Remote and on-the-fly: artificial intelligence driven science in laboratories and central facilities.

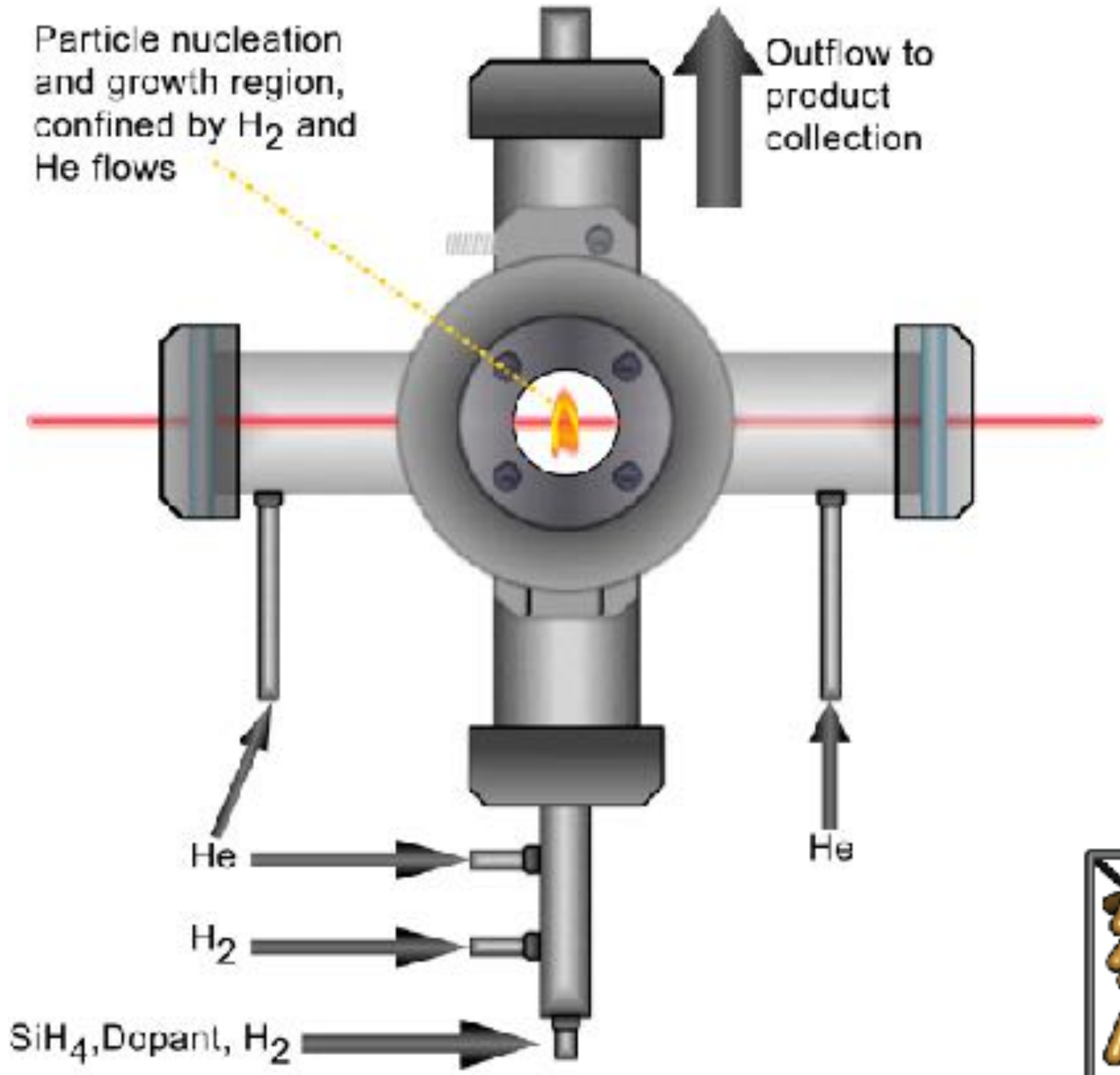
Dr. Phil Maffettone

NYSDS Oct 2021

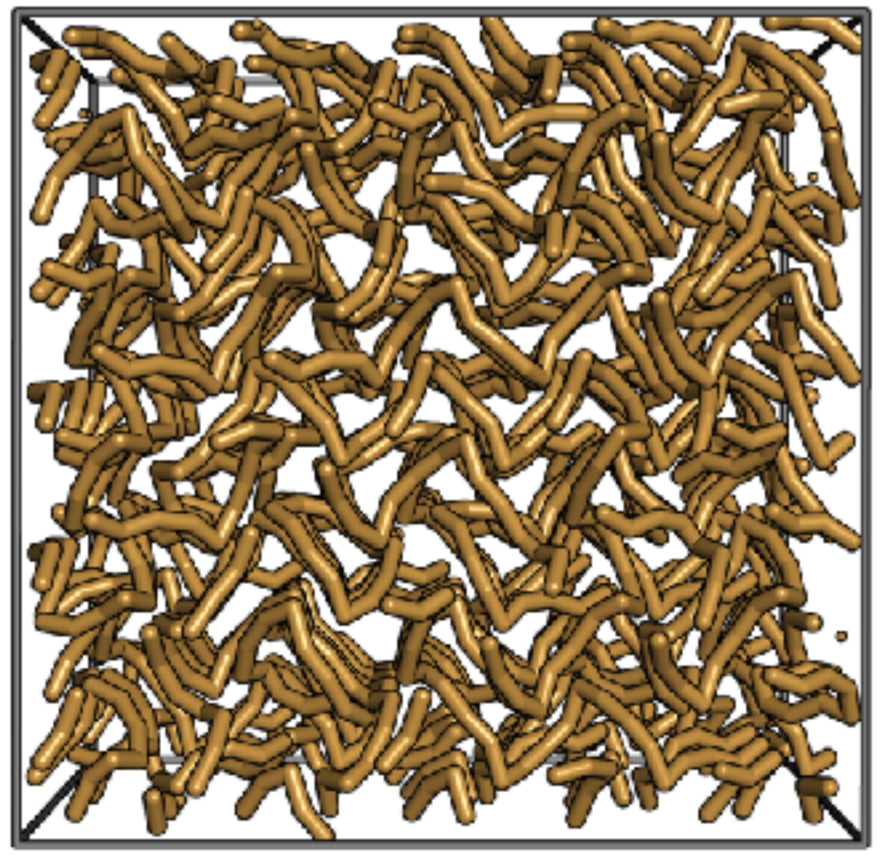
U.S. DEPARTMENT OF ENERGY | Office of Science

BROOKHAVEN NATIONAL LABORATORY

maffettone@dssi:~/$ whoami



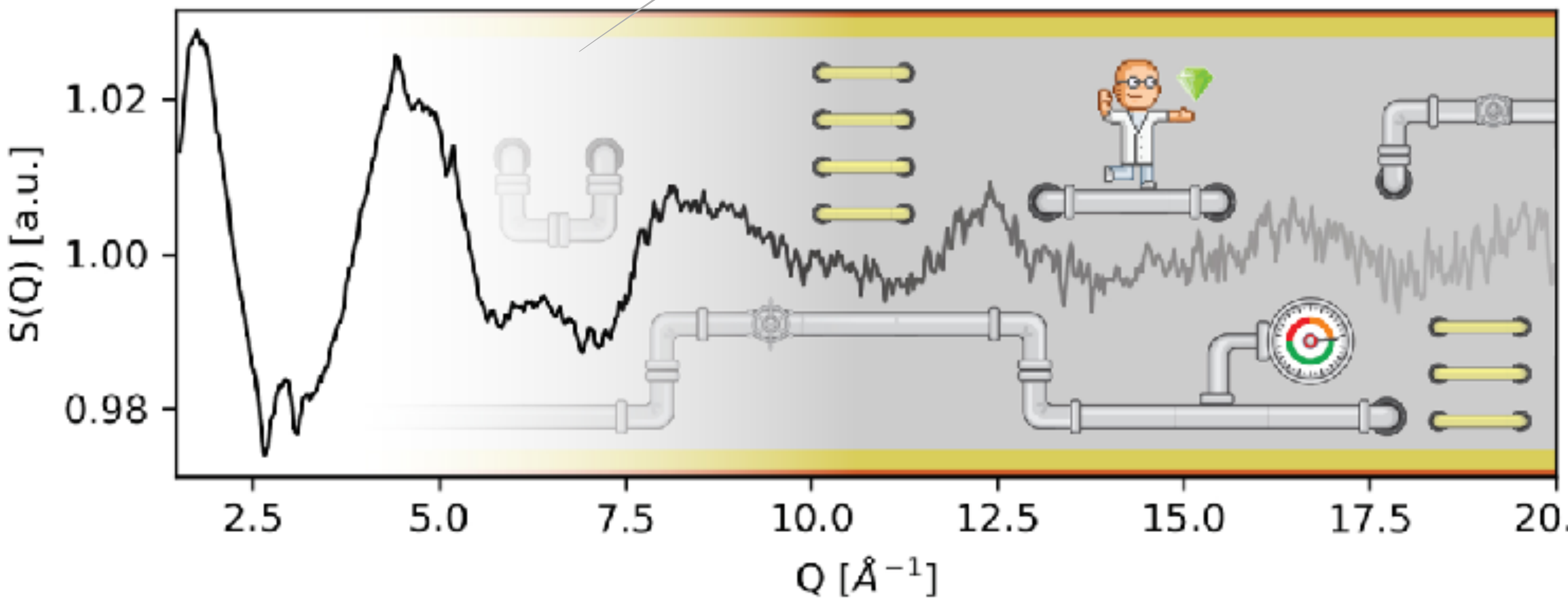
UNIVERSITY OF OXFORD
DOMINVS ILLVMINATIO MEA



