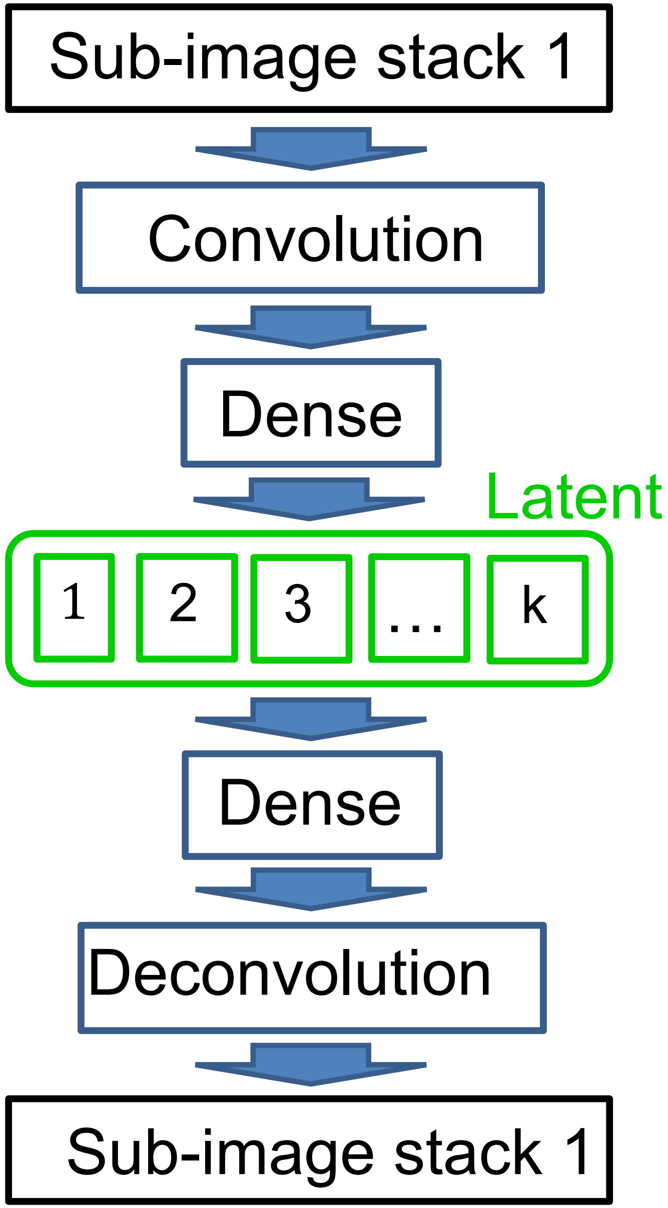
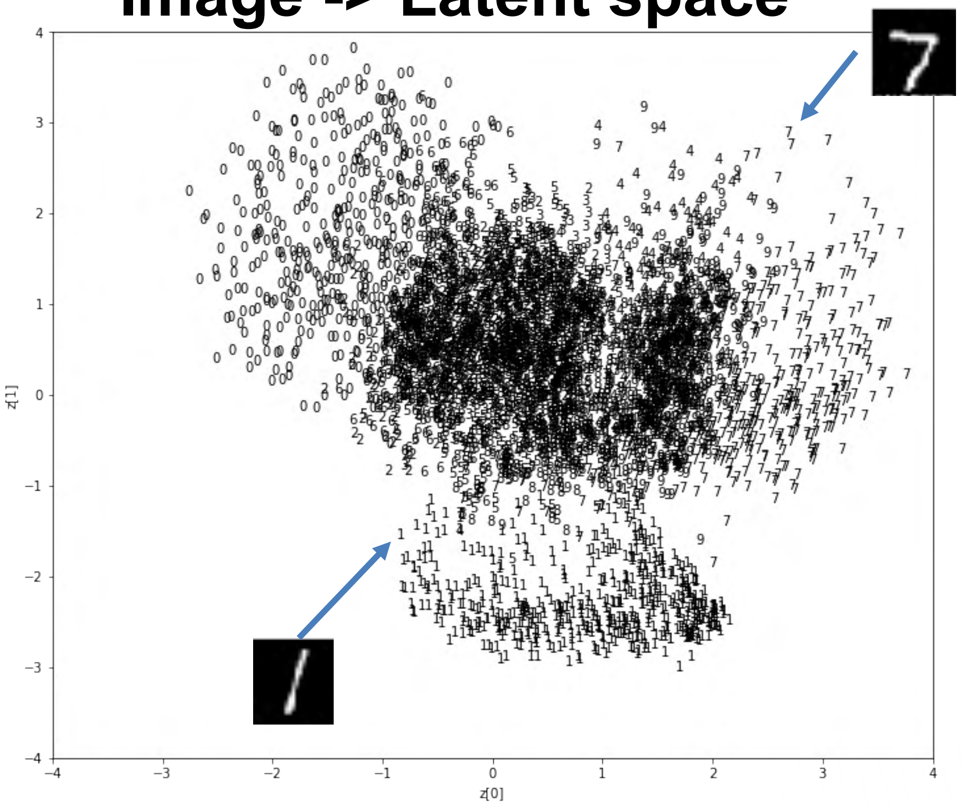
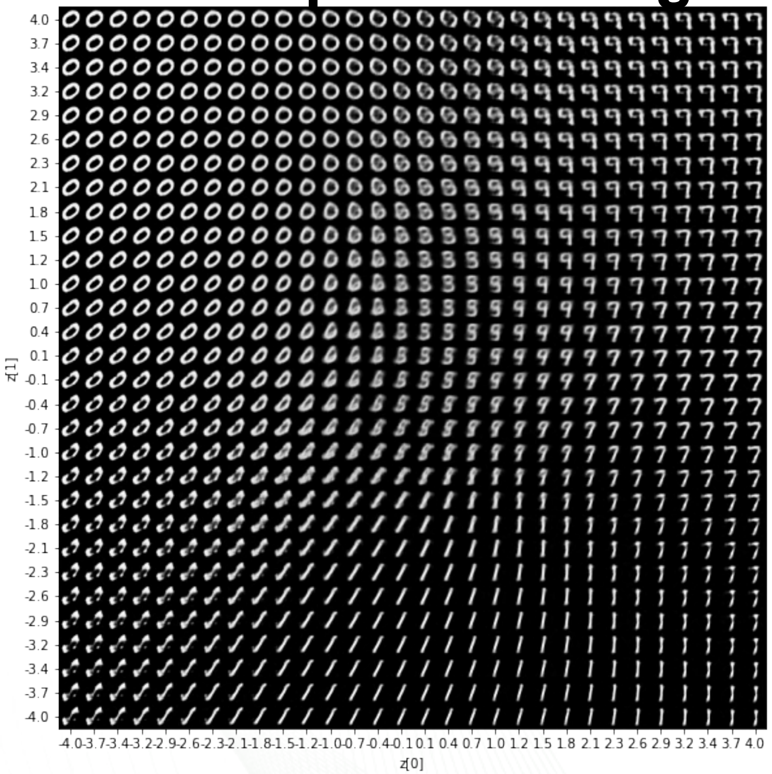


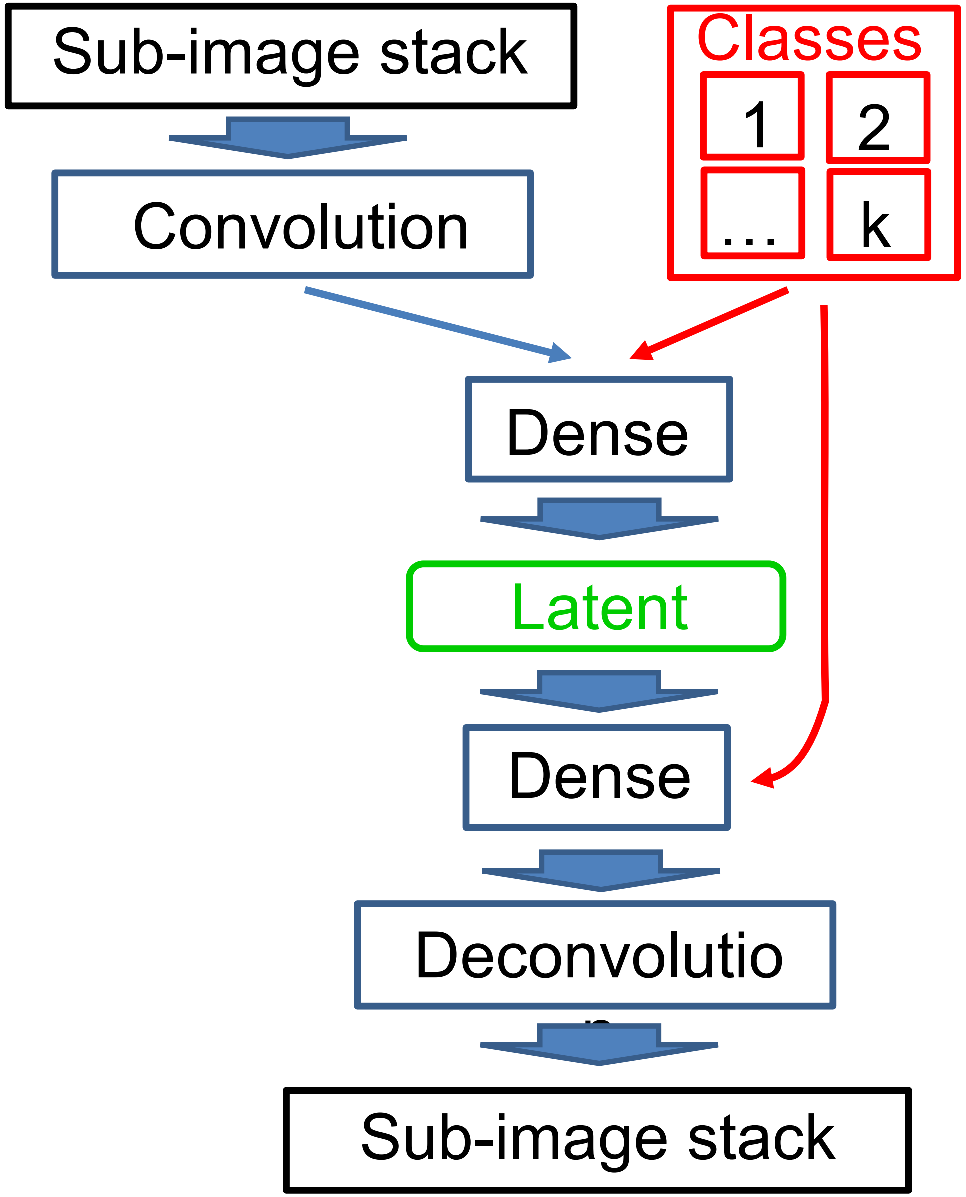
Image -> Latent space

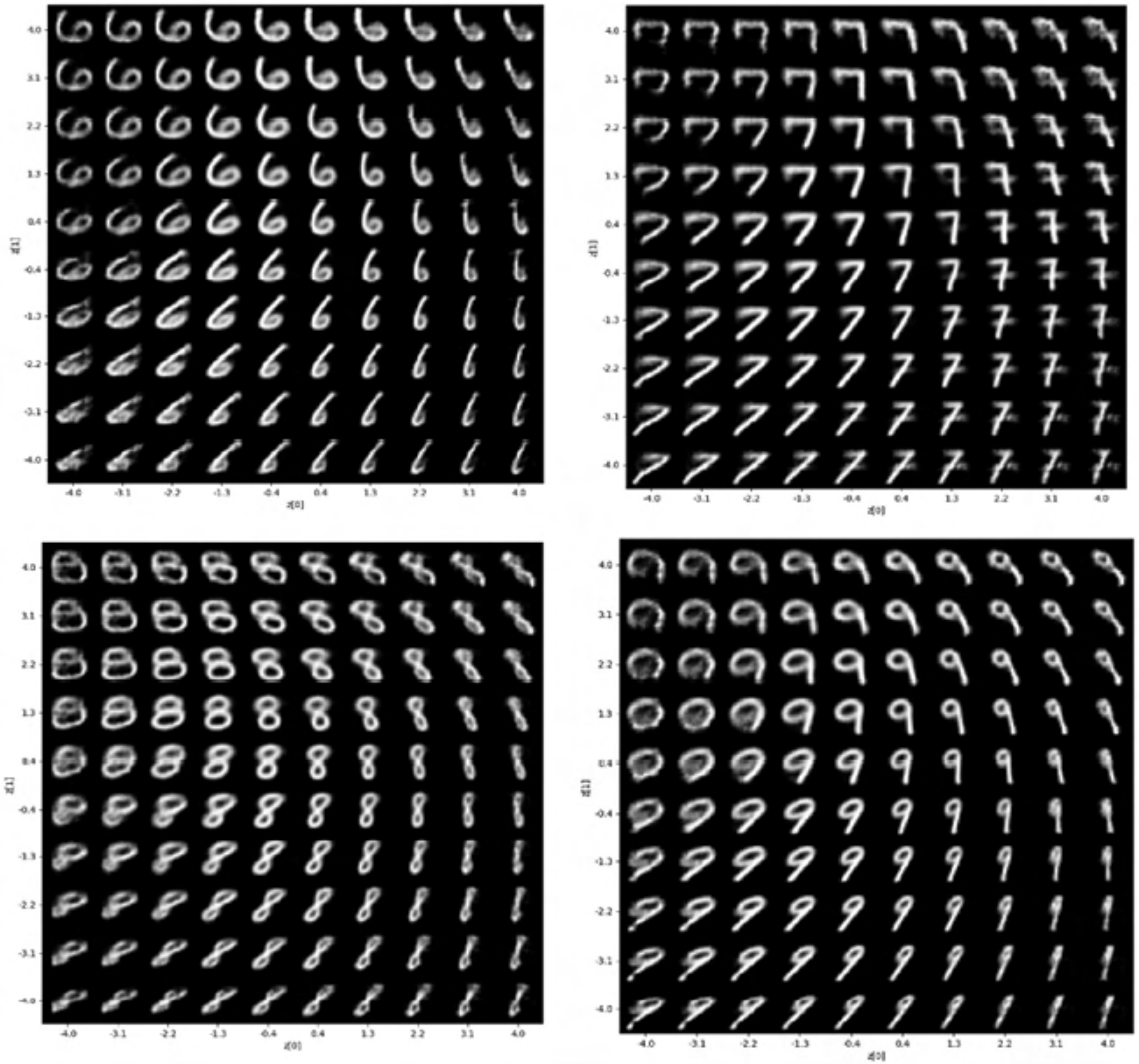


Latent space -> Image

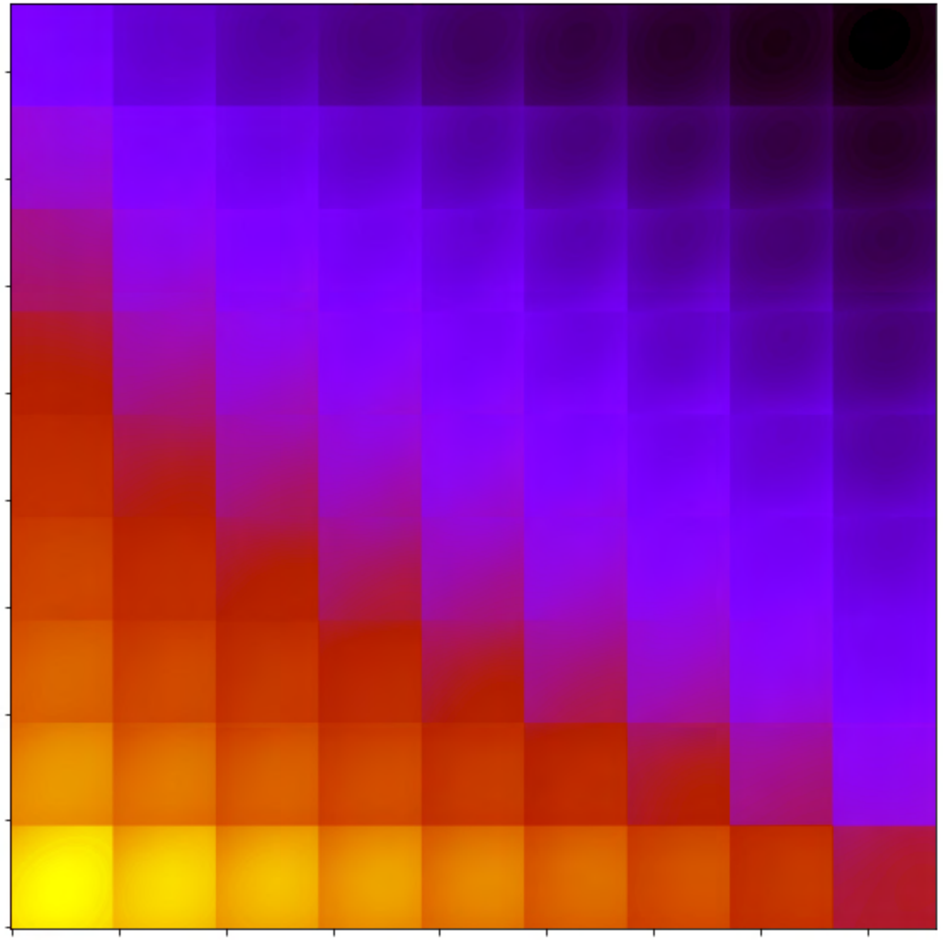


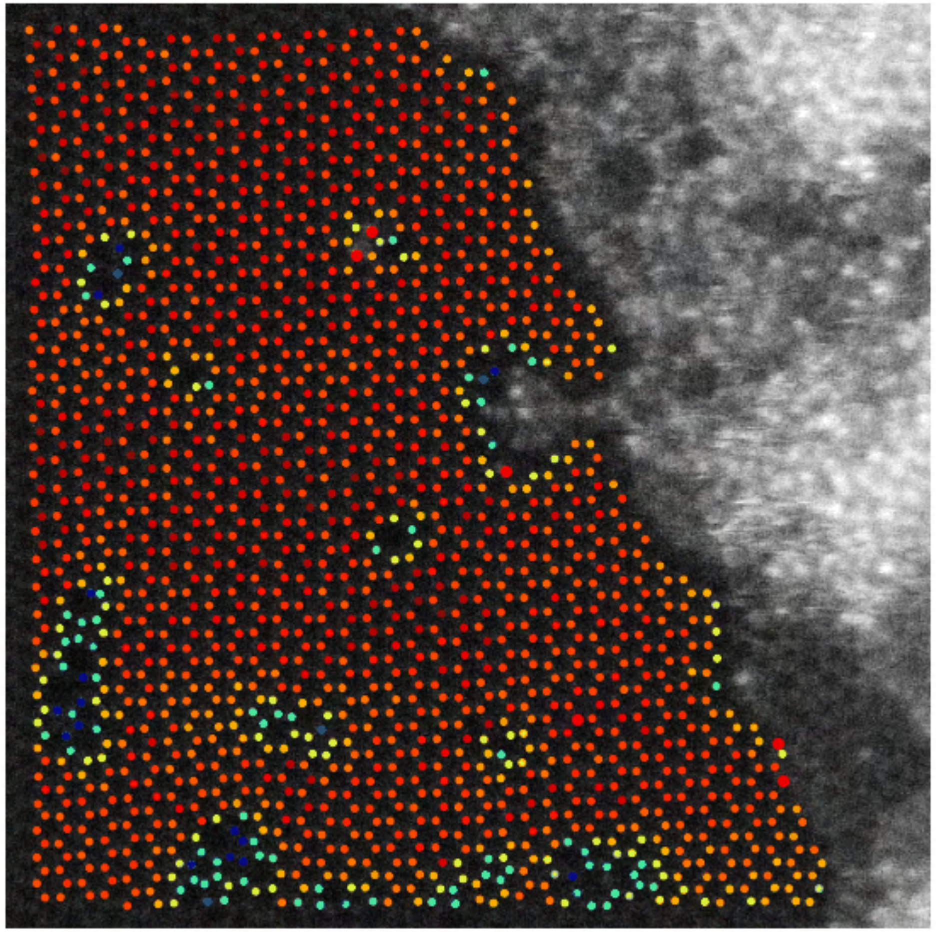
Conditional variational autoencoder



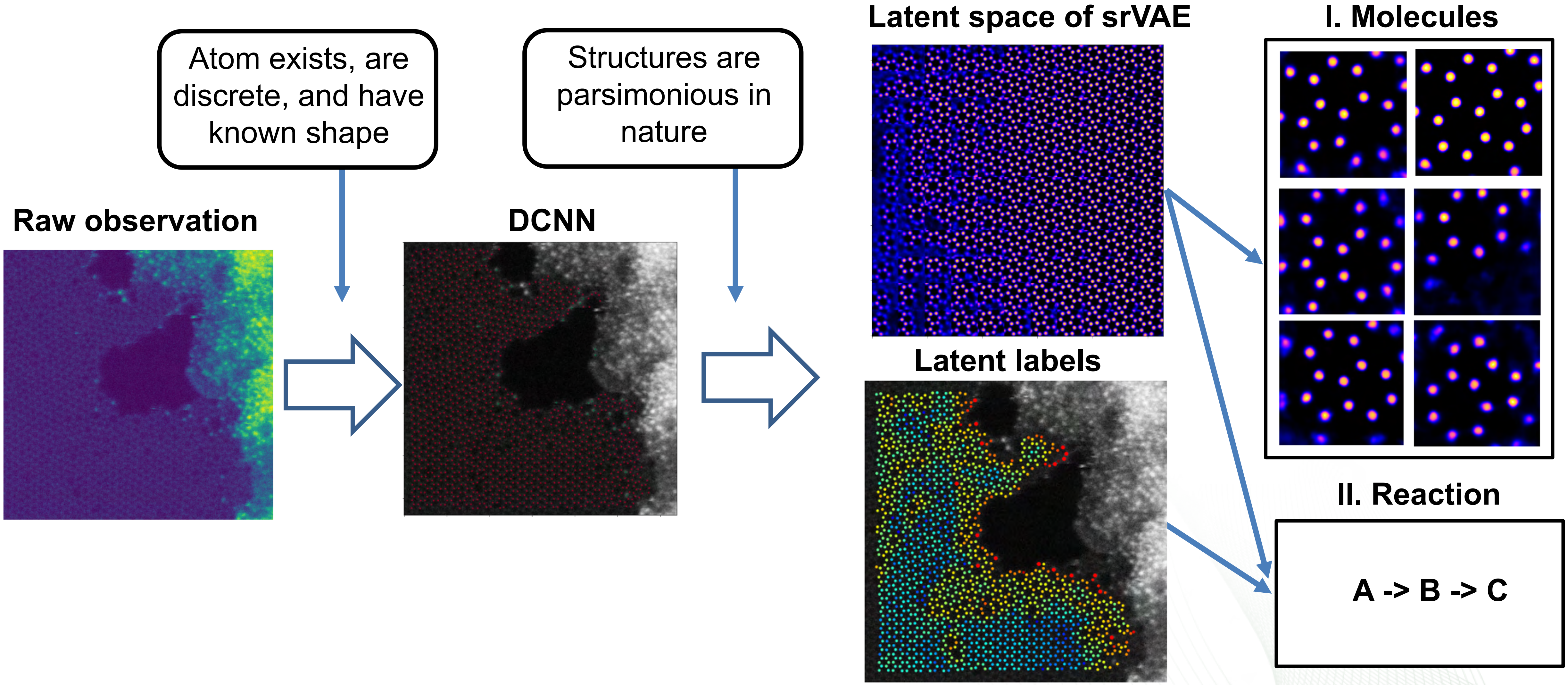


Rotational Variational Autoencoders

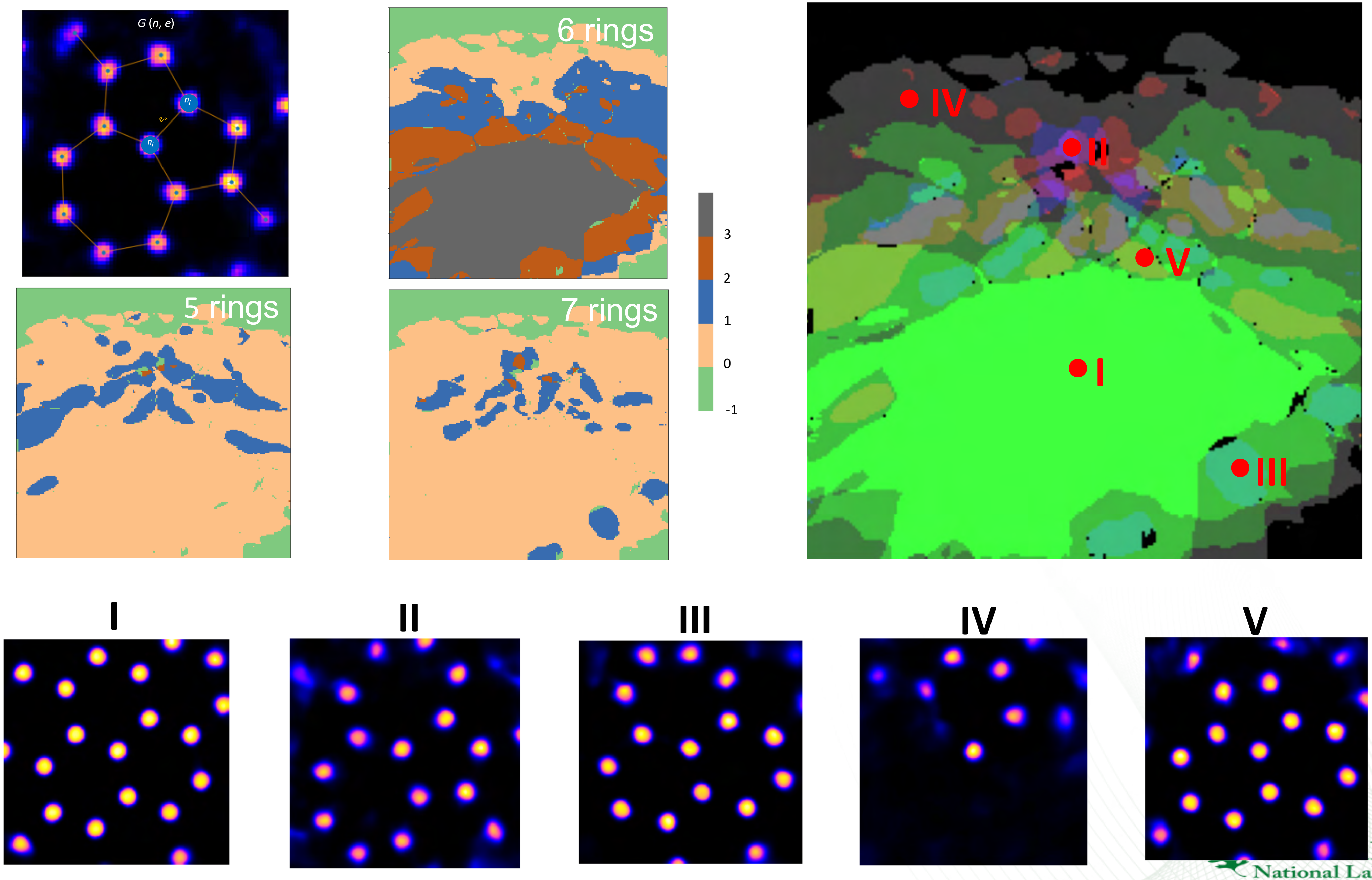




Discovering Molecular Fragments

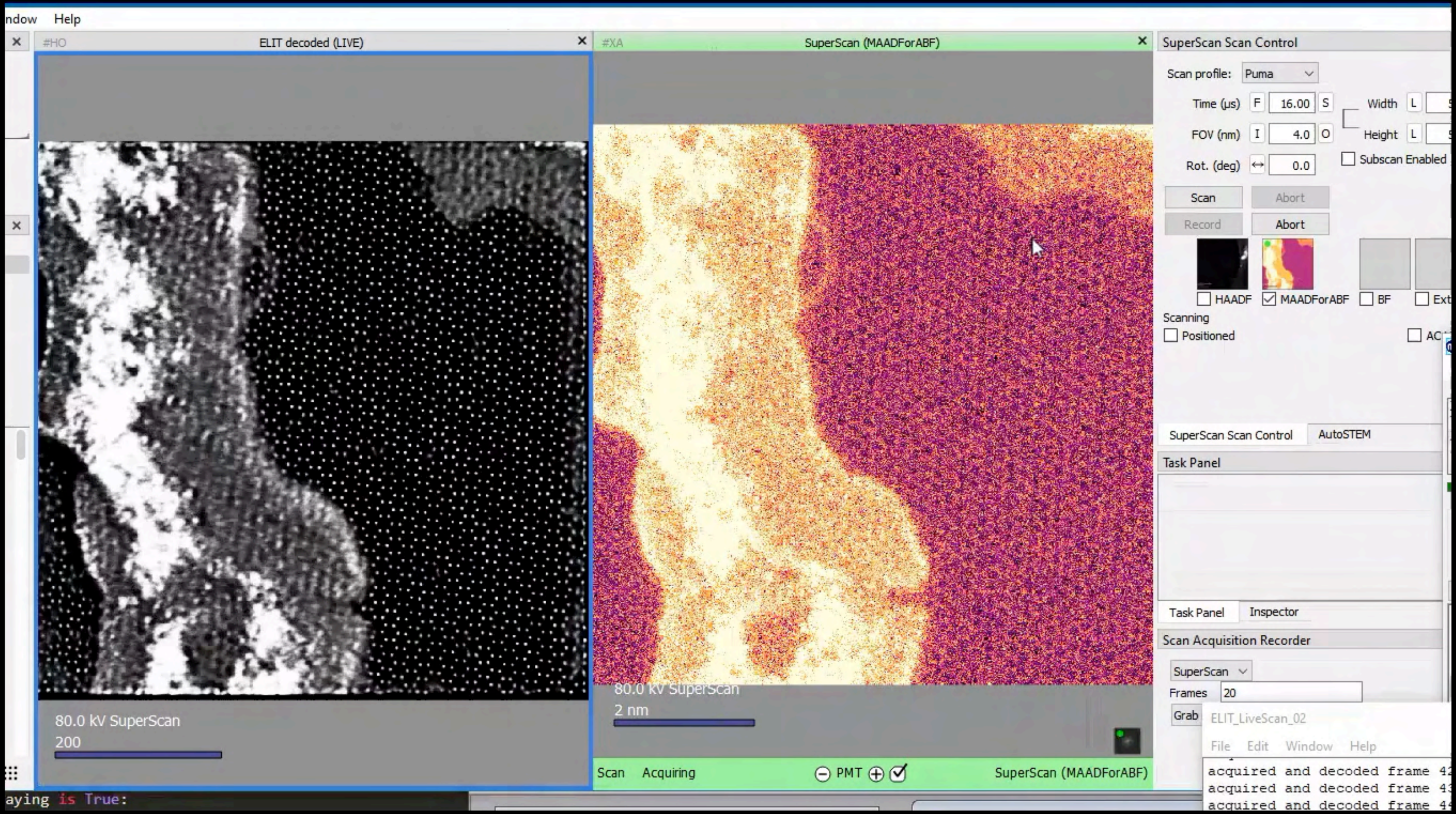


Discovering Molecular Fragments: Chemical Space Map



On-the-fly implementation

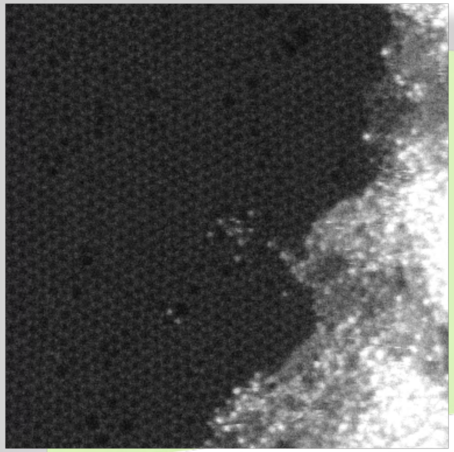
Implementation: Kevin Roccapriore, Ayana Ghosh, Sergei V. Kalinin & Maxim Ziatdinov



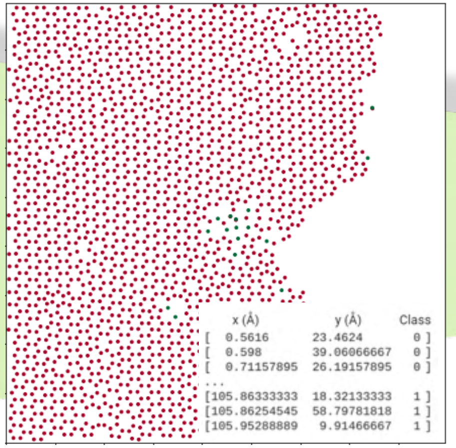
From Microscope to MD

Stage 1 | Stage 2 | Stage 3

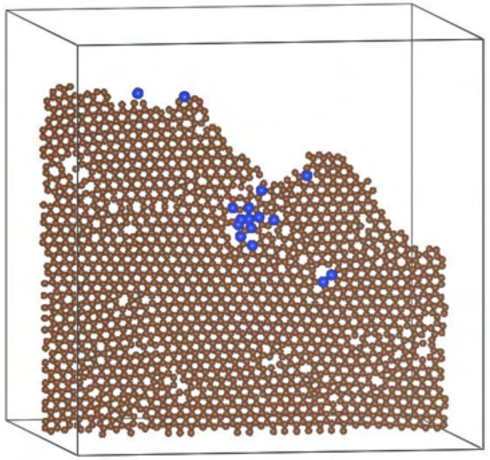
Microscopic images → **Atomic Types & Coordinates** → **Atomistic Simulations** → **Physical Phenomena**



DCNN



Transform

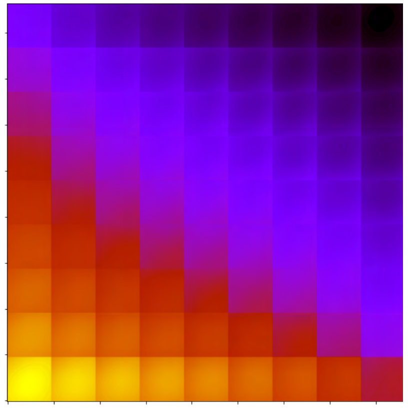


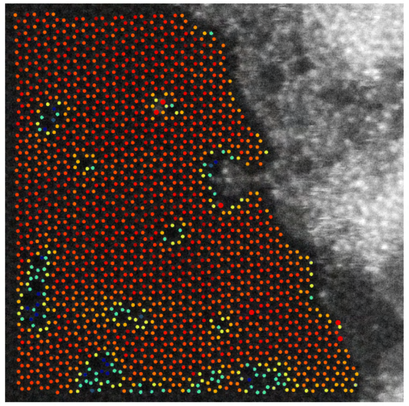
Materials Properties

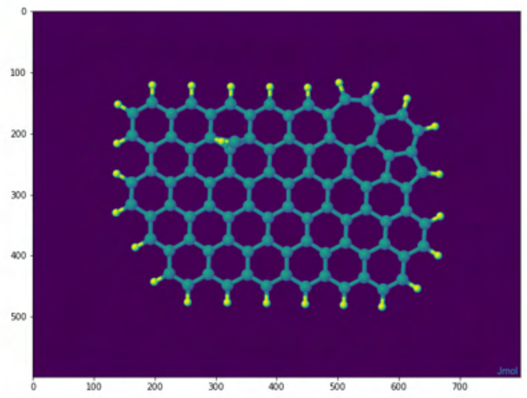
Feature Finding & Physics-informed learning

Data **Machine Learning** **Human-in-the-Loop** **Learn Physics**

- DCNNs for atom finding
- Invariant VAEs for physics discovery
- Conditional VAEs
- GP reconstructions
- Piping data in DFT/MD environment





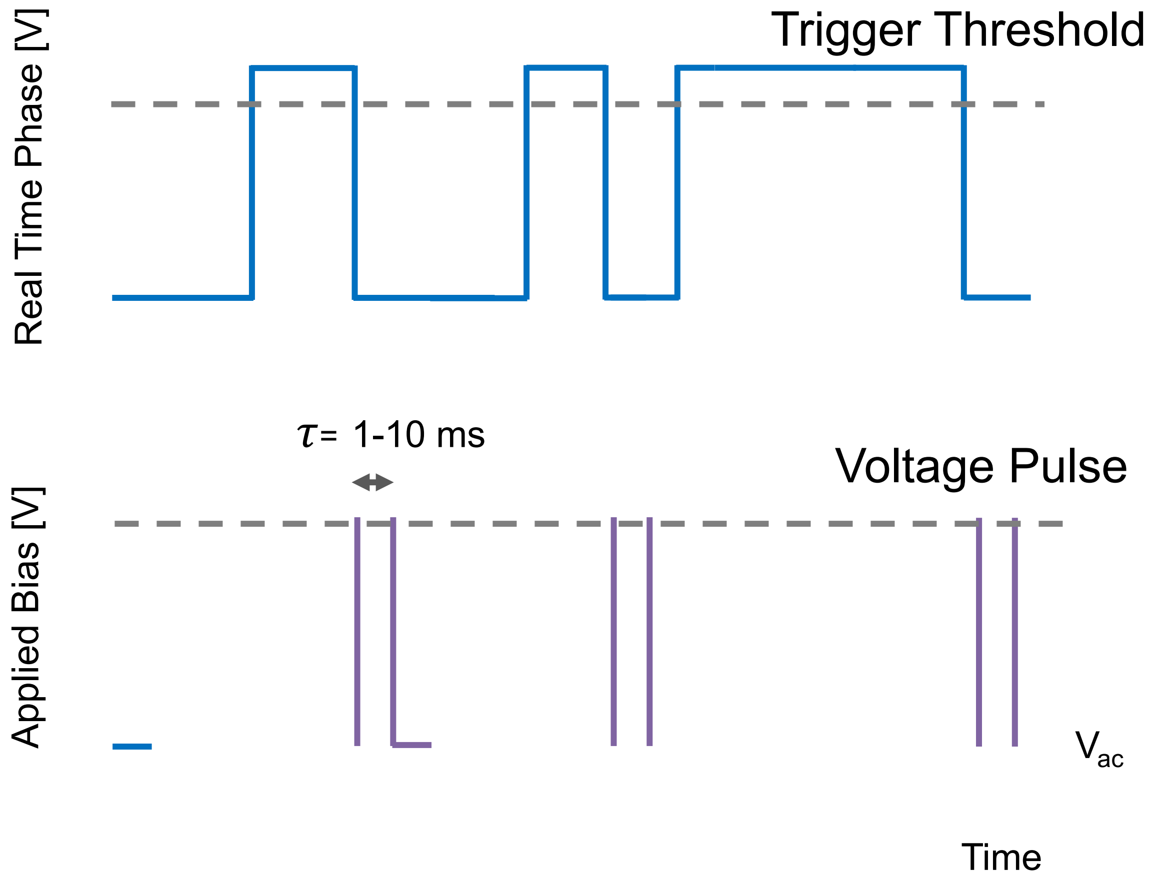


Automated Experiment:

almost easy…. If you know what you are looking for

FerroBOT: Image-based feedback

Real-time Feedback



Example PFM Phase

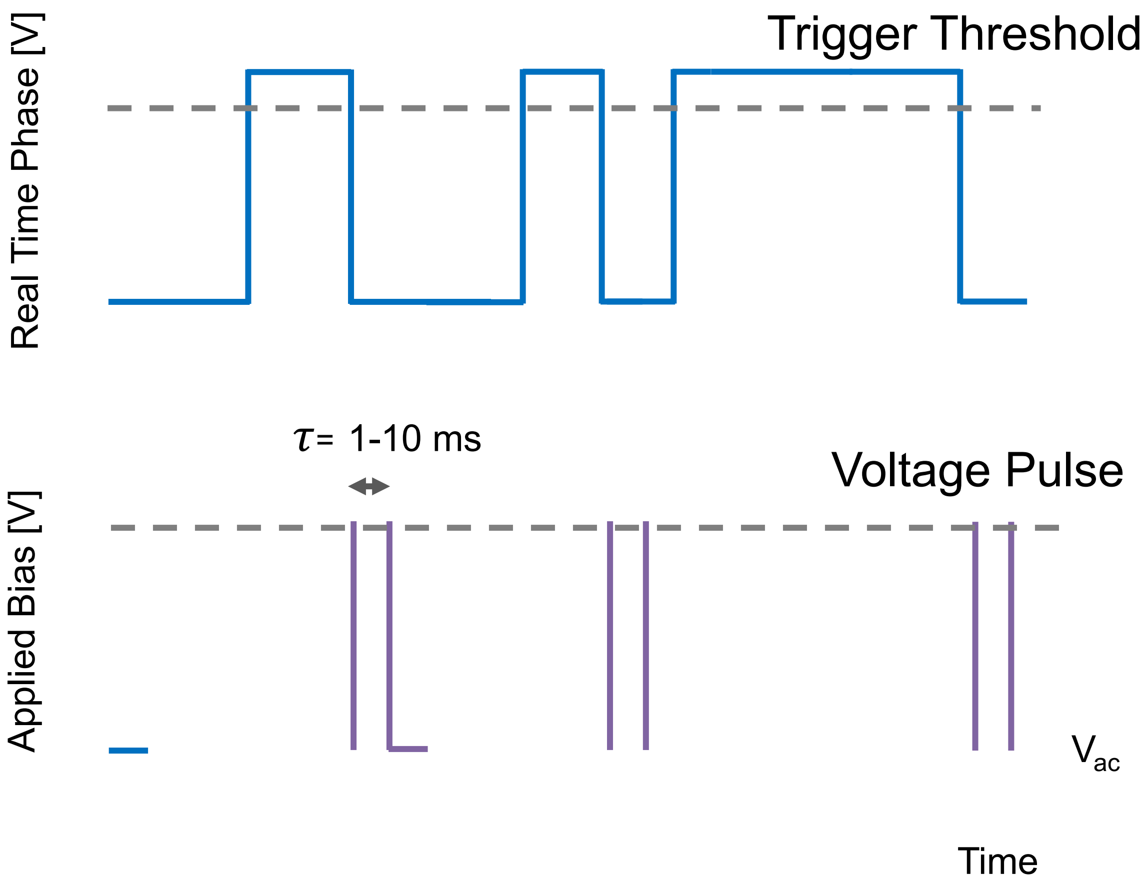
- FPGA – Labview – Matlab framework
  - Driving with AFM

Phase [deg.]