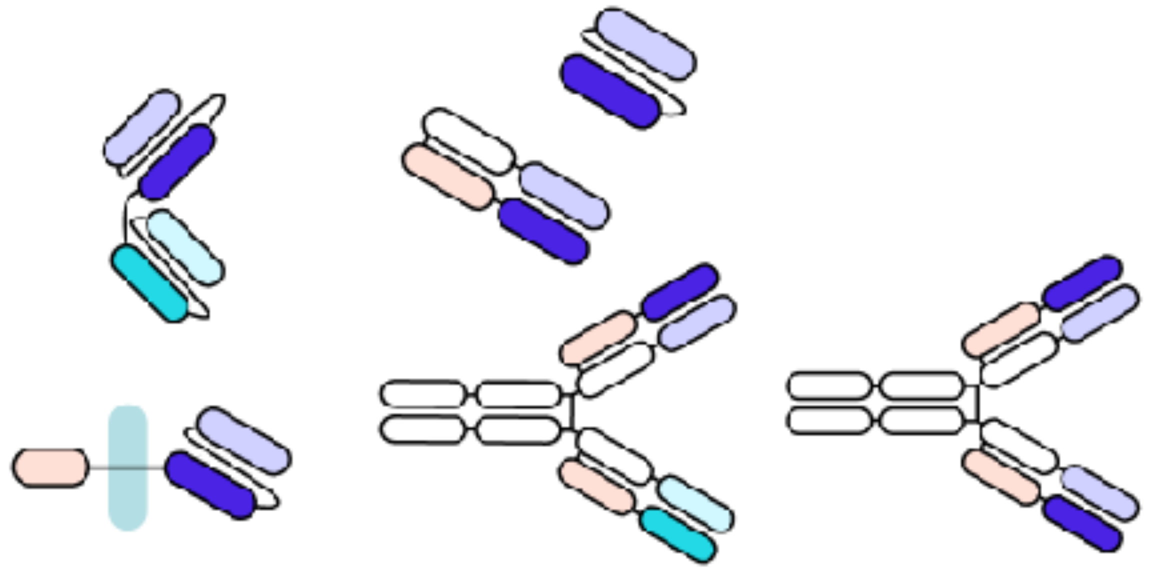
UB

UNIVERSITY OF LIVERPOOL

BROOKHAVEN
NATIONAL LABORATORY



BigHat
BIOSCIENCES



maffettone@dssi:~/$ members {ALL}

University of Liverpool

Prof. Andy Cooper

Peng Cui

Dr. Marc Little

Dr. Linjiang Chen

Xiaobo Liu

Dr. Tao Liu

Yu Che

Dr. Vladimir Gusev

Dr. Benjamin Burger

NSLS-II

Dr. Daniel Olds

Clara Cook

Dr. Thomas Caswell

Joshua Lynch

Dr. Stuart Campbell

Dr. Tatiana Konstantinova

Dr. Andi Barbour

Dr. Bruce Ravel

Rhur University