Reduction of ”purple” domain

FerroBOT: single action table

Real-time Feedback



Real-time Manipulation

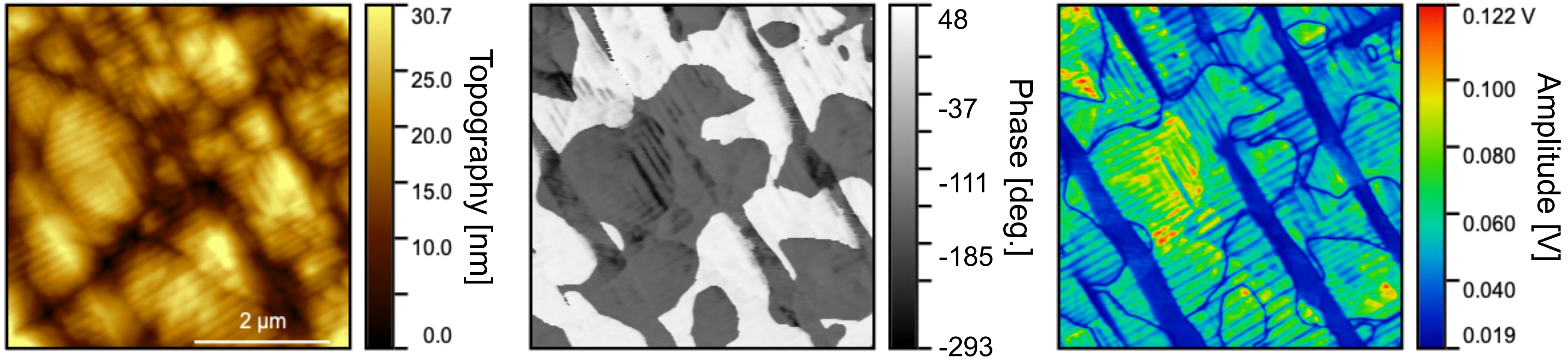
- FPGA – Labview – Matlab framework
  - Driving with AFM
- $BiFeO_3$: map DW energy landscape
  - Pulsing 5 ms, 1.5V, vacuum, domain wall growth

Amplitude Phase

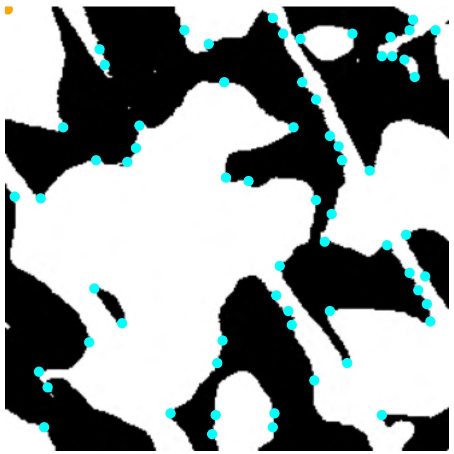
Realtime Feature Finding: $PbTiO_3$

- Cypher microscope
  - Ethernet connection for transmitting locations
- Skimage corner finding
  - Thresholded image feed
- Python – LabView framework
- Outlook
  - IV curves on domain walls
  - Ferroelastic domain wall probing

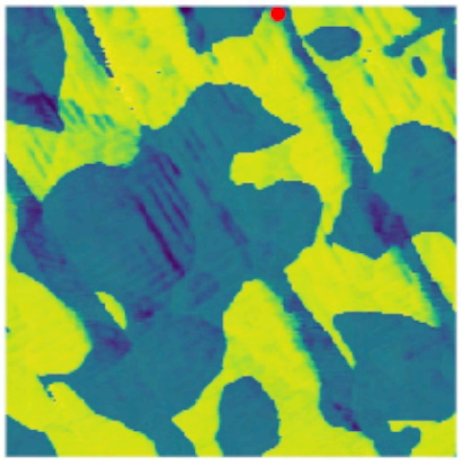
Initial State via Vertical Piezoresponse Force Microscopy



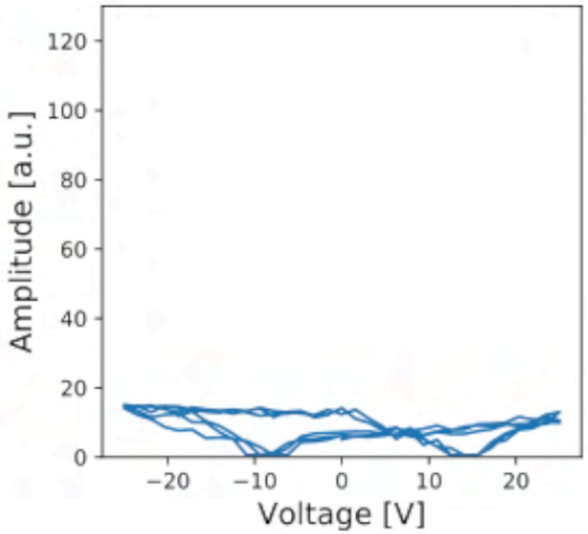
Feature Recognition



Tip Location



BEPS at Features

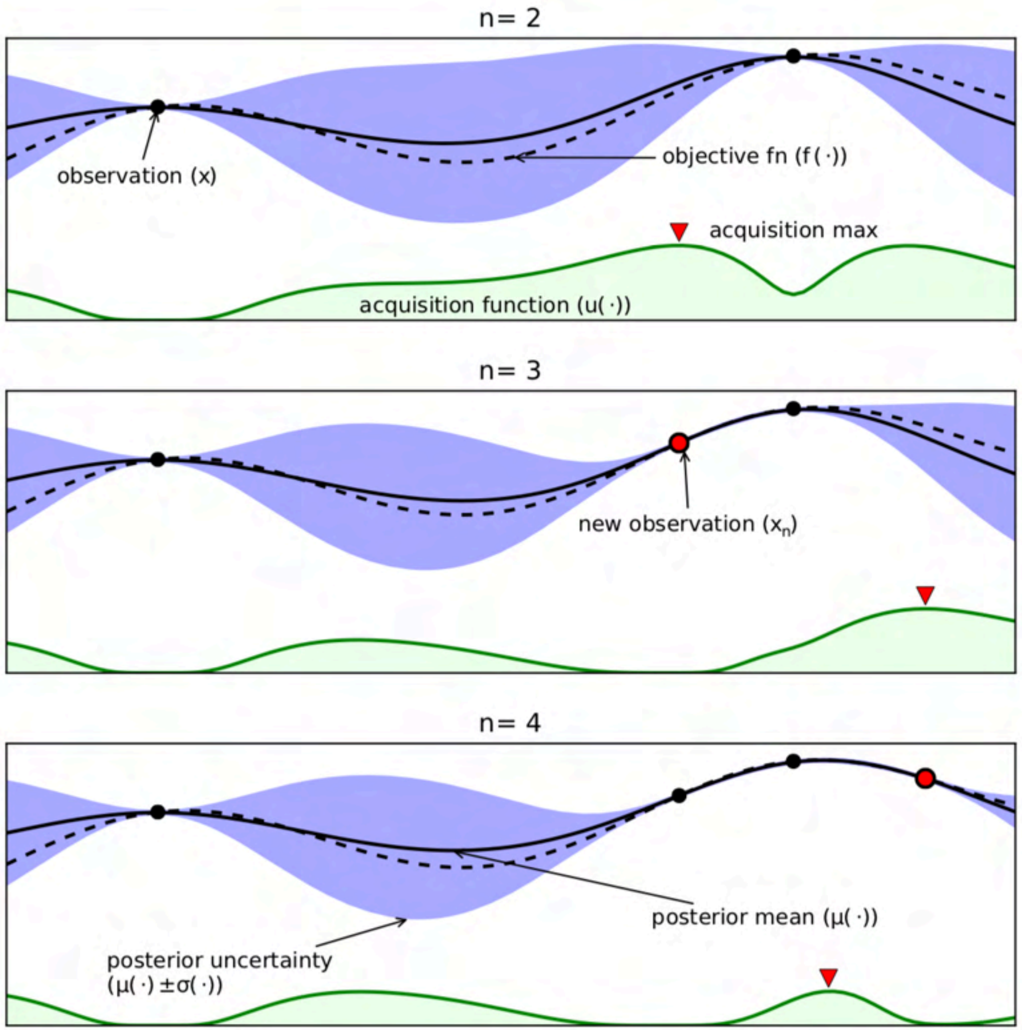


Here, we use simple computer vision algorithms to explore polarization switching in the regions predefined by operator based on prior knowledge

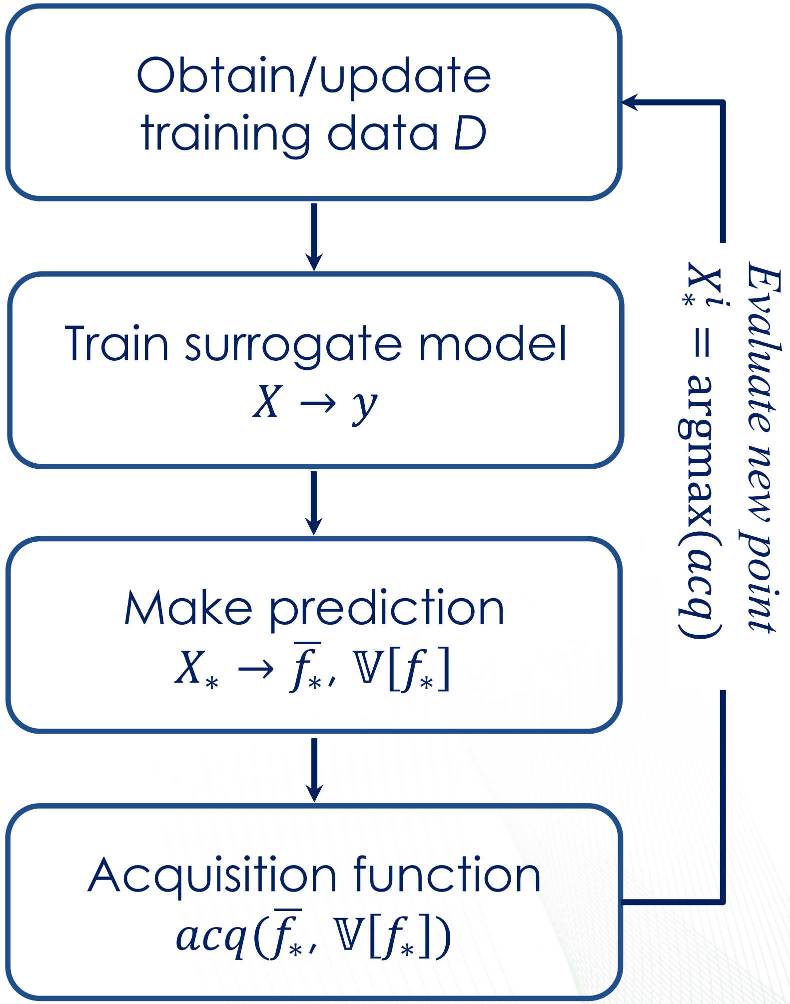
Automated Experiment:

… with John Snow priors…

The basics: Bayesian Optimization

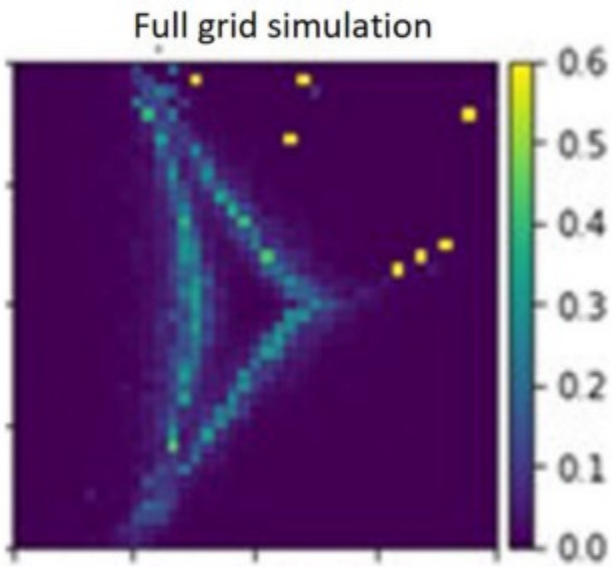


X, y: (sparse) Training data
X_*: New (not yet evaluated) points



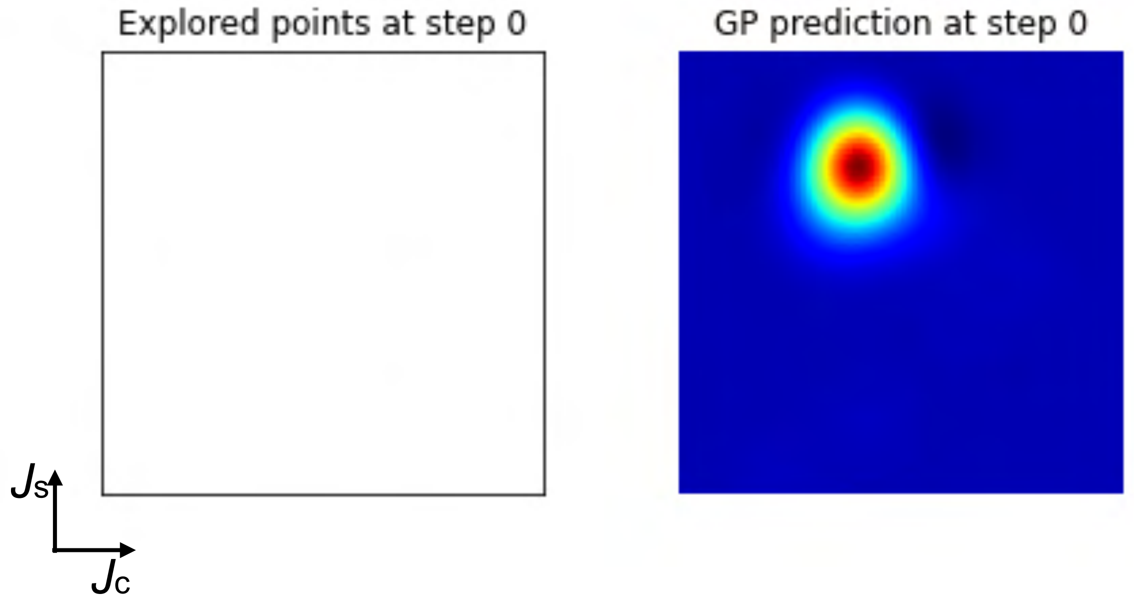
N. de Freitas et al., Taking the Human Out of the Loop: A Review of Bayesian Optimization, *Proceedings of the IEEE* **104**, 148 (2015)

Bayesian Optimization for physical discovery



- Started to work in August 2019
- First, we planned to apply BO for autonomous experimentation at CNMS. Then, COVID happened…
- “So what if we use BO to explore the parameter spaces of theoretical models?”

Discovering regions where heat capacity is maximized in NNN Ising model



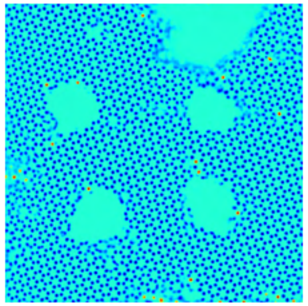
Implementation in GPim

```python
def acquisition_function(gpmodel, X_full, X_sparse):
    mean, sd = gpmodel.predict(X_full, verbose=0) # leave it as is
    acq = 5 * sd + 10 * mean # Good for heat capacity
    return acq, (mean, sd) # leave it as is
```

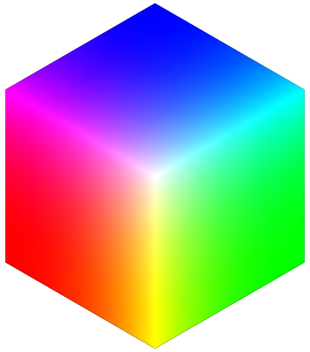
```python
# Initialize Bayesian optimizer with a custom acquisition function
boptim = gpim.boptimizer(
    X_sparse, Z_sparse, X_full,
    J2_to_S_func, acquisition_function=acquisition_function,
    batch_size=1000, lengthscale=[1., 40.], dscale=4,  exit_strategy=1,
    exploration_steps=650, use_gpu=True, verbose=1, save_checkpoints=True,
    filename='/content/drive/My Drive/research/Ising_BO/bo_ising_heat_acq1_e')
# Run Bayesian optimization
boptim.run()
```

Automated experiment workflows

SPM or STEM image



EELS or SPM datacube



- **Sliding window/linear transform**
- **Keras DCNN**
- rVAE (rotational invariance)
- rcVAE (plus classification)

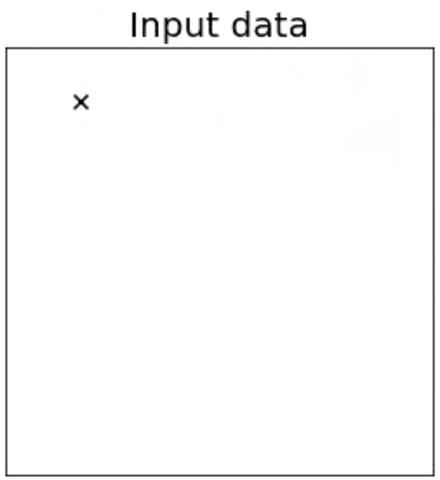
Descriptor

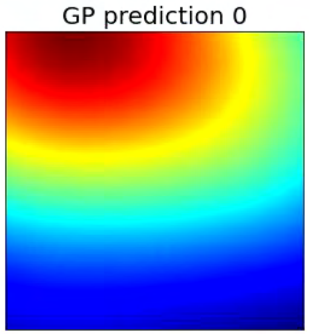
Gaussian processing

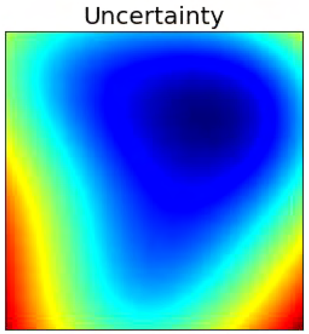
- **Integrated intensity**
- **Keras DCNN**
- Spec2im autoencoder
- (im,spec)2(spec,im)
- CycleGAN

- **Acquisition functions**
- **Pathfinder functions**
- **Kernel control**

GPim library (M. Ziatdinov)



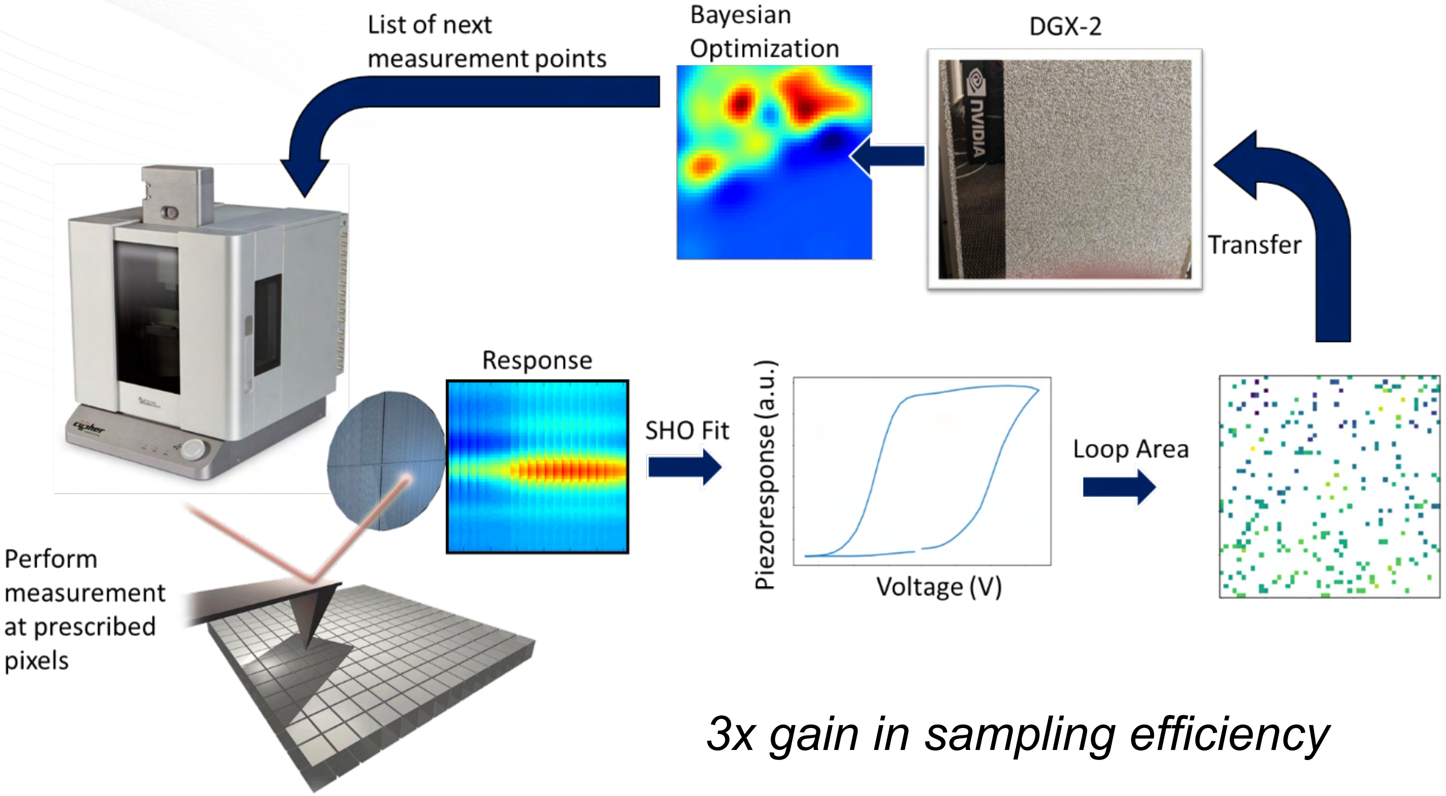




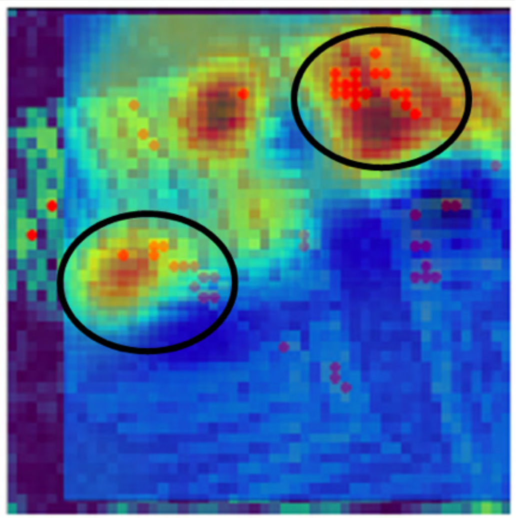
- **AE based on structural analysis for STEM data**
- **AE based on spectral data in PFM**
- **AE based on DL for EELS data**
- Feature of interest finding for mesoscopic images

Bayesian Optimization for Self-Driving Microscopy

Then, COVID restrictions got relaxed and Rama Vasudevan realized a “self-driving” PFM