Lars Banko

Dr. Yury Lysogorskiy

Prof. Alfred Ludwig

NYU

Nate Gruver

Sam Stanton

Marc Finzi

Polina Kirichenko

Prof. Andrew Gordon Wilson

Columbia University

Dr. Boyan Penkov

Prof. Ken Shepard

Flatiron Institute

Dr. Aidan Daly

BigHat Biosciences

Dr. Peyton Greenside

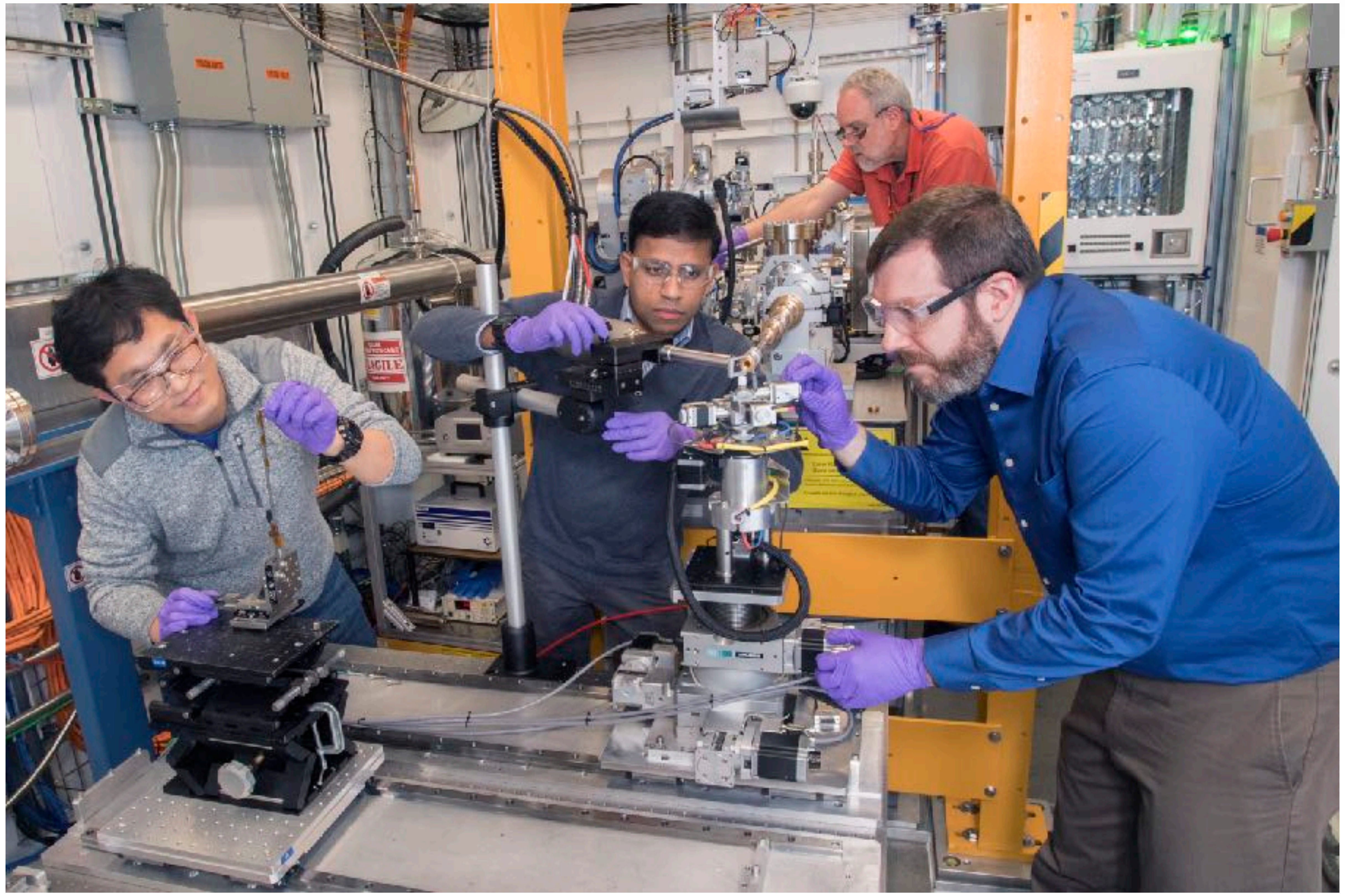
Dr. Emily Delany

Vivek Myers

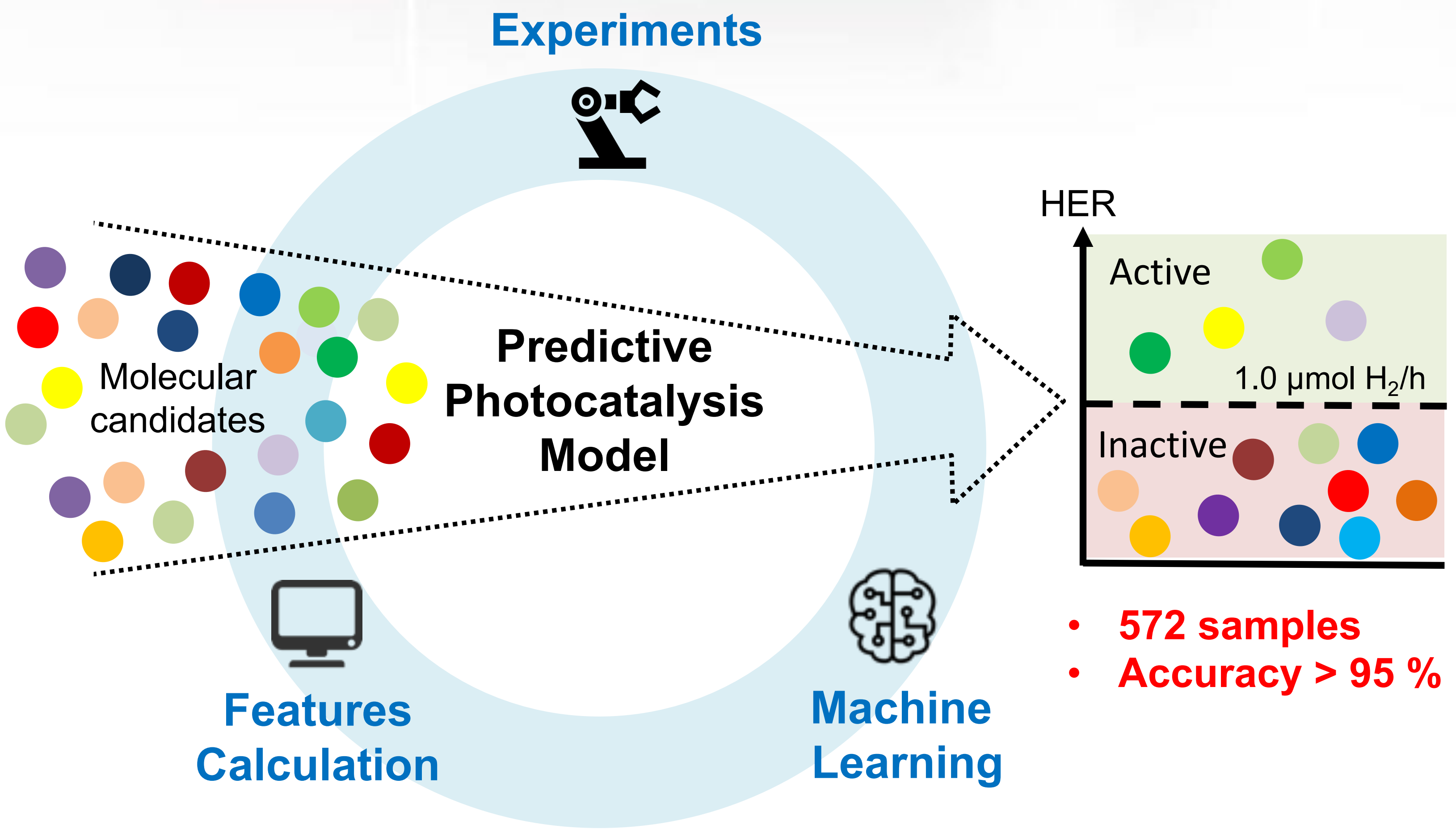
Aaron Solomon

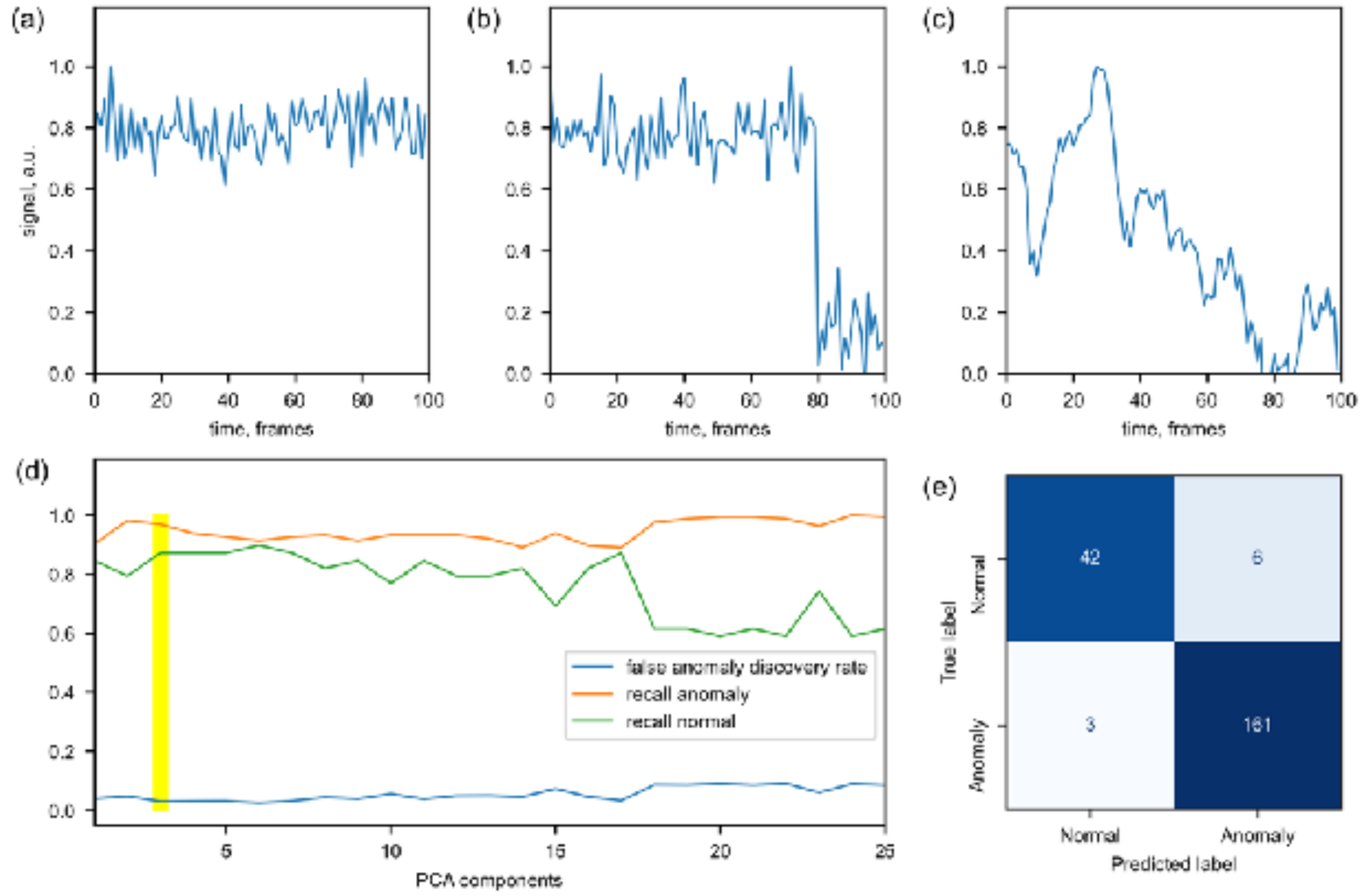
Federated AI, data streaming, and pragmatic engineering

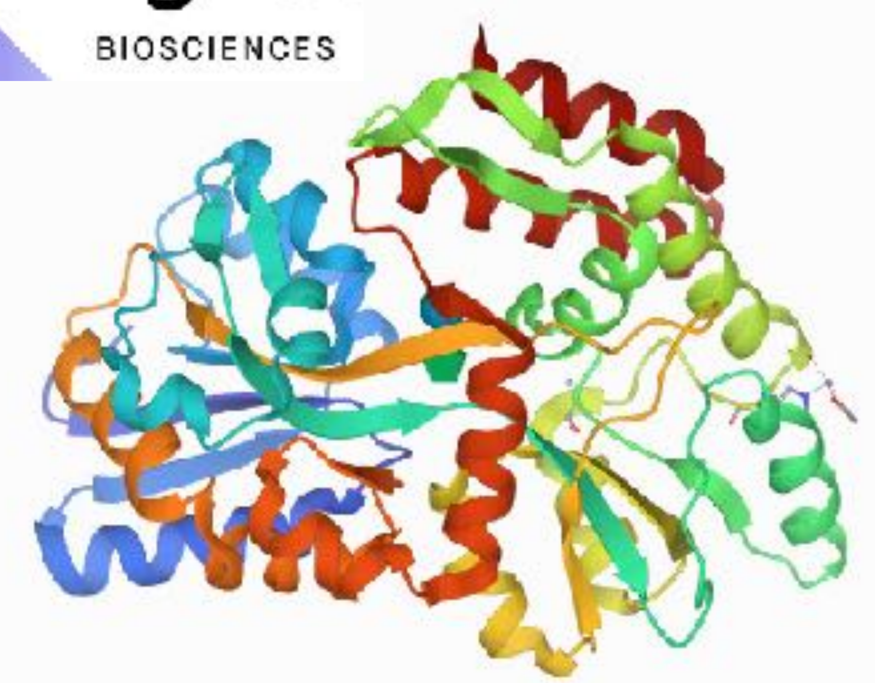
- There is no one-size-fits-all AI/ML approach for any science.
- Federations of agents can solve different tasks asynchronously.
- Data streaming enables this.
- Collaboration drawing on domain knowledge and AI/ML expertise results in the most impactful projects.

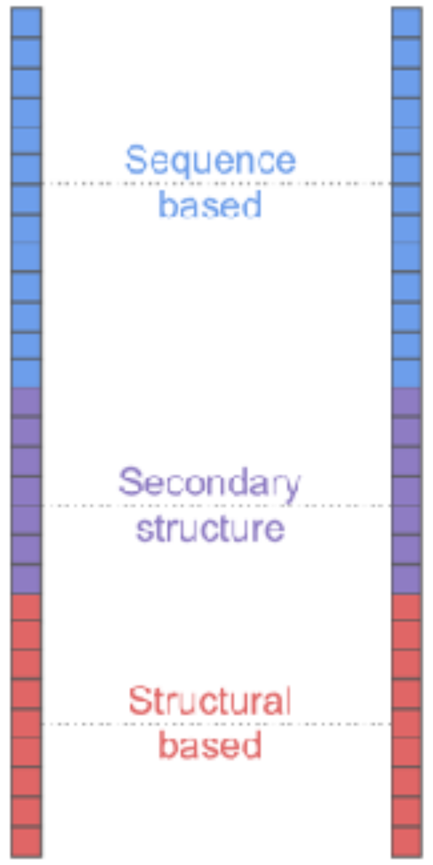


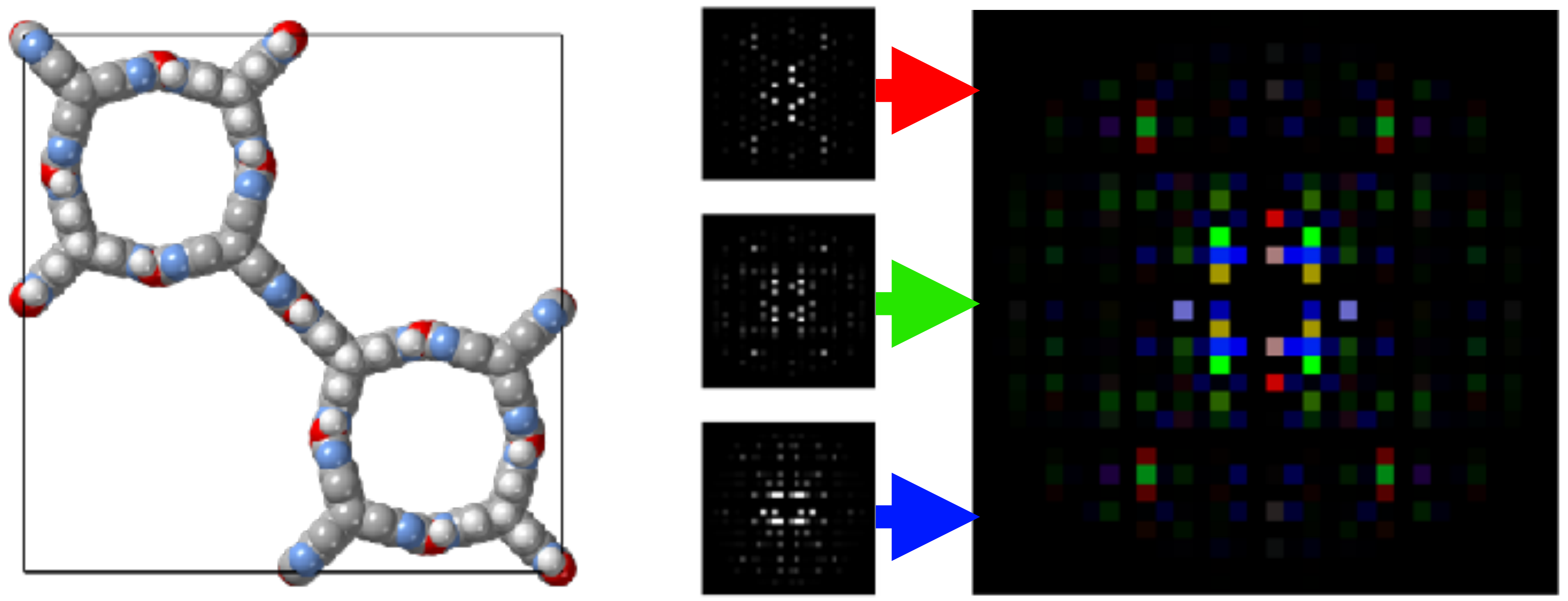
Feature engineering can be incredibly effective, and requires collaboration with domain experts.