Comparison with “ground truth”

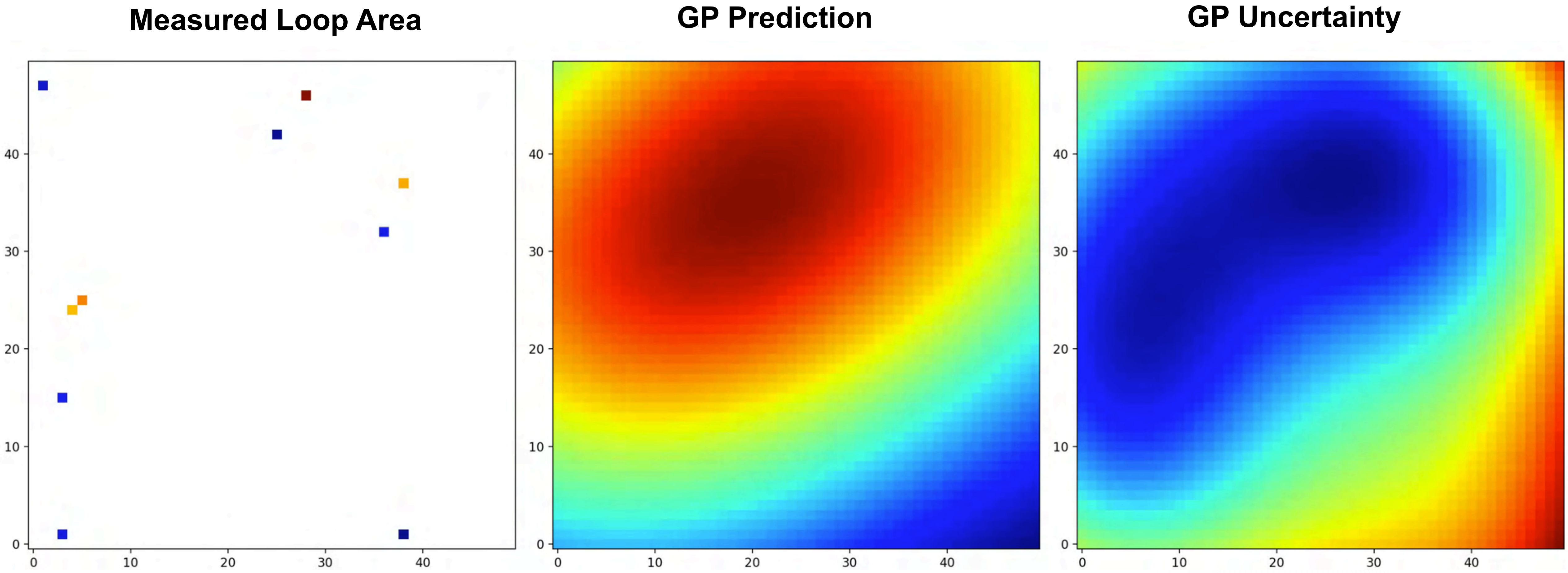


3x gain in sampling efficiency

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

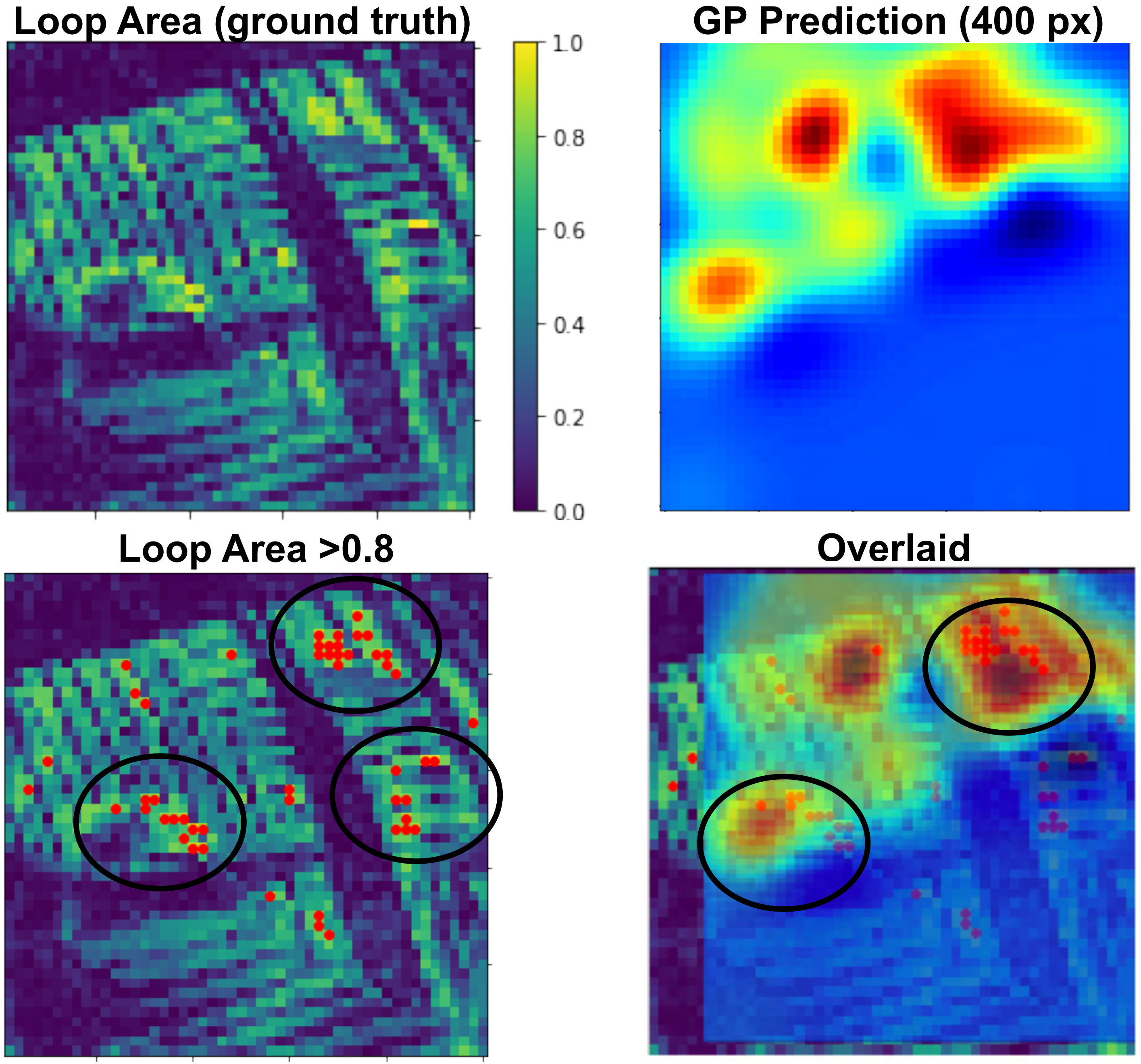
OAK RIDGE National Laboratory

Putting it together: GP optimized experiments



But what if we do not know a priori what elements of domain structure are we interested in?

- First step – Gaussian Processing towards exploration of specific behavior
- Here, we explore regions with maximal area under hysteresis loop
- For *N* measured points, the GP reconstructs the loop area map and uncertainties of the reconstruction
- Based on these, next locations for measurements are selected



arXiv:2103.12165

arXiv:2011.13050

Next steps:

- Incorporate prior knowledge of domain structure
- Factor in generative physics of ferroelectric domain structures
- These are complex ML problems

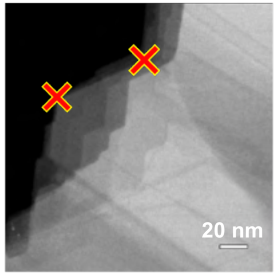
But: the bridge is built!

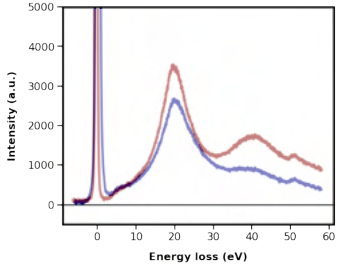
Automated Experiment:

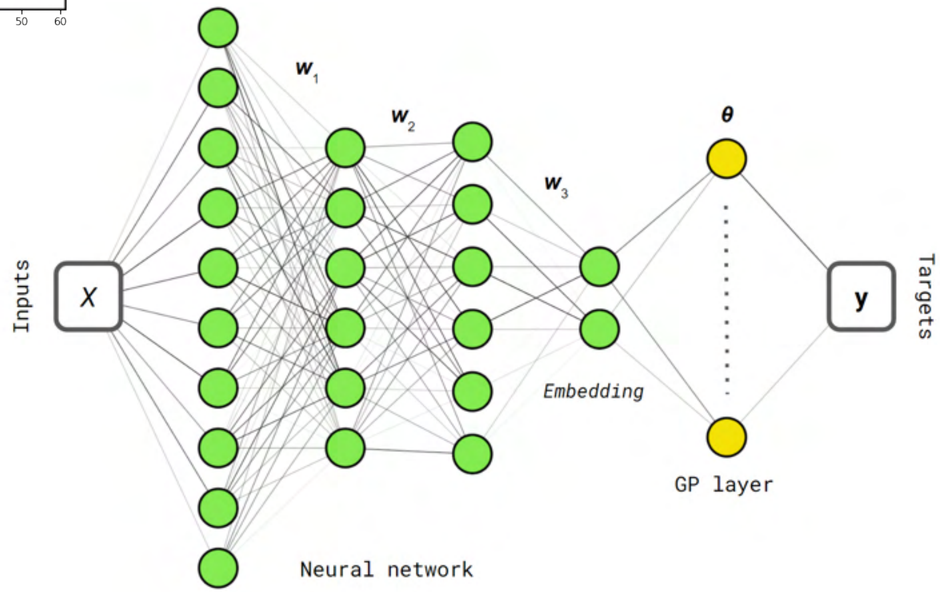
… as a scientist…

Physics-based feature engineering:
Deep kernel learning – Bayesian optimization

Active learning







Specify physics criteria

Acquire **structural data**

Measure a **spectrum**

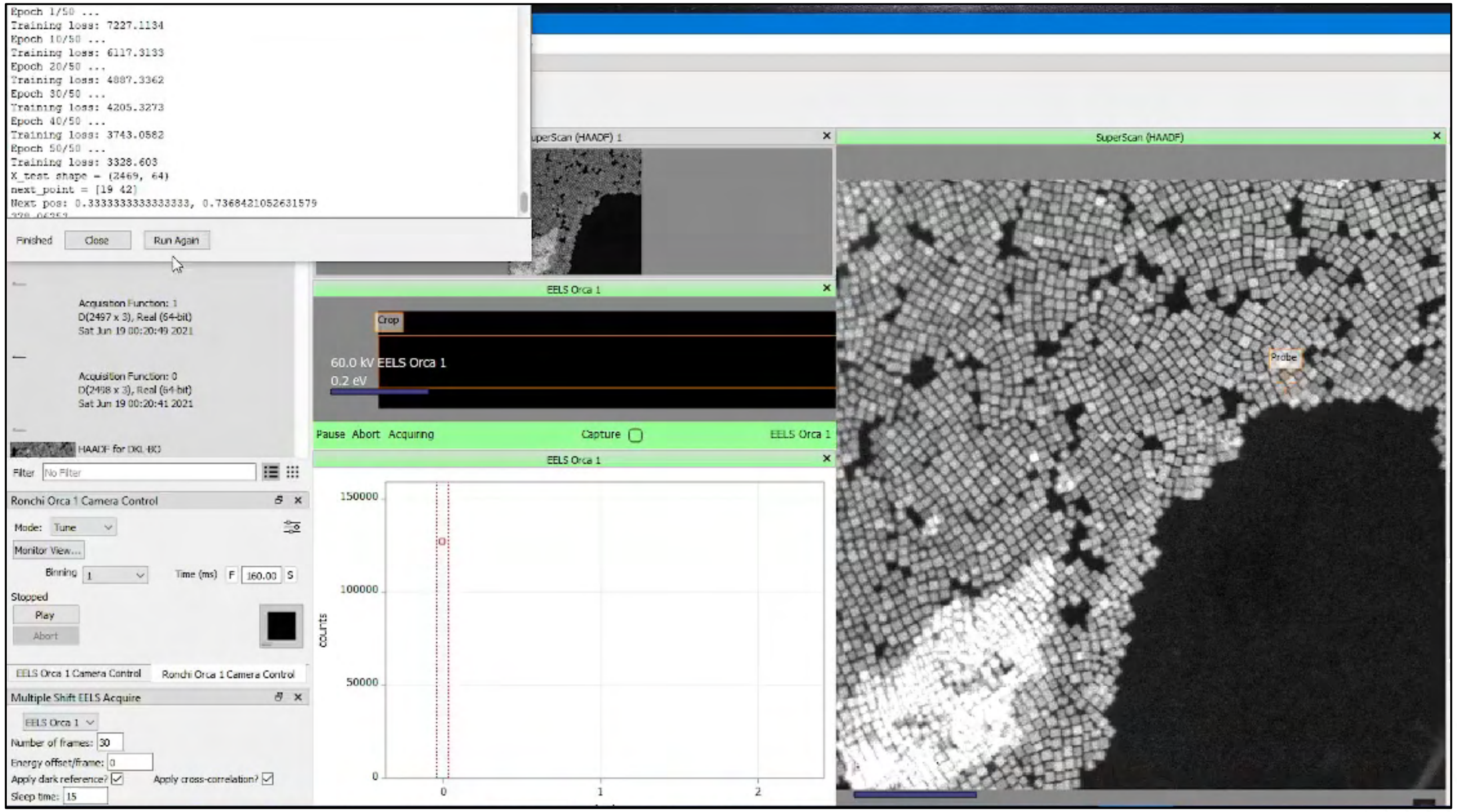
Train DKL model with new data

Decide next position (optimize physics criteria)

Allows navigation of the system to search for physics

Physics-based feature engineering:
Microscope Operation

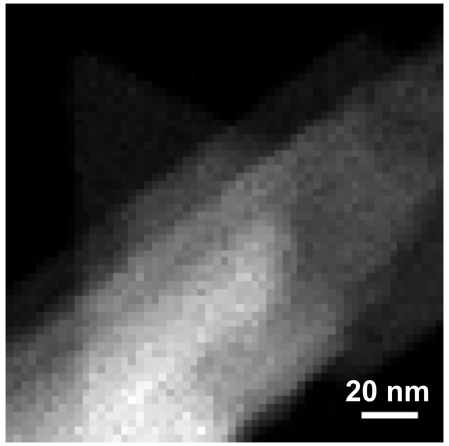
Structural image + Set of spectra



Opportunity alert

- Between measurements, the beam is optionally blanked or placed in a safe position.
- This is an **excellent** opportunity for "smart EELS" with beam sensitive materials

Physics-based feature engineering: MnPS3

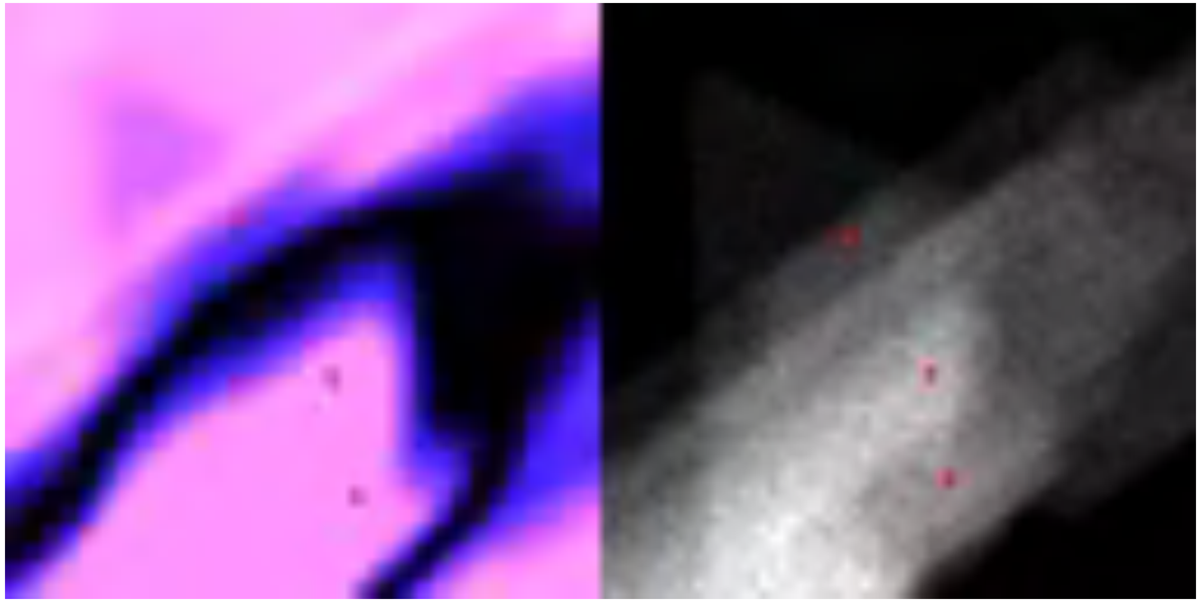


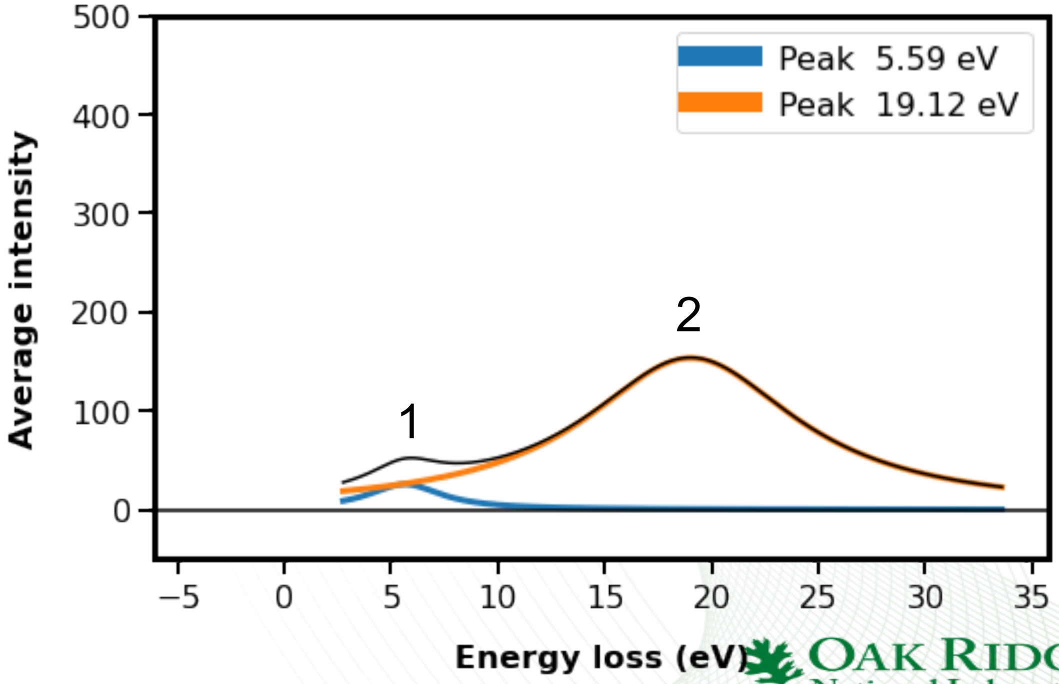
- **Discovering physics in a “new” material $MnPS_3$**
- **Curve fitting** to help enforce physical processes

Physics search criteria:

$$\boldsymbol{Ratio = Peak\,1\,/\,peak\,2}$$

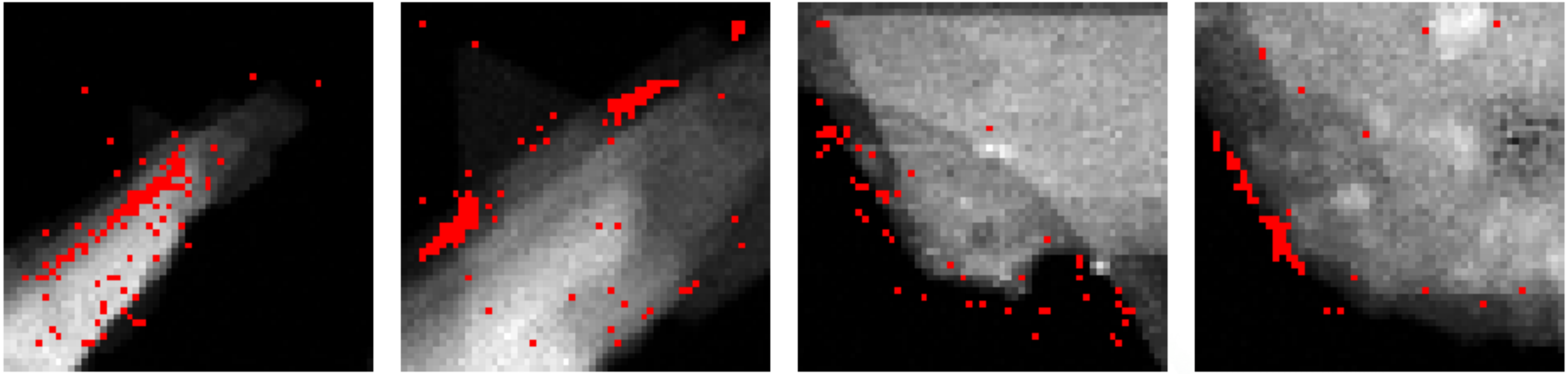
“Acquisition function” HAADF-STEM



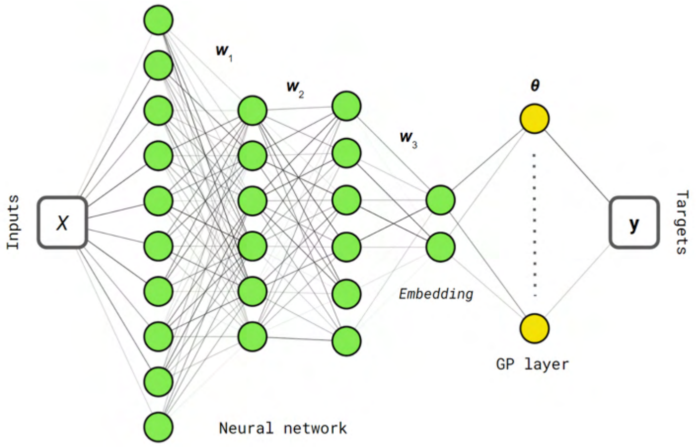


More examples of physics discovery

- Very similar behavior when searching for the same criteria!
- Success!