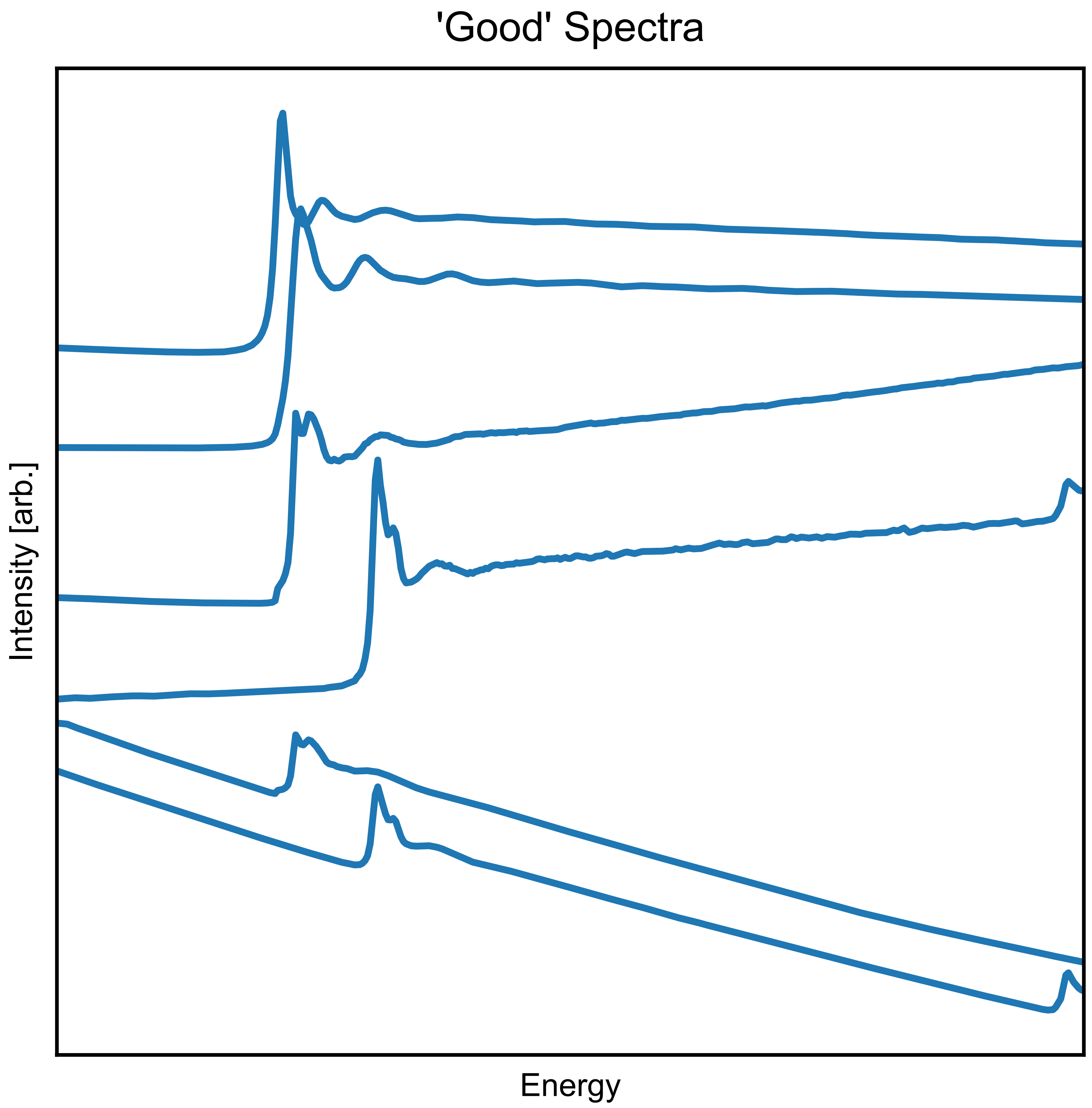


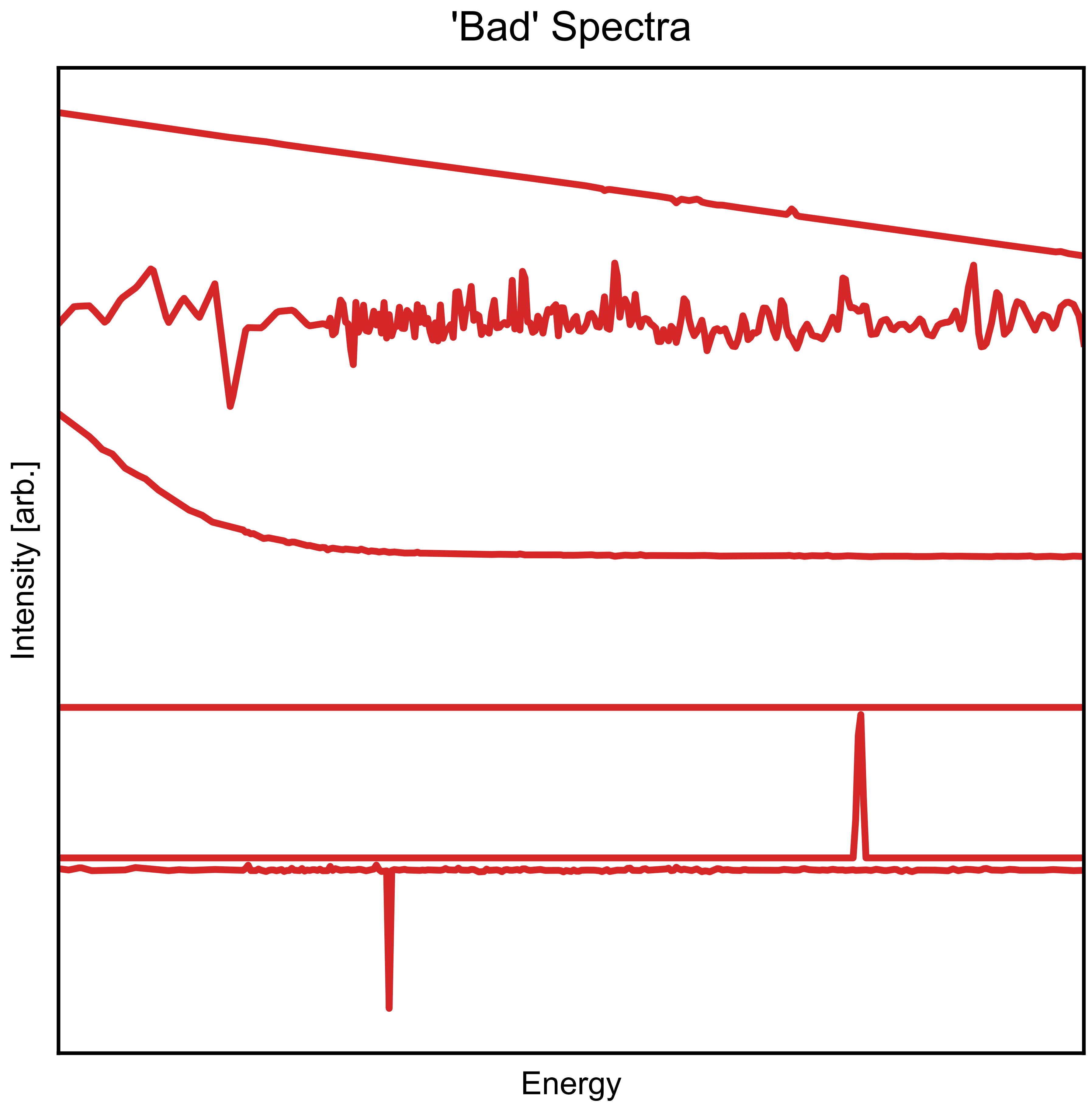


Supervised learning:

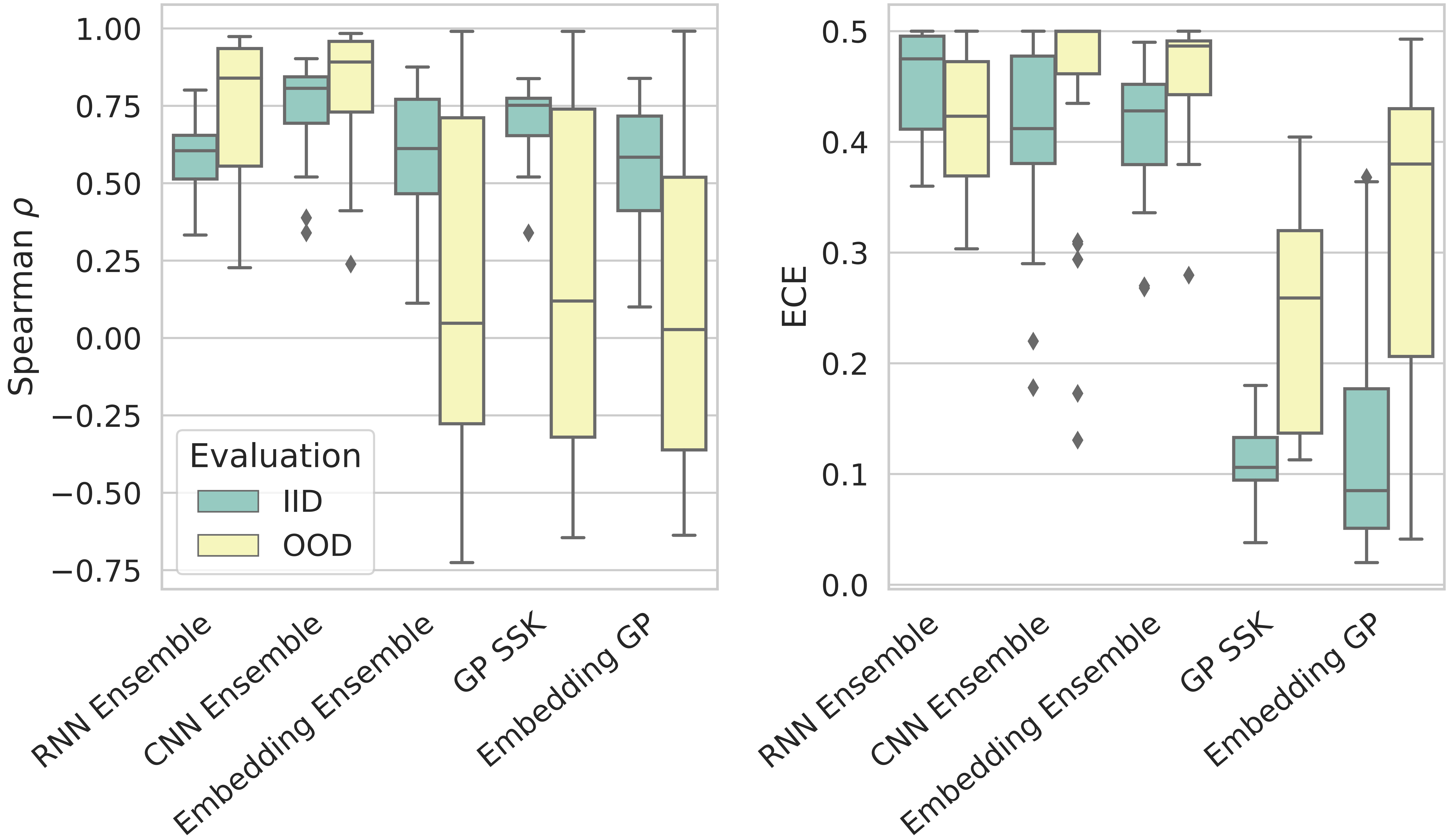
Predicting labels for data when we have–or can create–labeled datasets.

Identifying experimental failures at BMM.



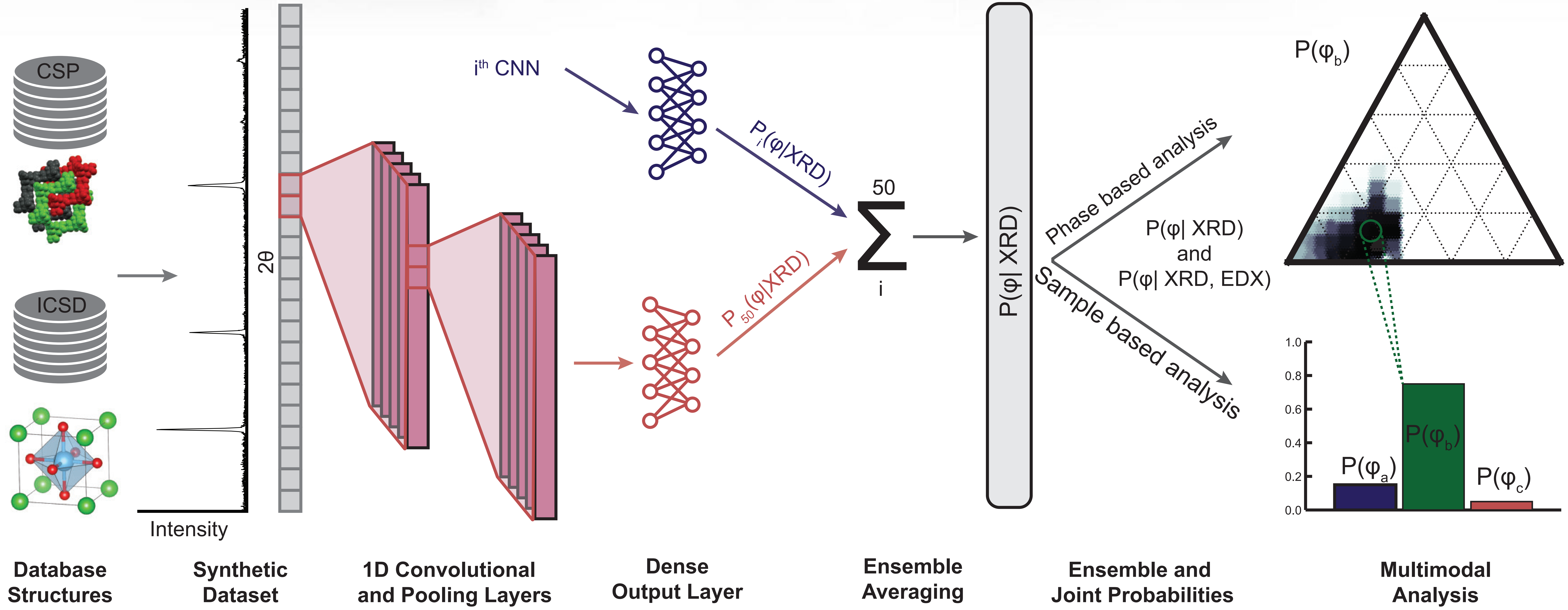


Testing a suite of algorithms and featurizations for describing the functional response of a fluorescent protein.



Gruver, N. *et al., ICML 2020*

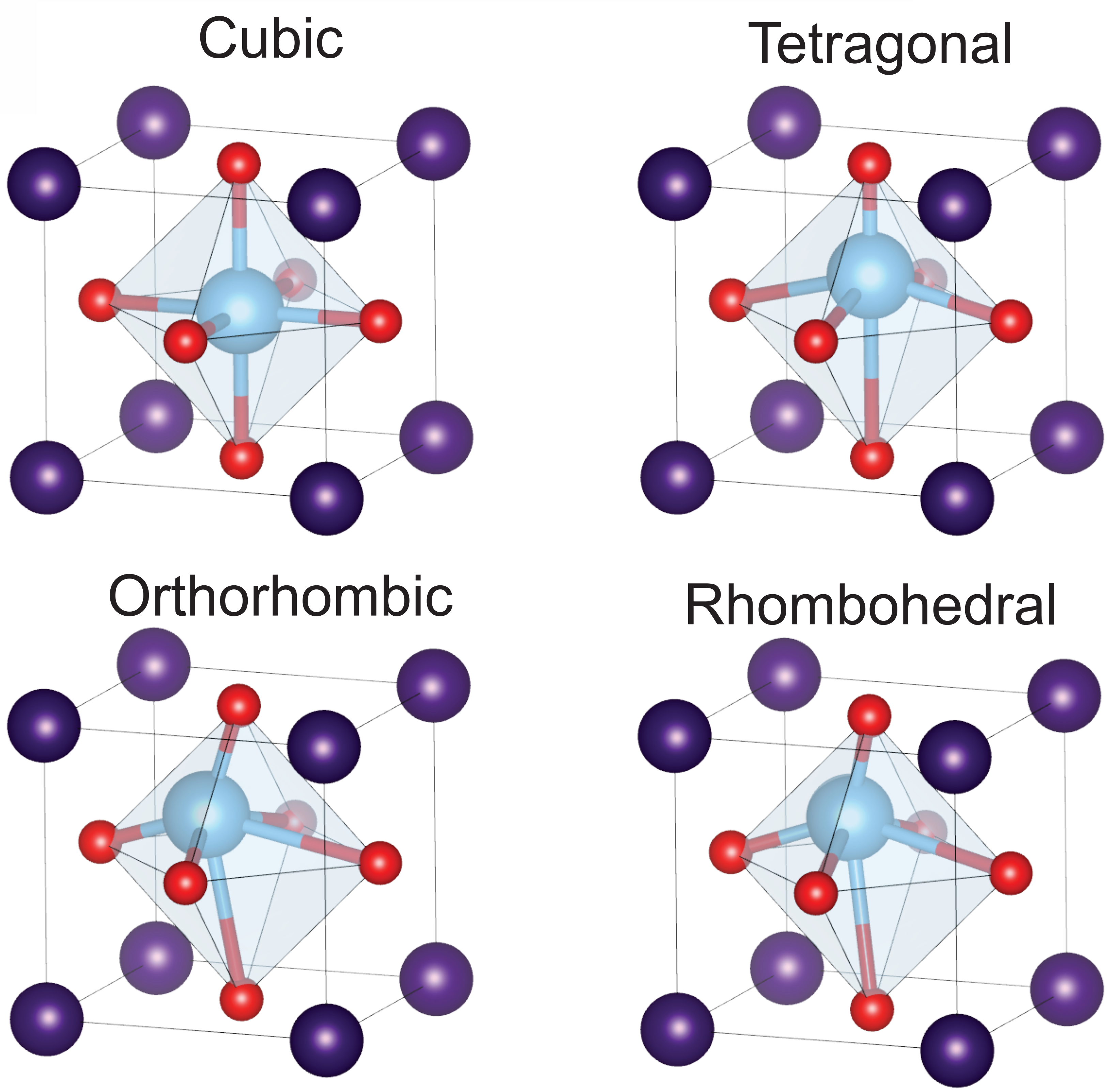
A fully synthetic data pipeline can train an accurate probabilistic model prior to the experiment.