Automated Experiments in 4D STEM

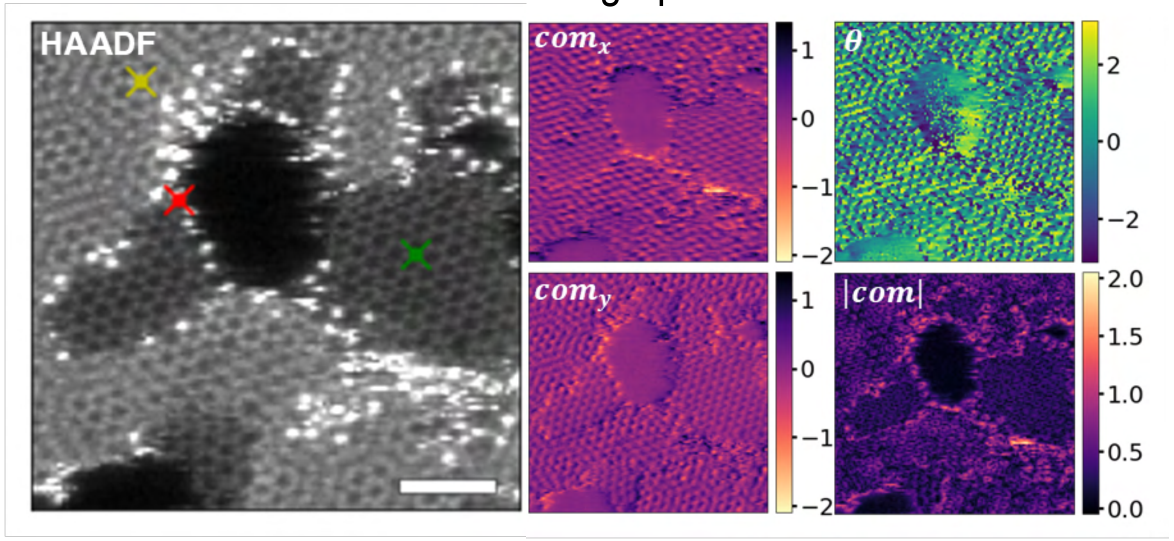


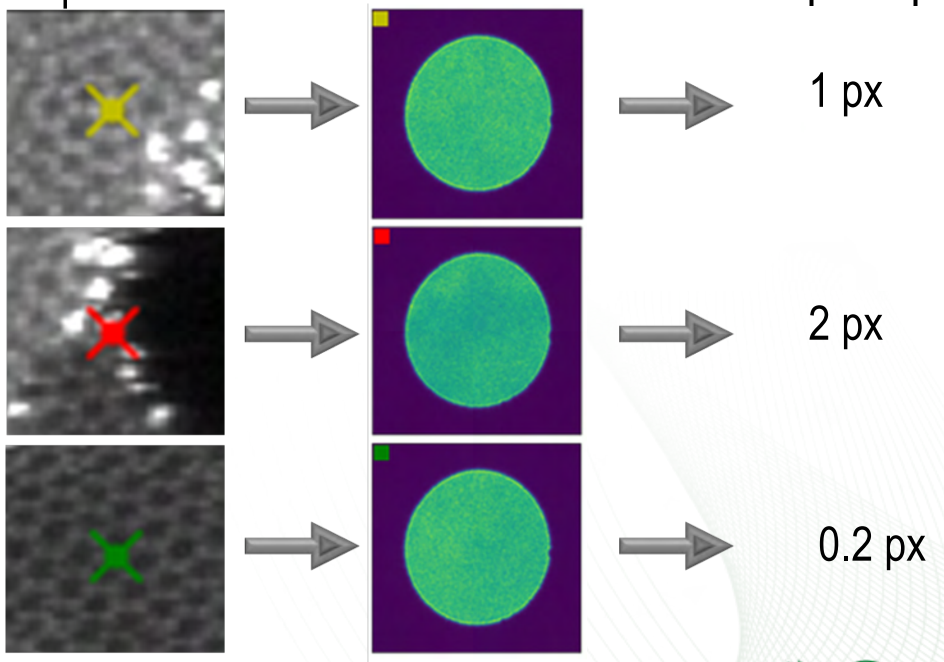
Quantities to explore

- ➢ Electric field
- ➢ Potential
- ➢ Charge density
- ➢ Strain

Choose **explorable quantity**

Traditional 4D STEM: graphene "DPC"





Automated Experiments in 4D STEM DPC example

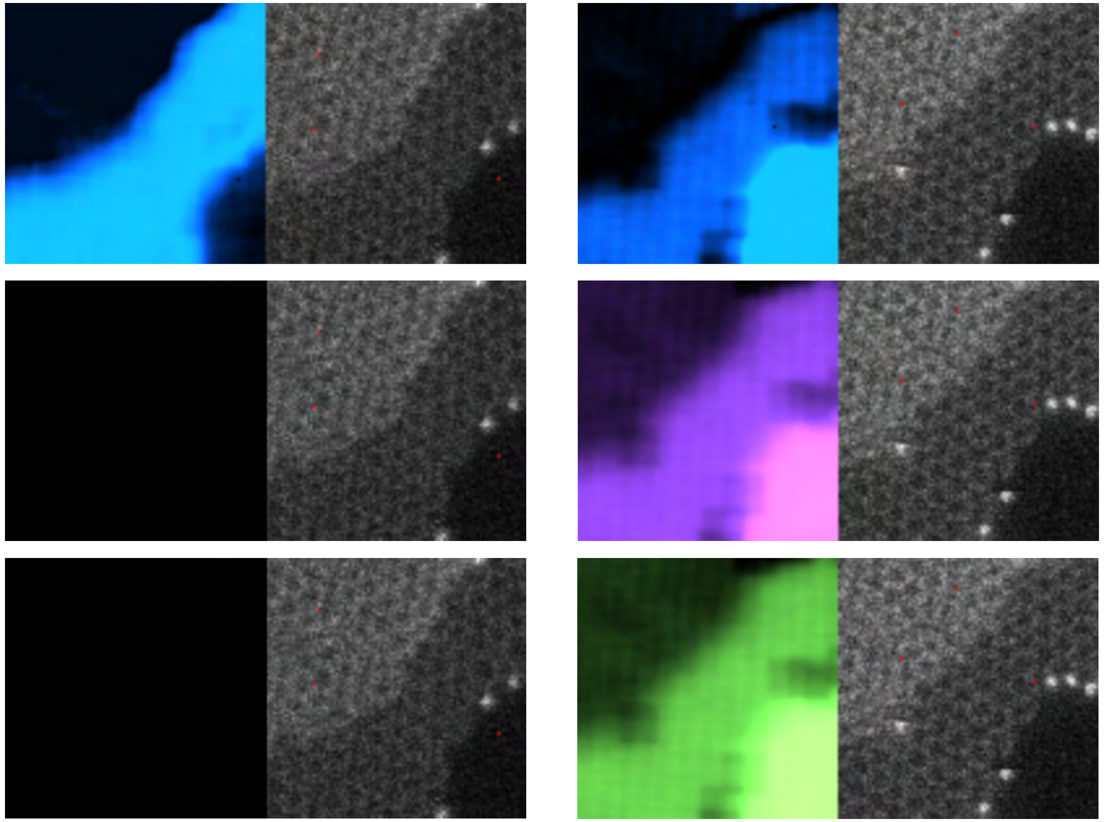
Choice of scalarizer (pre-acquired data)

CoM angle | CoM magnitude

Acquisition function

Prediction map

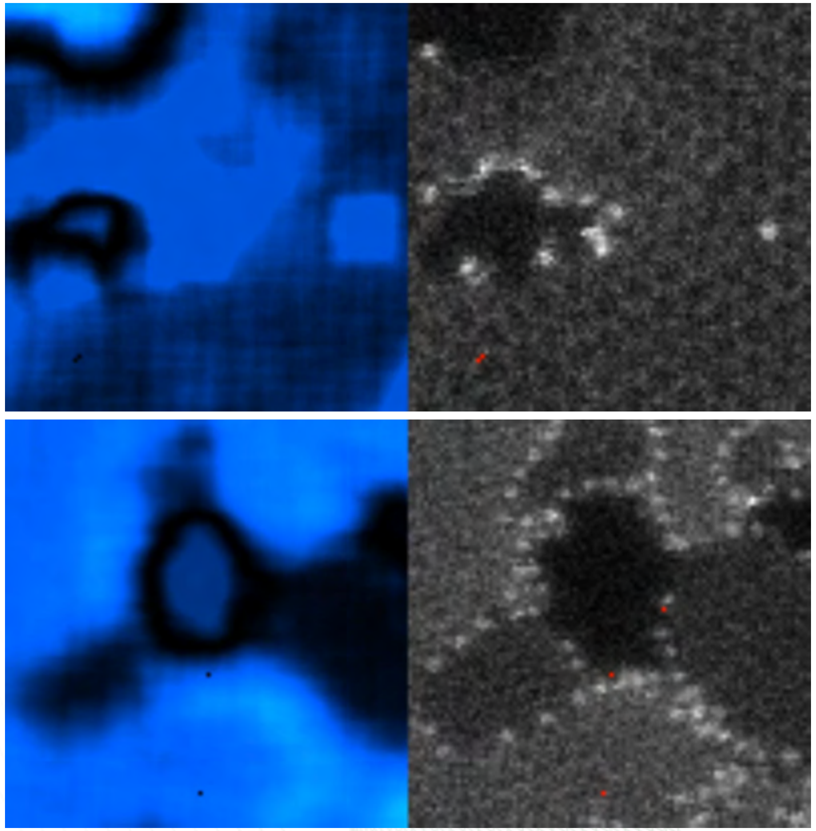
Uncertainty map



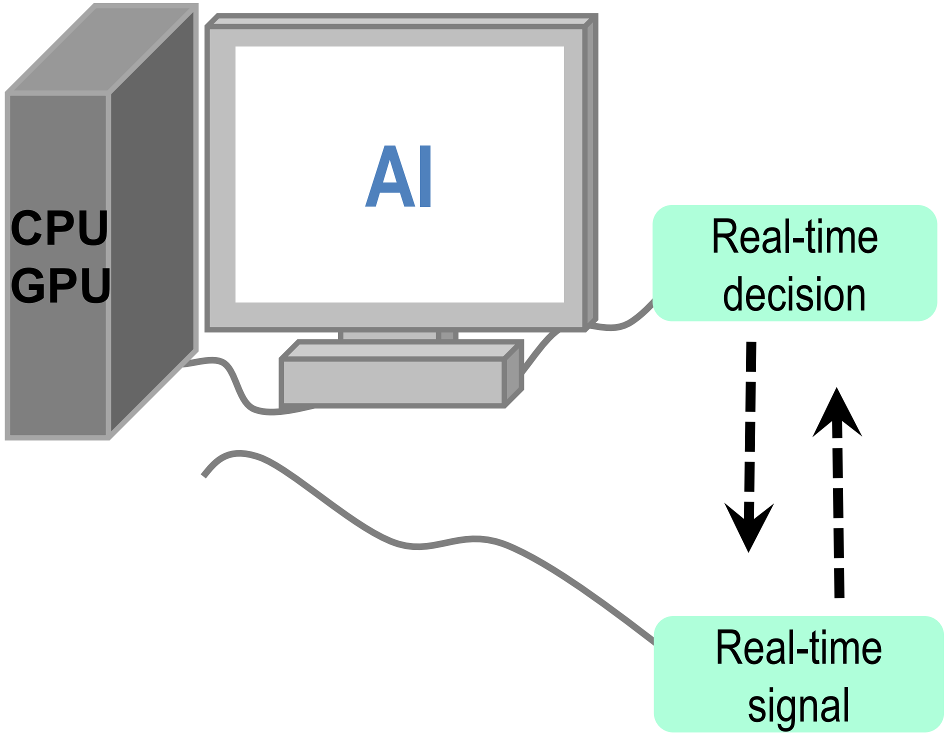
Real experiment

Next measurement map

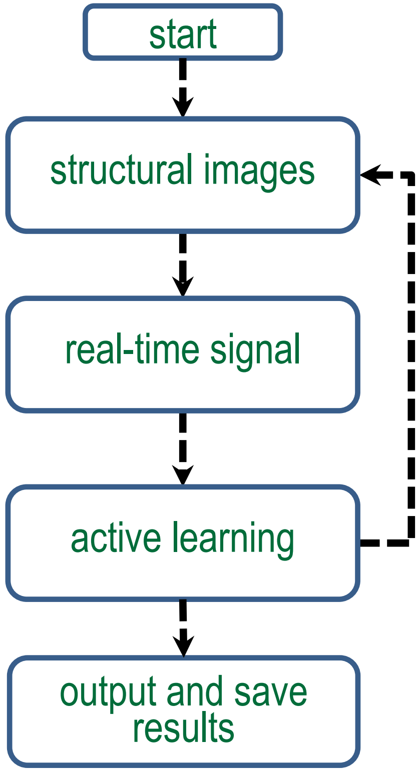
Visited points

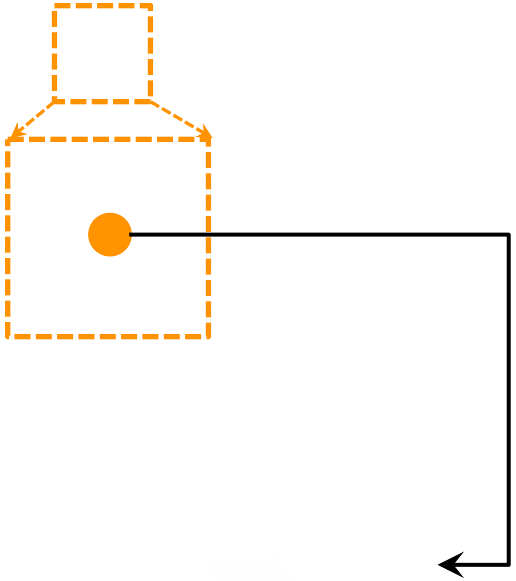


Deep Kernel Learning for PFM



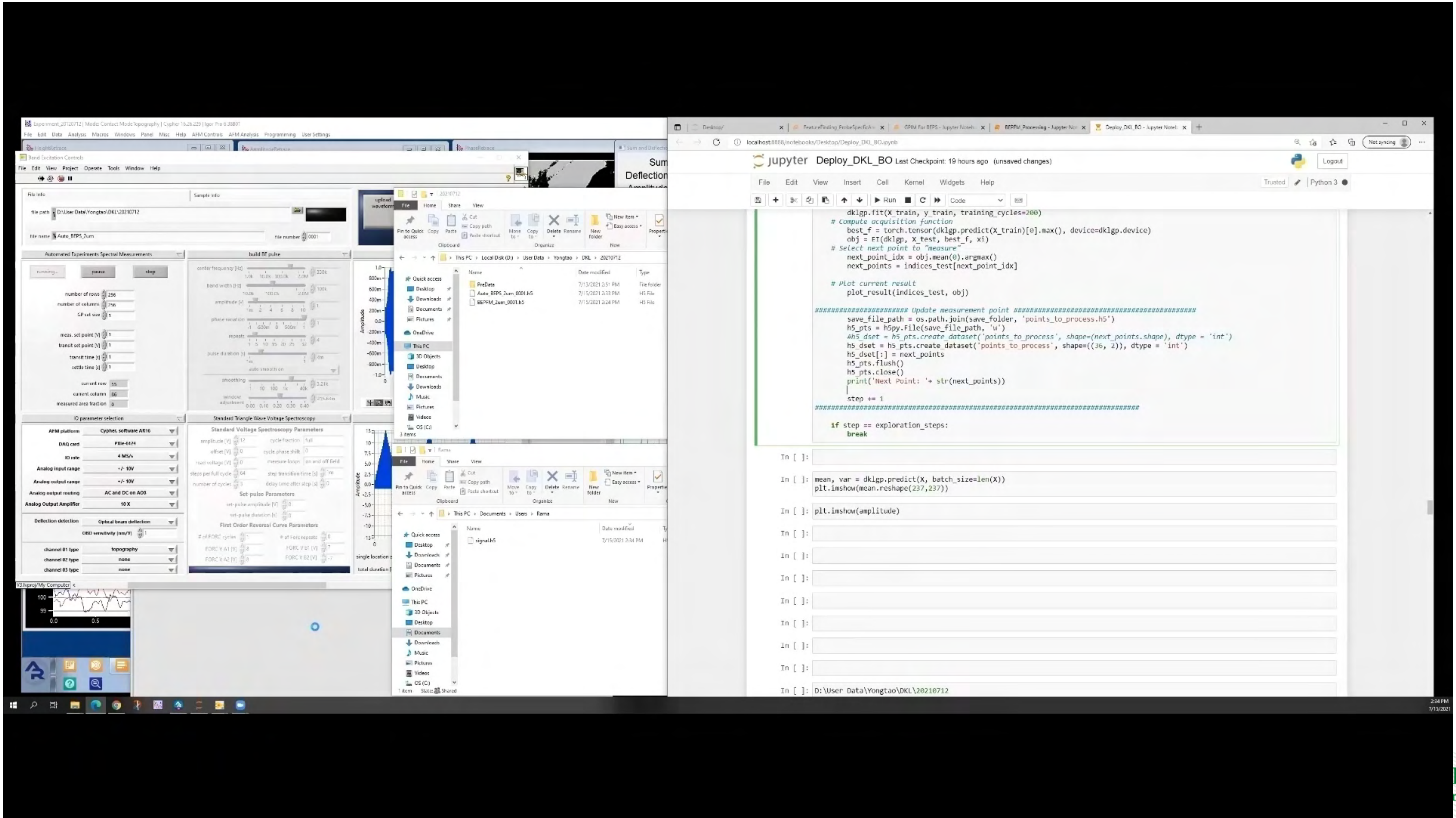
Active Learning





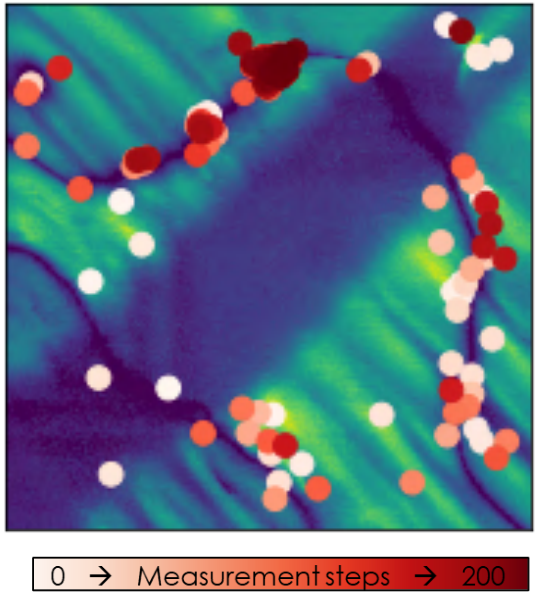
- ➢ (Deep Kernel Learning) Active learning of structure-property correlation.

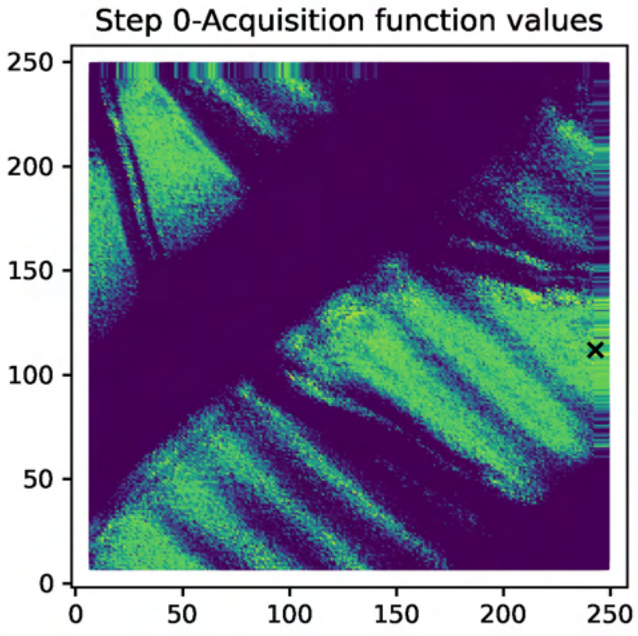
Deep Kernel Learning AE



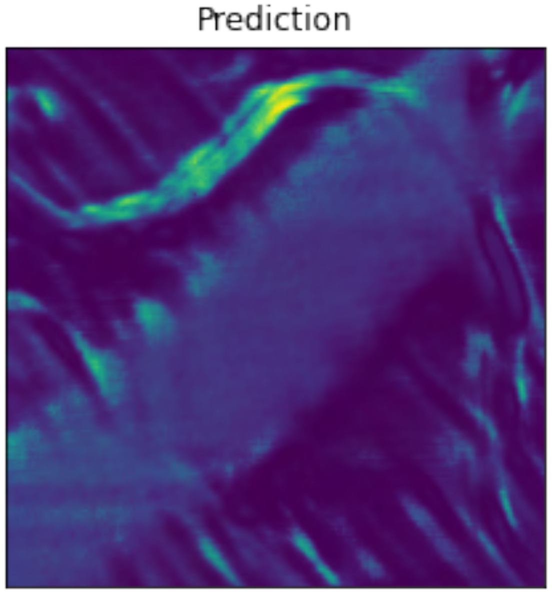
DKL AE

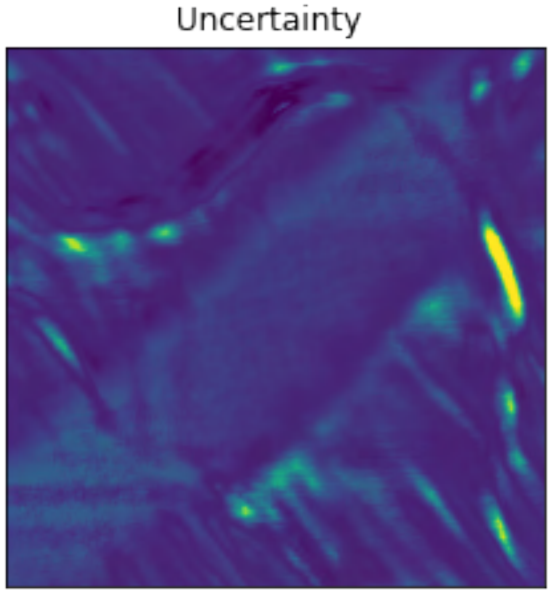
Guided by: Off field loop area





Results: DKL predicted loop area map





- Large loop opening corresponding 180^o domain walls probably due to the large polarization mobility of 180^o walls.
- Future work: DKL-nonlinearity study of HZO; CIPS domain walls.

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 system's 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

Probabilistic model

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

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

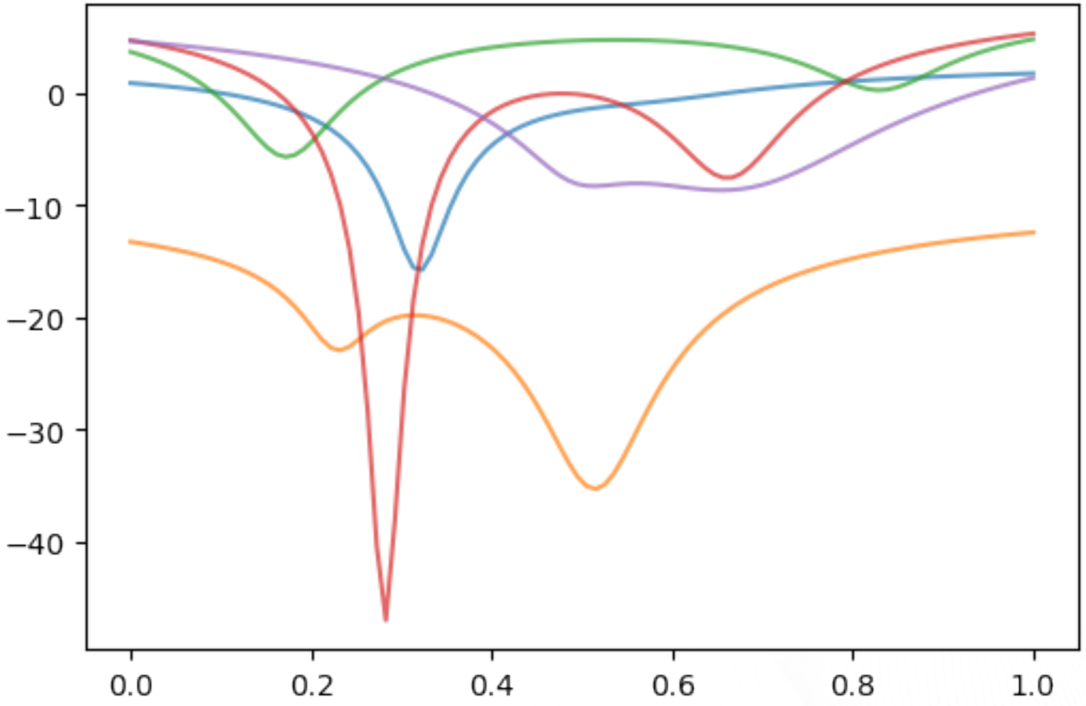
$$L_n \sim \frac{A_n}{\sqrt{(x - x_n^0)^2 + w_n^2)}}$$

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

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

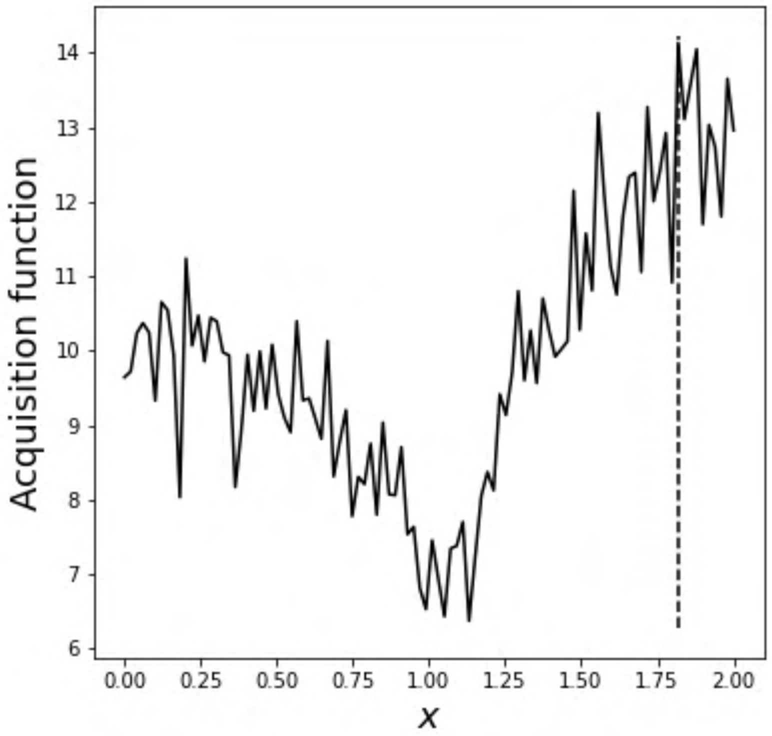
$$x_n^0 \sim Uniform(0, 1)$$

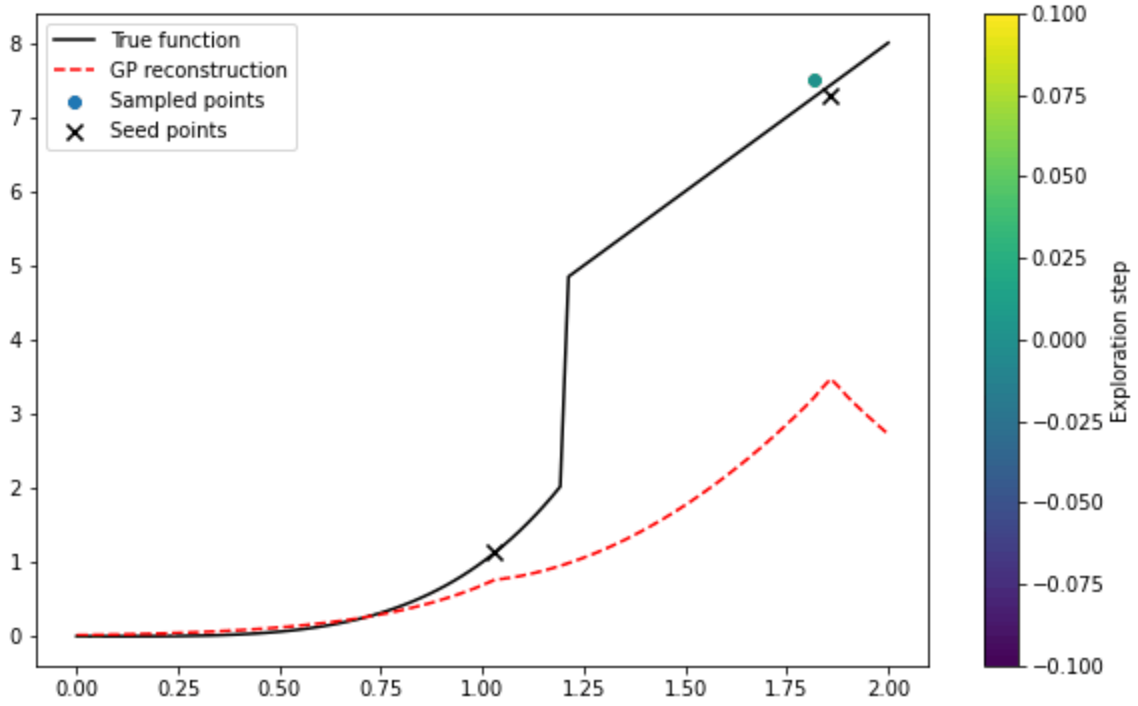
Prior predictive distribution



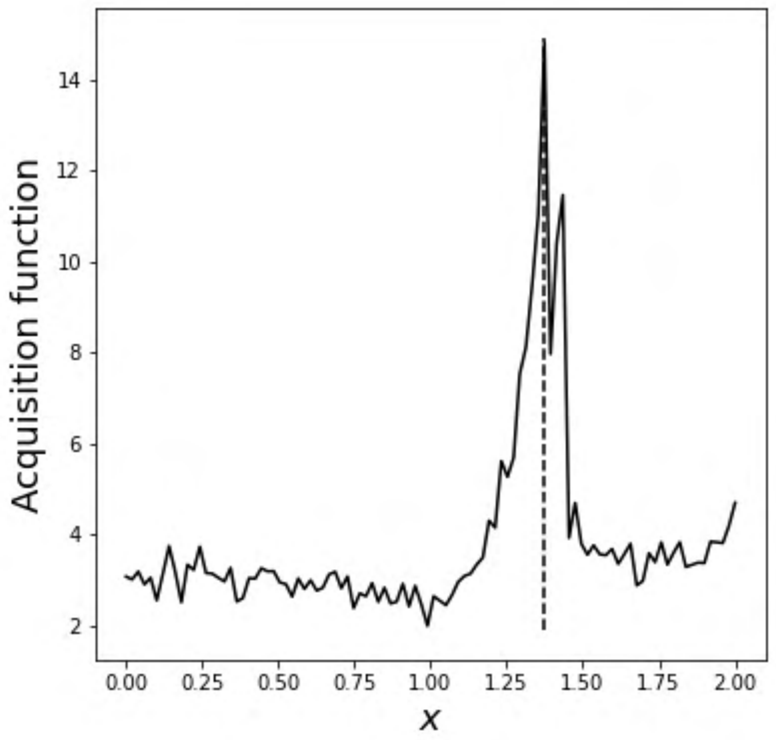
This model simply tells us that there are two minima in our data but does not assume to have any prior knowledge about their relative depth, width, or distance

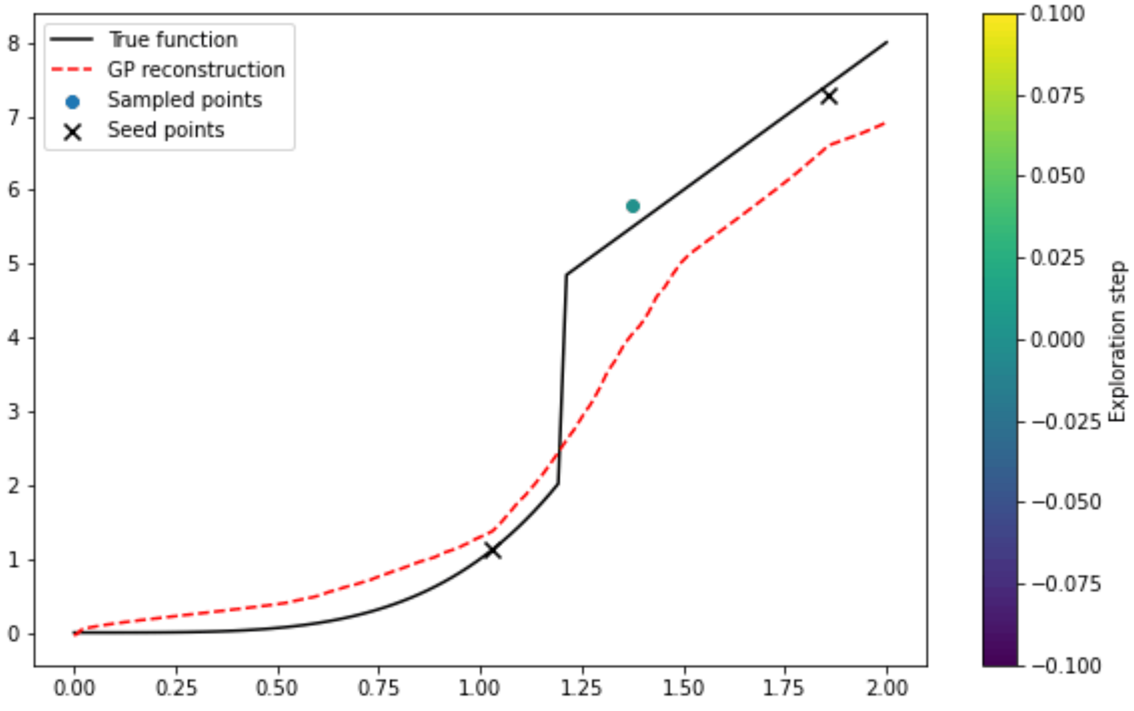
Simple GP search





Structured GP search



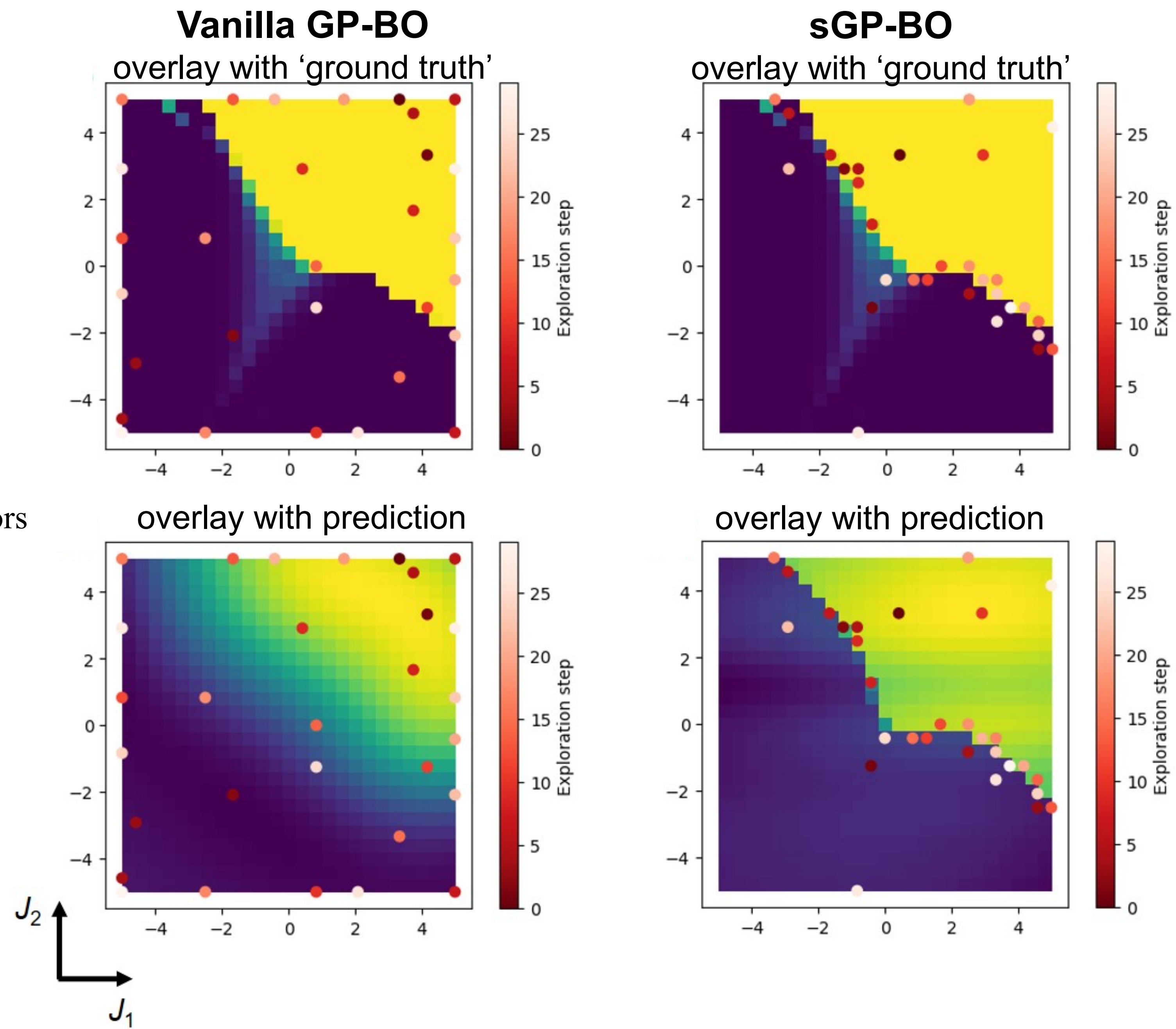


Application to Ising model

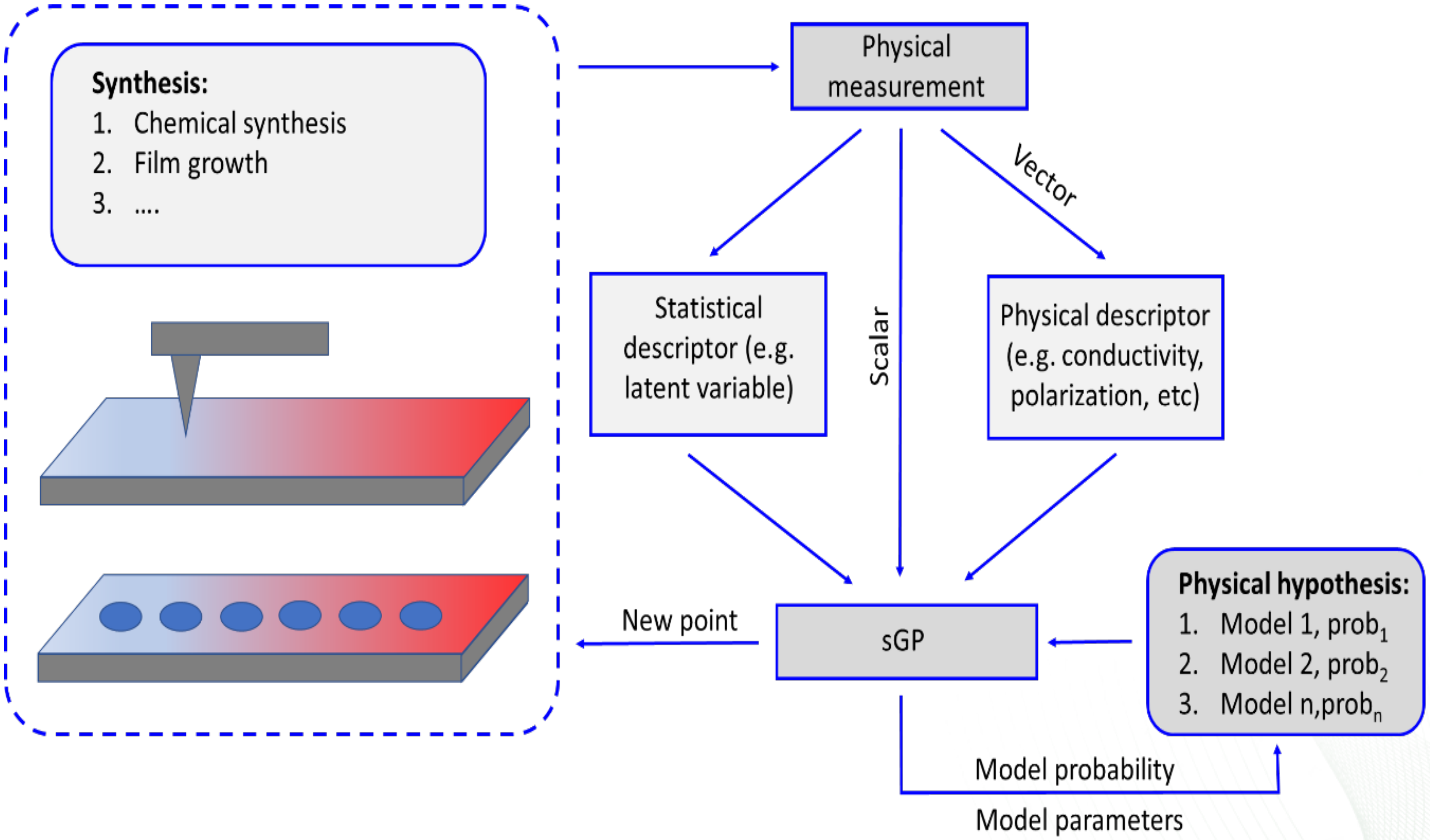
Probabilistic model

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

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

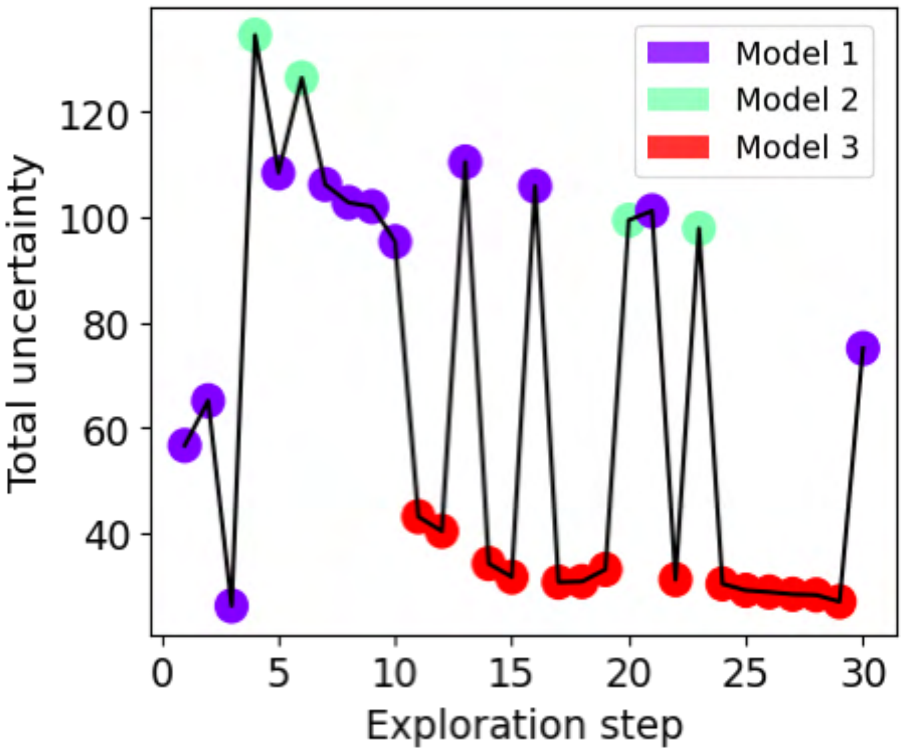


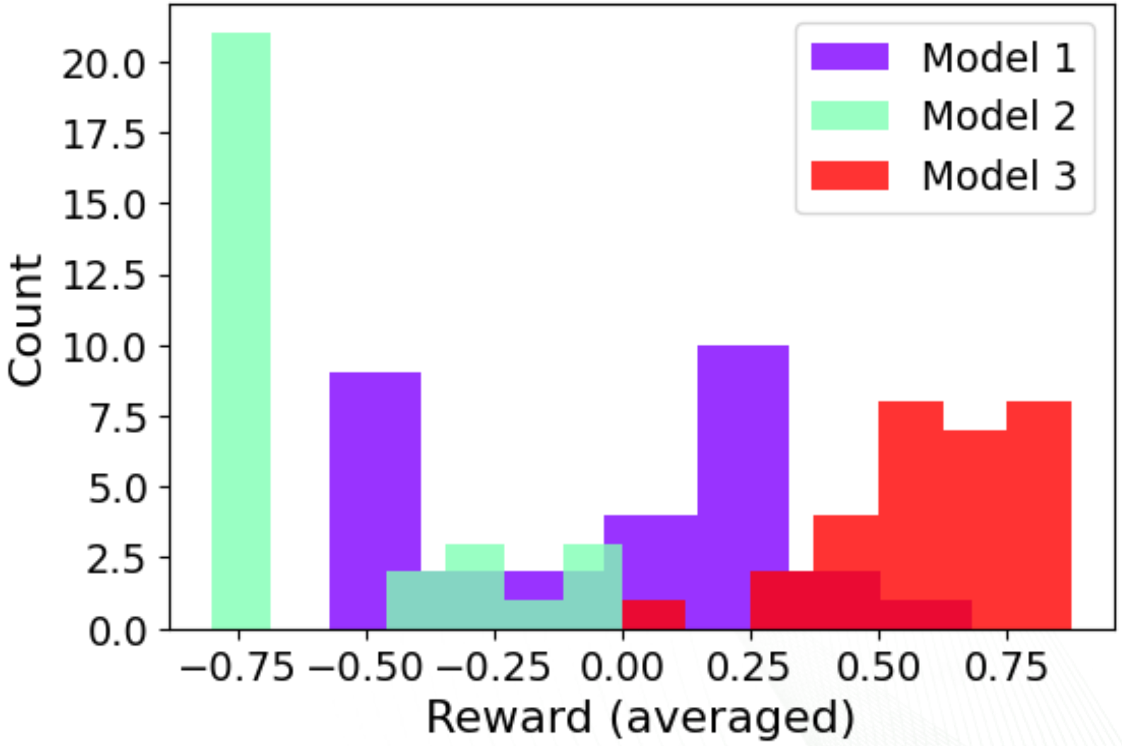
Hypothesis active learning: hypoAL



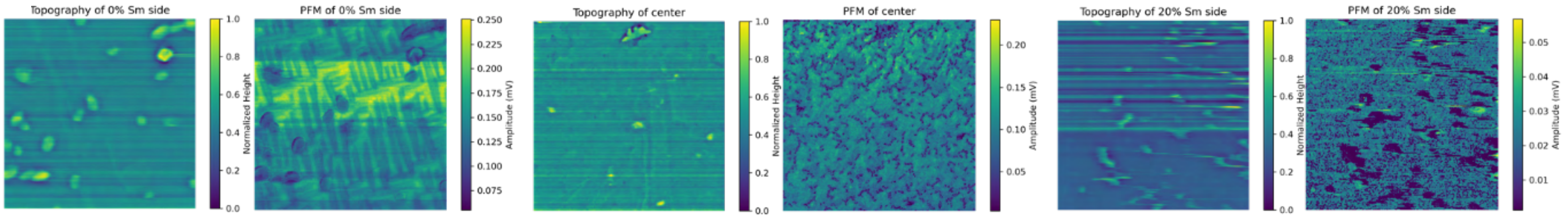
46

Next step: active model selection



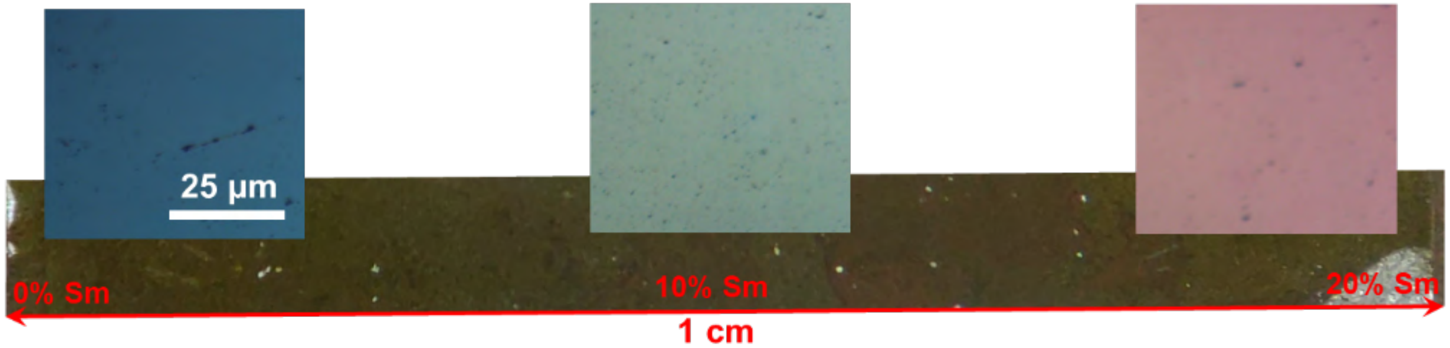


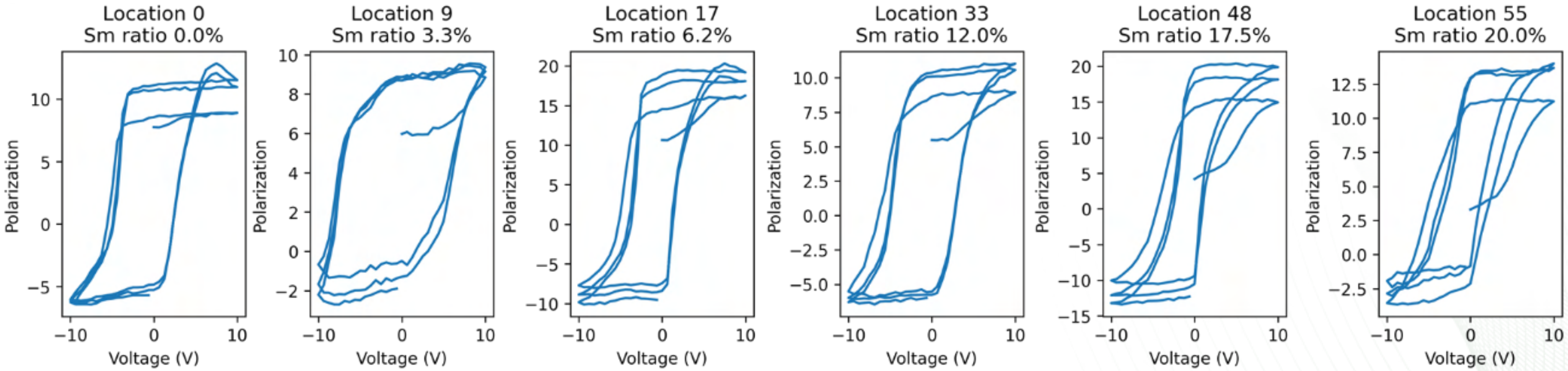
Back to combinatorial libraries:



optical images

sample picture





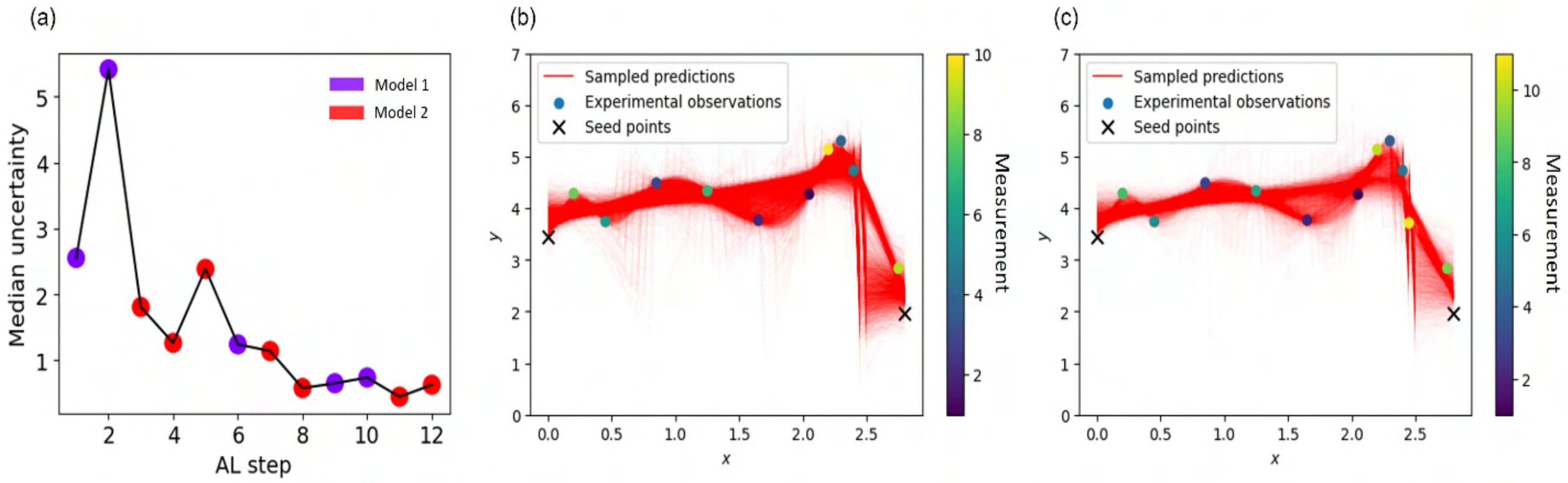
Hypothesis selection for ferroelectric response

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):

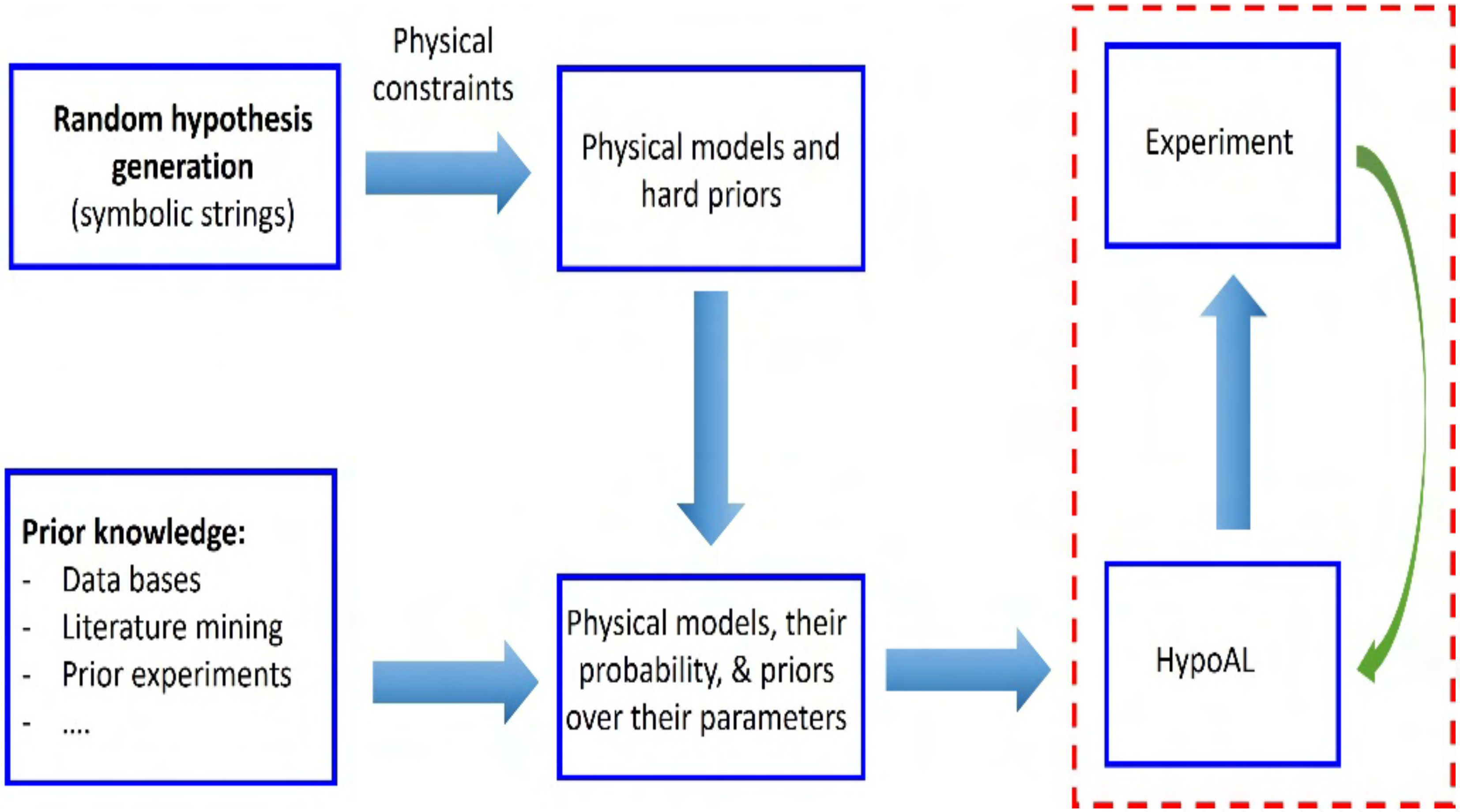
$$S = \begin{cases} S_0\left(1 - \dfrac{x}{x_0}\right)^{\frac{5}{4}} + C_0, & x \le x_c, \\ C_1, & x > x_c \end{cases}$$



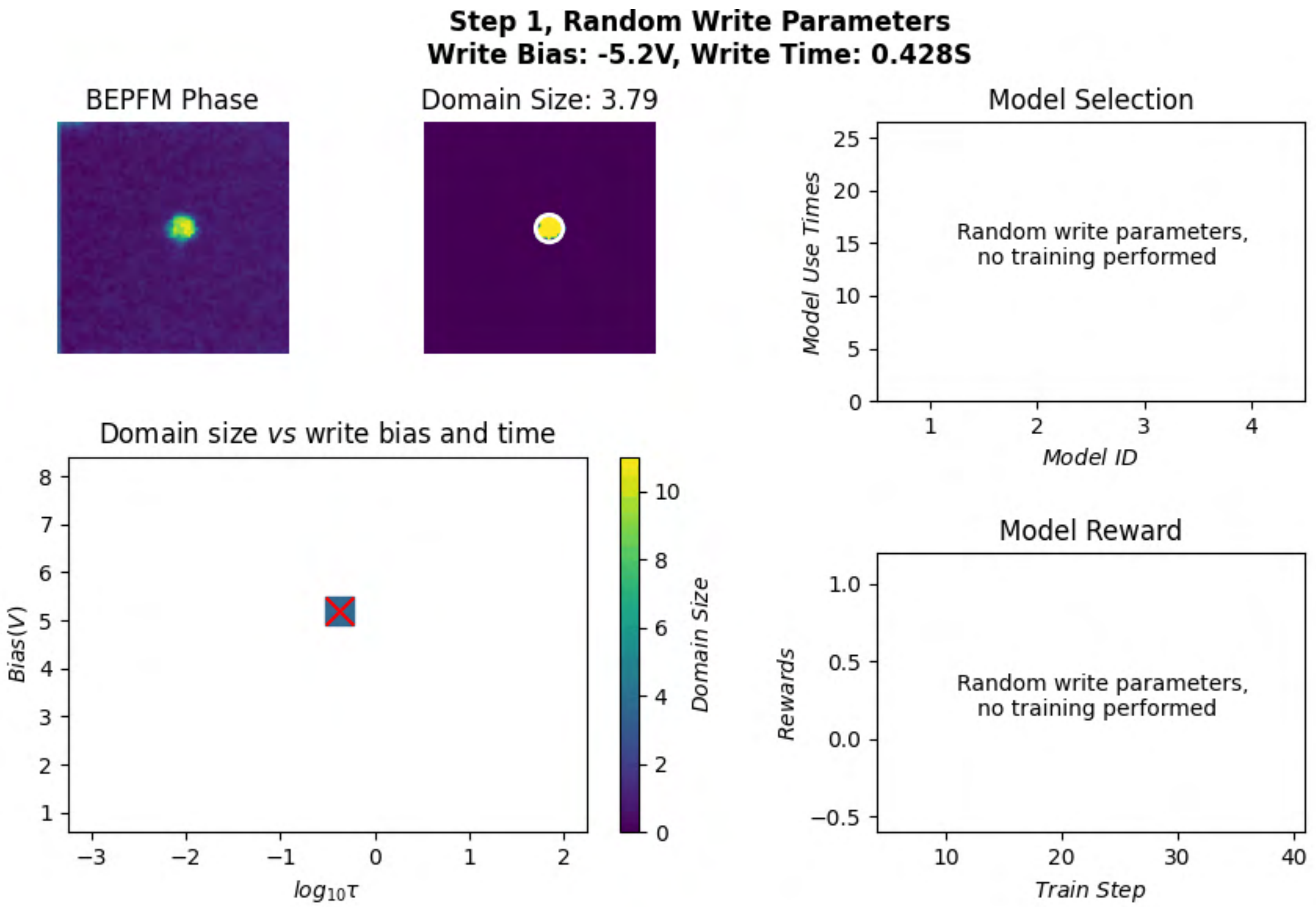
OAK RIDGE National Laboratory

Towards hypothesis learning

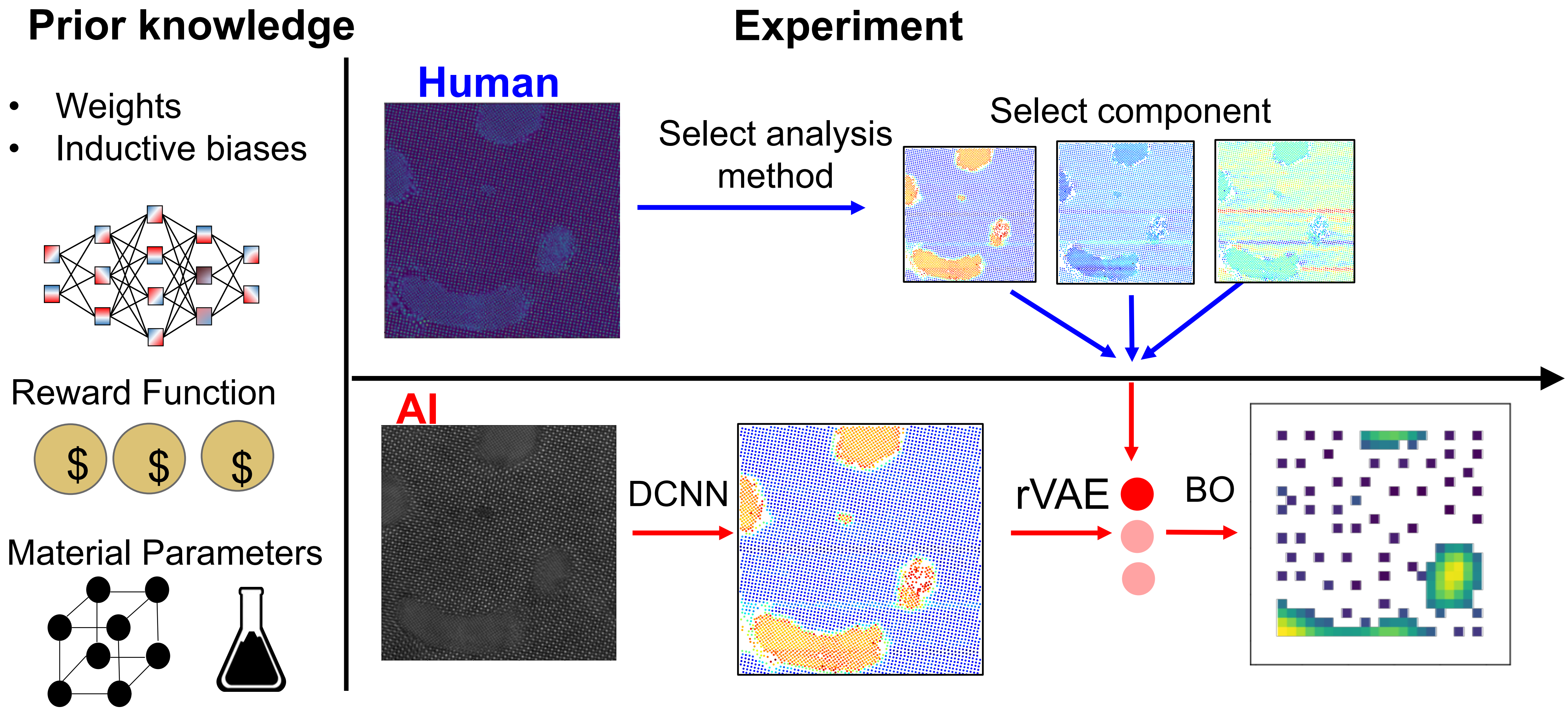
- Data already exists: Eureka, SISSO, SinDY, etc.
- What if we make hypothesis learning a part of active experiment?
- Need policies for hypothesis generation



Hypothesis learning in experiment

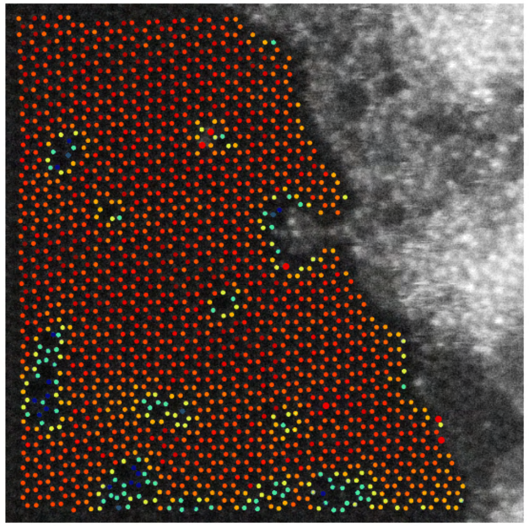


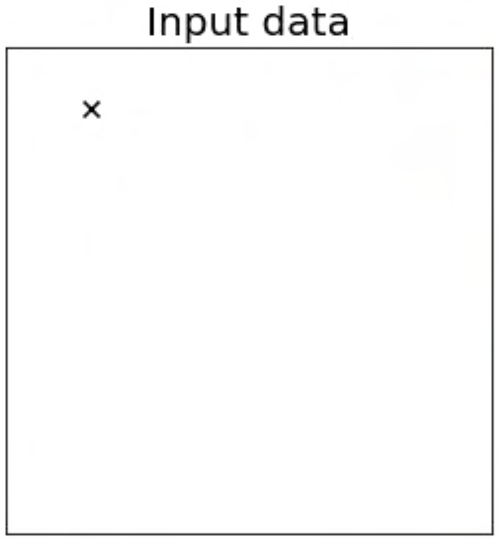
More then ML: Human in the Loop



What if I want to do it myself?

- **AtomAI:** comprehensive toolbox for DCNN-based supervised exploration of STEM and SPM Data:
- **PyroVED:** building structure-property correlations and unsupervised and semi-supervised physical discovery
- **GPim:** Gaussian processing toolbox for image analytics and automated experiment
- **gpax:** hypothesis-driven structured Gaussian Processes
- **PyCroscopy:** General data formats, workflows, and image analytics





YouTube: M*N: Microscopy, Machine Learning, Materials
Medium: https://ziatdinovmax.medium.com/

Concluding:

- **Machine learning is great, but**
  - Requires domain expertise
  - Ease of use for deployment
  - Some ML knowledge

- **Microscope is a laboratory:**
  - Engineering controls
  - Algorithms
  - Connection to domain expertise

OAK RIDGE National Laboratory | CENTER FOR NANOPHASE MATERIALS SCIENCES

Sergei V. Kalinin
sergei2@ornl.gov

Connect!

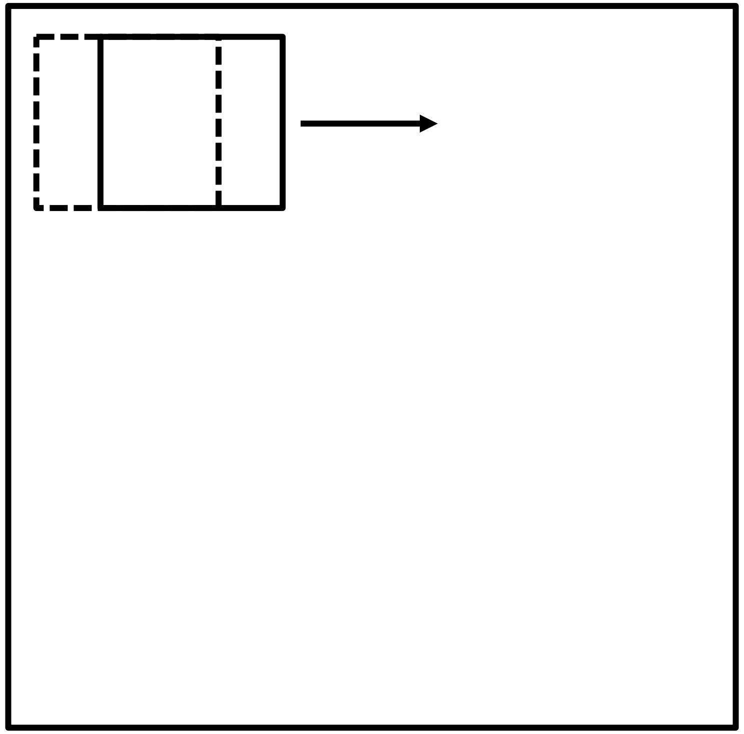
YouTube

OAK RIDGE National Laboratory

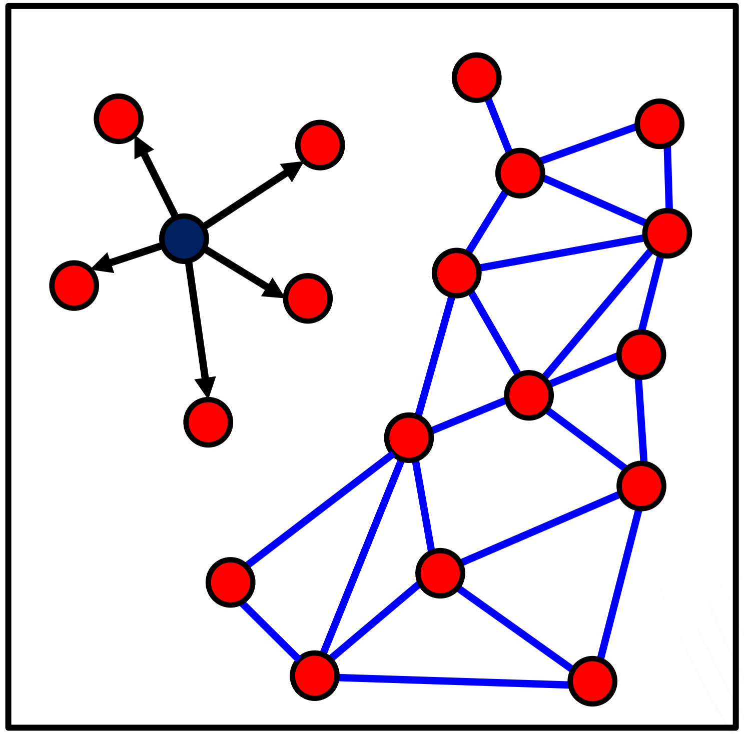
Describing the building blocks

- The classical physical descriptions (symmetry, etc) can be defined locally only in Bayesian sense
- We can argue that local descriptors are simple, if not necessarily known
- And the rules that guide their emergence are also simple, if not known

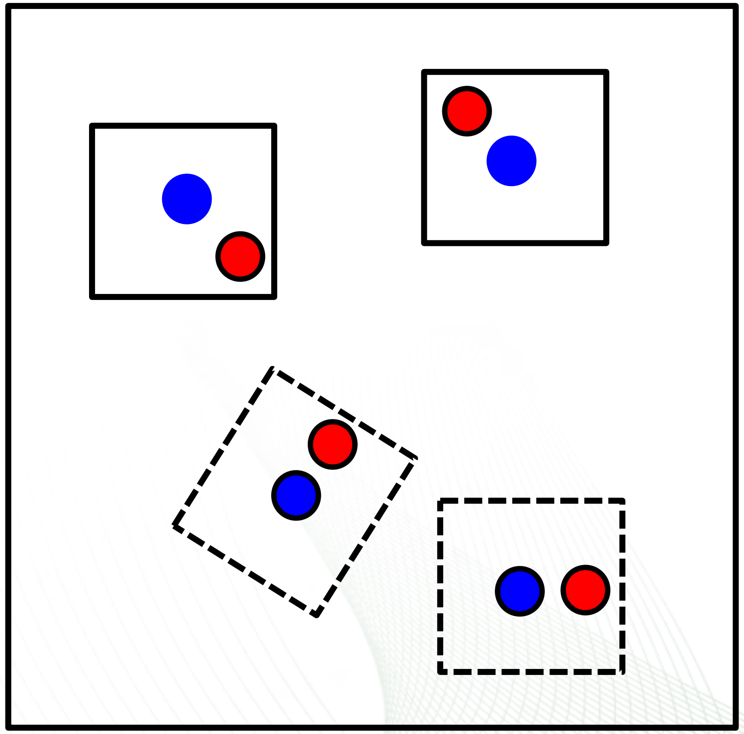
Continuous translational symmetry



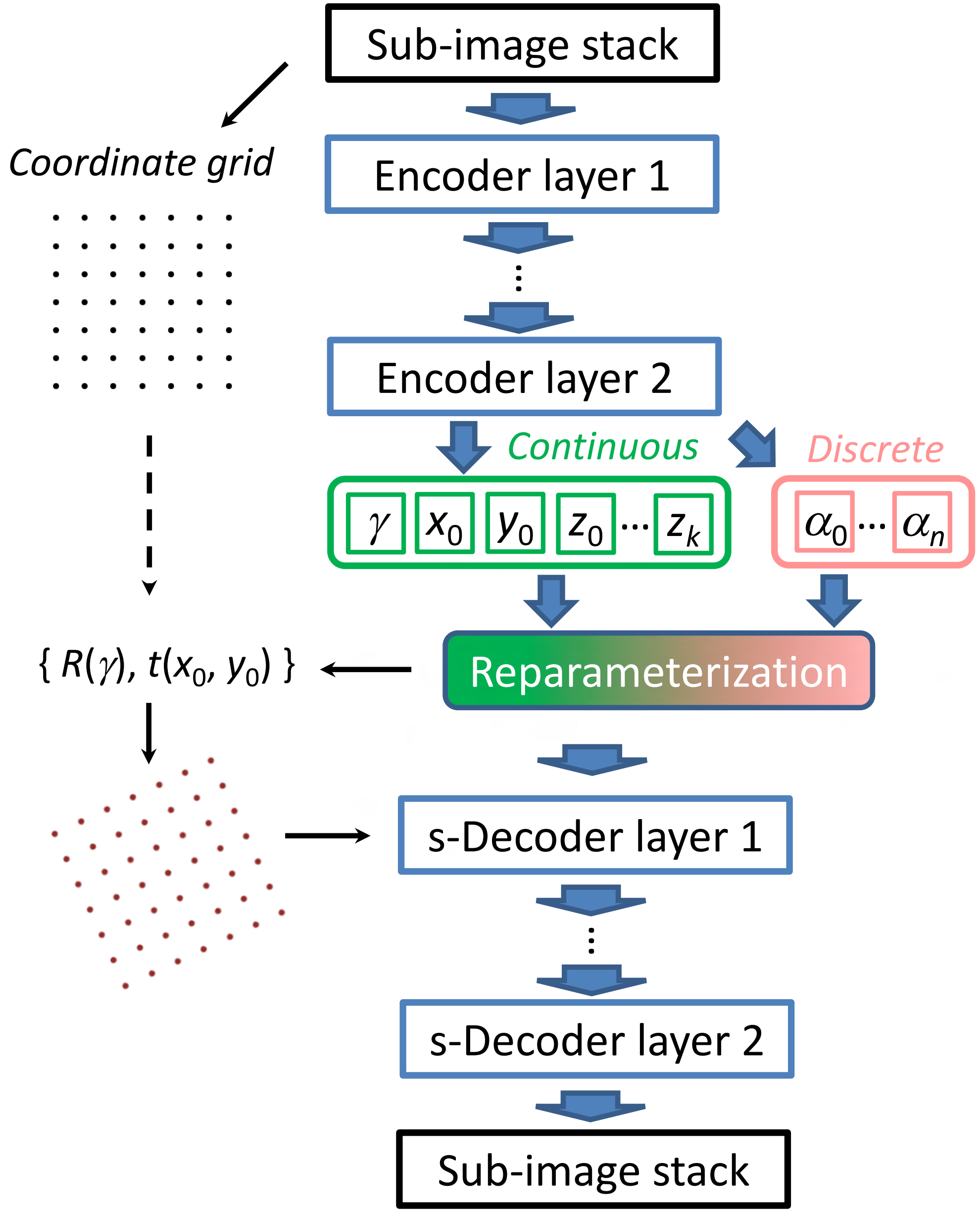
Atom based descriptions



Localized sub-images



jrVAE



- Generative model is a function of spatial coordinate (*e.g.*, via spatial broadcasting)
- 3 additional latent variables to absorb rotations and shifts
- Disentangles rotations and translations from image content
- Learns discrete classes in unsupervised fashion
- Well-suited for analyzing microscopy (sub-)images on atomic and molecular levels

$$ELBO =$$

$$-\,Reconstruction\ Loss$$

$$-\beta_c(t)\left|\left(D_{KL}\big(\underbrace{q(z|x)\|p(z)}_{\text{Structure}}\big) + D_{KL}\big(\underbrace{q(\gamma|x)\|p(\gamma)}_{\text{Rotation}}\big)\right) - C_z\right| \quad \text{Continuous}$$

$$-\beta_d(t)\left|D_{KL}\big(q(\alpha|x)\|p(\alpha)\big) - C_\alpha\right| \quad \text{Discrete}$$

+ physics-based “loss” ?

Theory in the loop

- Theory provides a “map” of the physical phenomena with the level of detail dependent on complexity and veracity of the theory. However, unless the map is very good, we cannot drive by the map
- Experiment gives us immediate feedback, but it may be unclear which direction to choose
- Theory experiment matching allows to combine general direction and local details
- We aim to implement automated experiment, both fully correlative (self-driving car), and model based (theory “map” is updated when driving

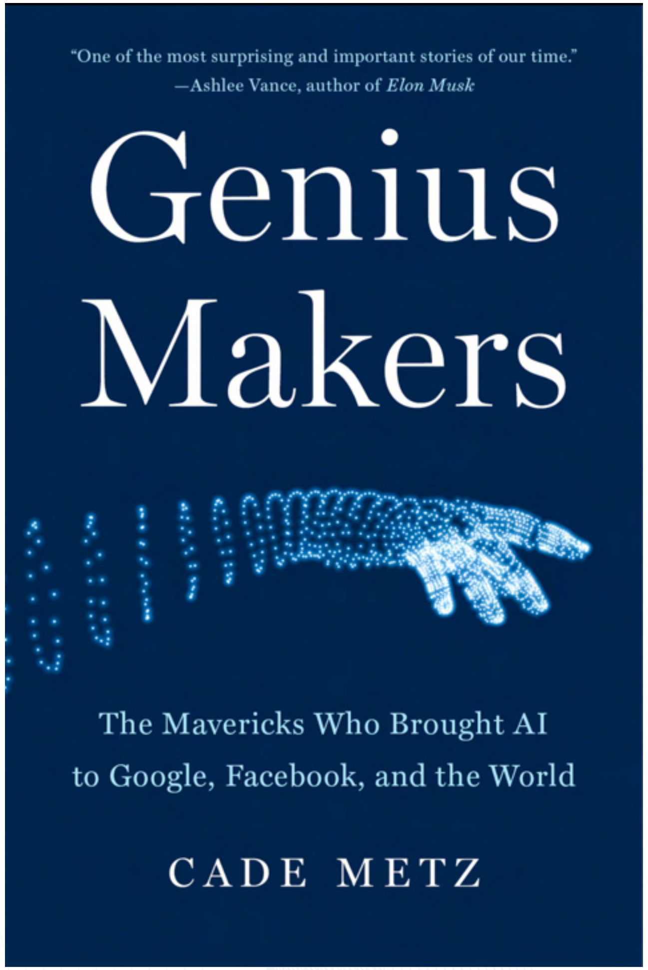
OAK RIDGE National Laboratory

Enter the Scientific Workflows

"*One of the things that is not understood at all by the conventional forces in society, if you will, is that America benefits enormously from these global platforms—global platforms that are built in America, whether it's the Internet itself, email, Android, the iPhone, et cetera,*"

Schmidt says. If a company, or indeed a country, controlled the platform, it controlled what ran on top of it. A creation like Google's TensorFlow was the latest example.

"*It's a global platform competition, and it is extremely important that the platforms be invented in America. Platforms establish a base by which innovation occurs in the future.*"



Oak Ridge National Laboratory

What do we hope to achieve?

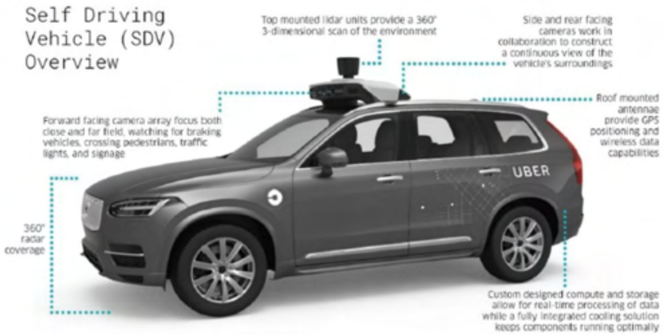
Microscopy today:

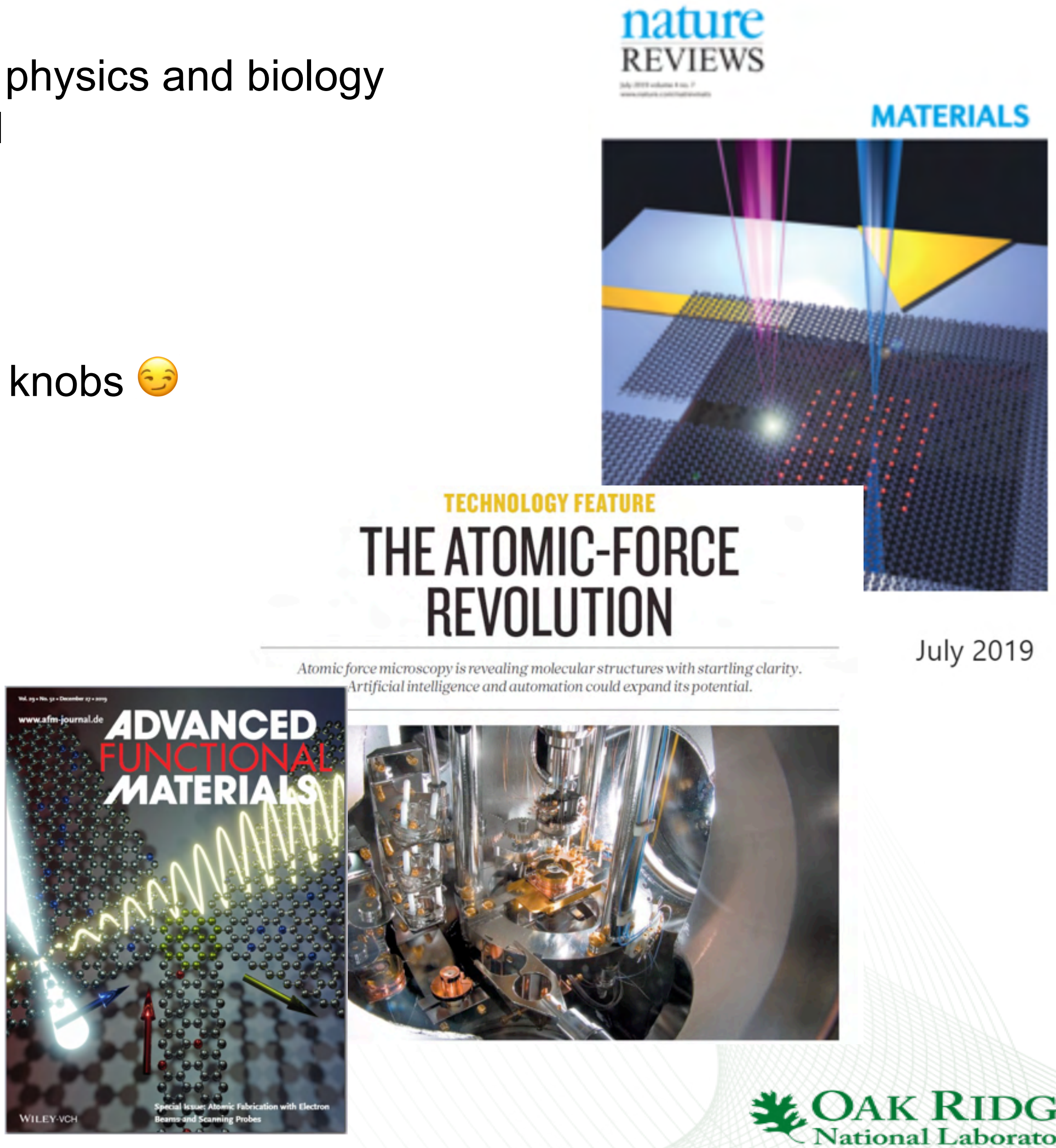
- Primary component of research in materials science, physics and biology
- 1000s of high-end (S)TEM platforms, ~10,000 overall
- 1000s of high-end UHV SPMs, >50,000 ambient
- Chemical and mass-spectrometric imaging

What do microscopists do?

- Most of the time - sit alone in the dark room and turn knobs 😏
- Limited amount of collected data
- Case for automation: CryoEM

Unsurprisingly, inspired by autonomous cars, etc. – multiple proposals to make automated microscopes!





July 2019