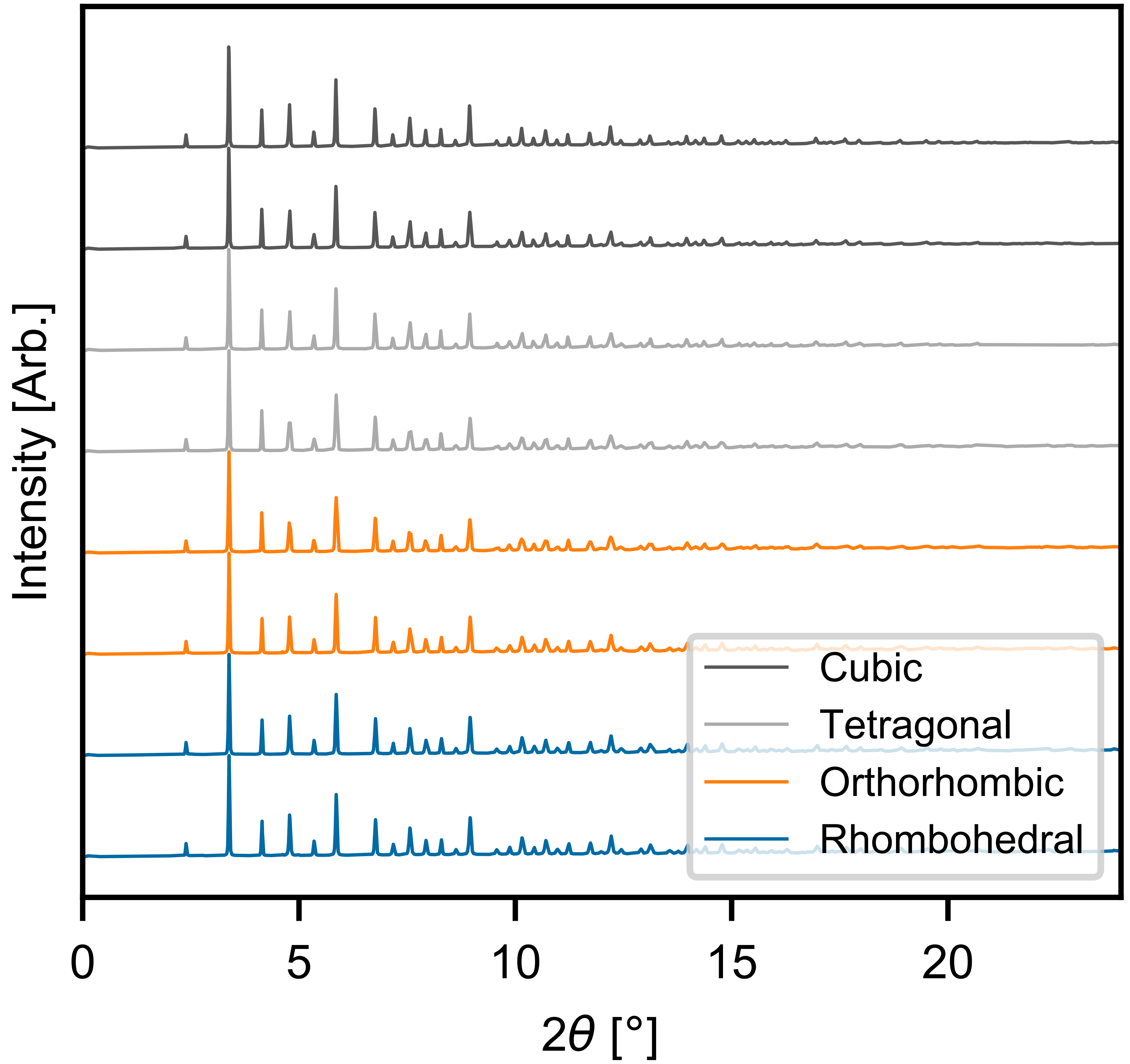


RUB UNIVERSITY OF LIVERPOOL U.S. DEPARTMENT OF ENERGY Office of Science BROOKHAVEN NATIONAL LABORATORY

Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

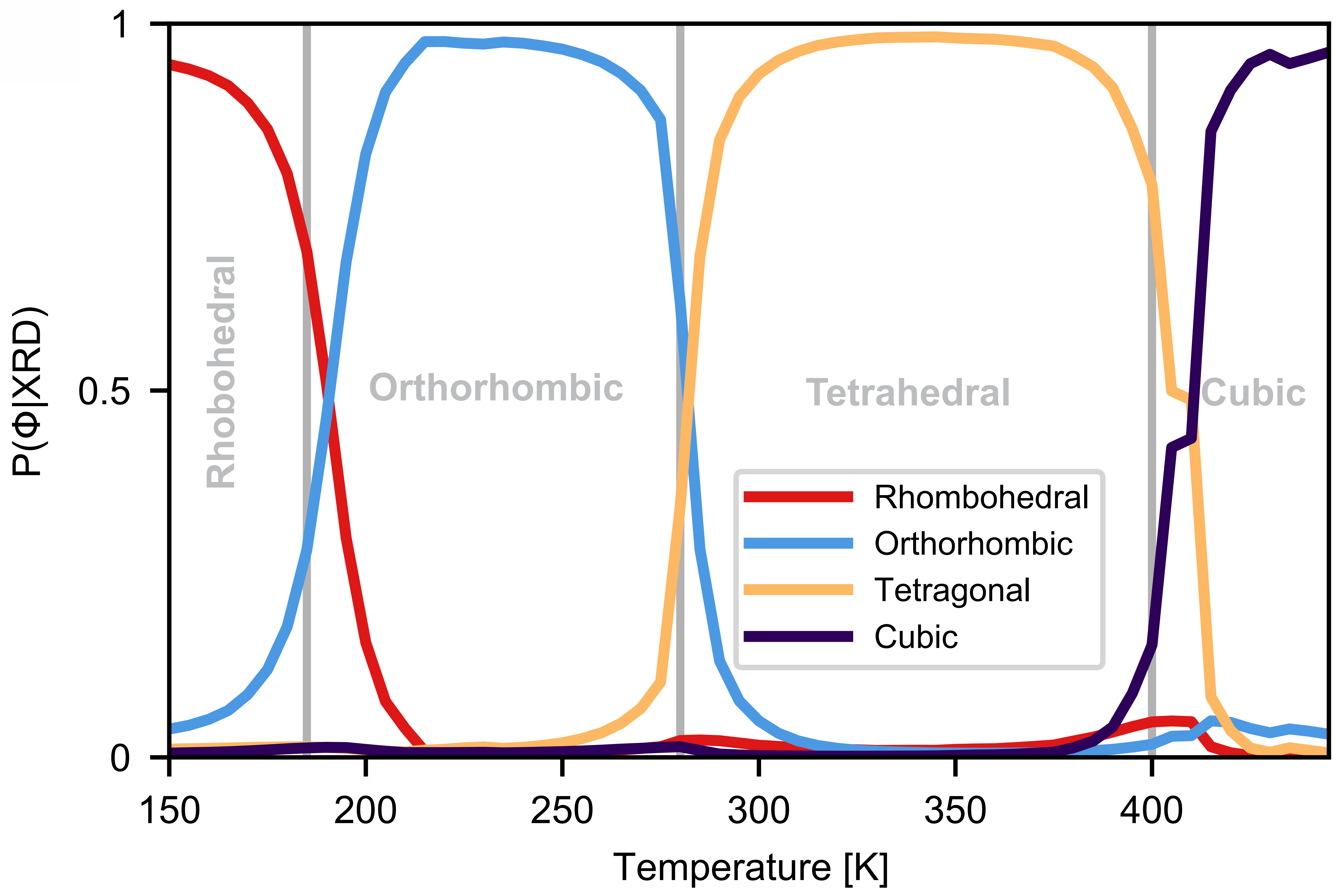
Classifying subtle phase transitions in $BaTiO_3$.





Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

Classifying subtle phase transitions in $BaTiO_3$.

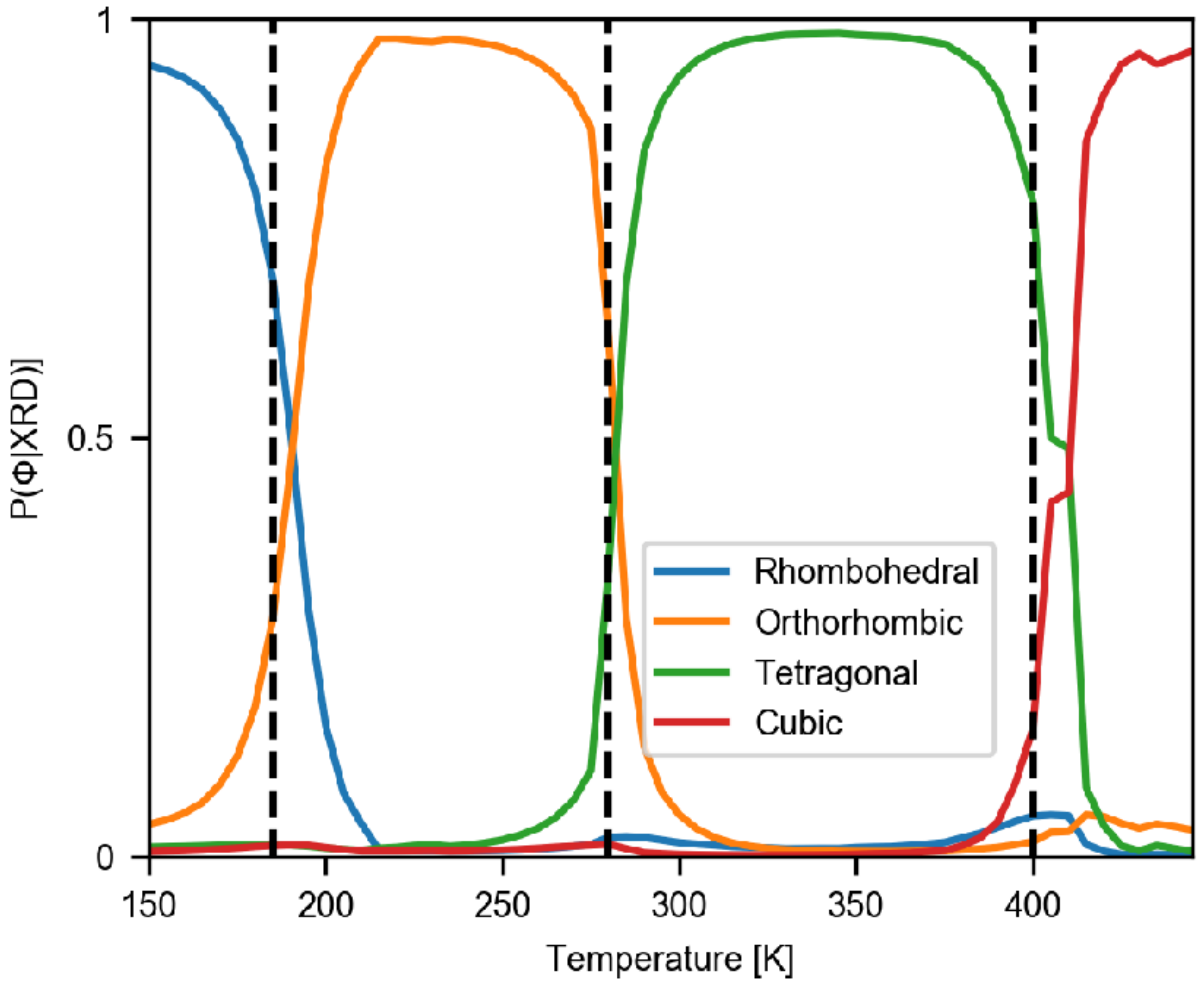


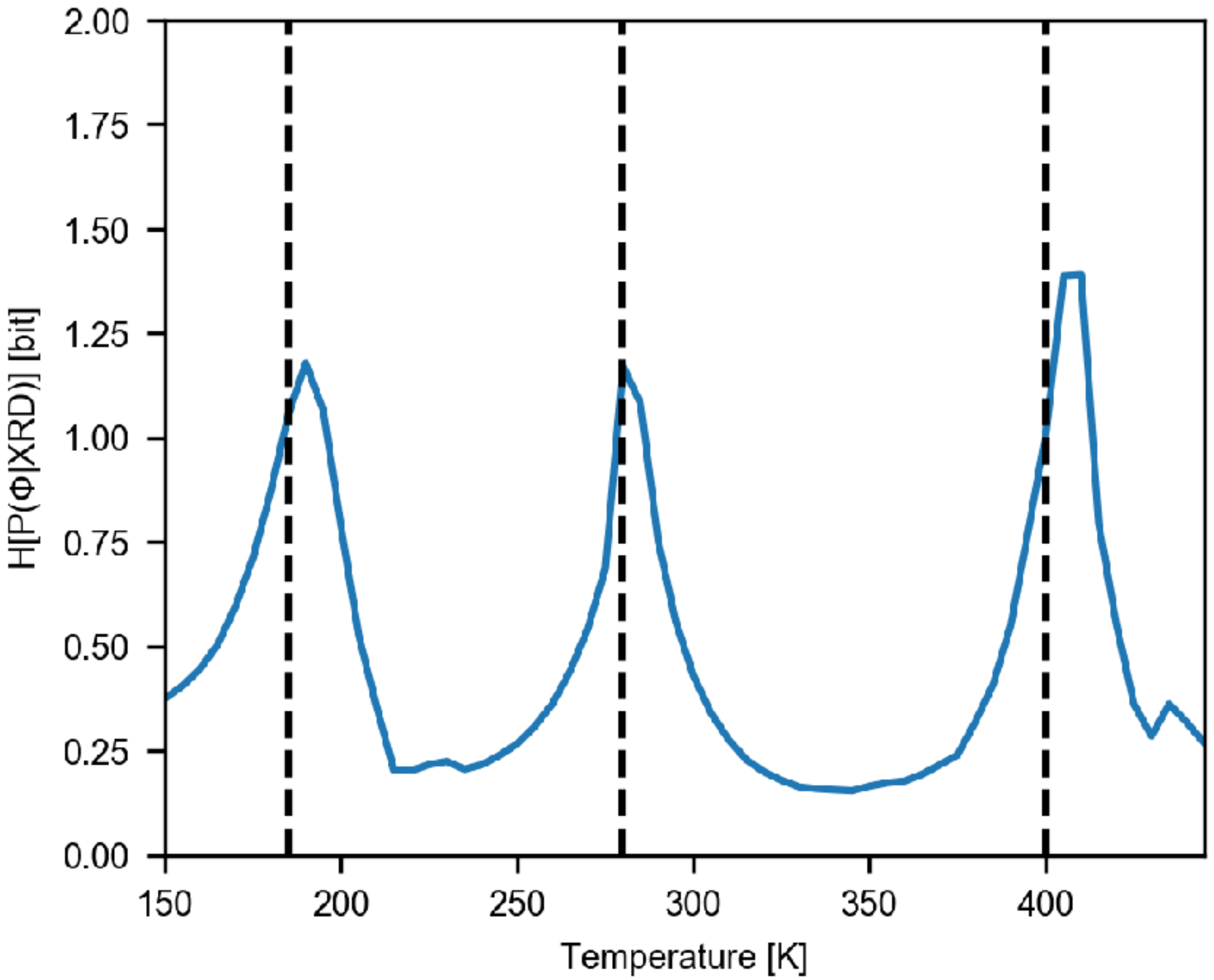
Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

Unsupervised learning:

How do we approach situations when we are exploring the unknown?

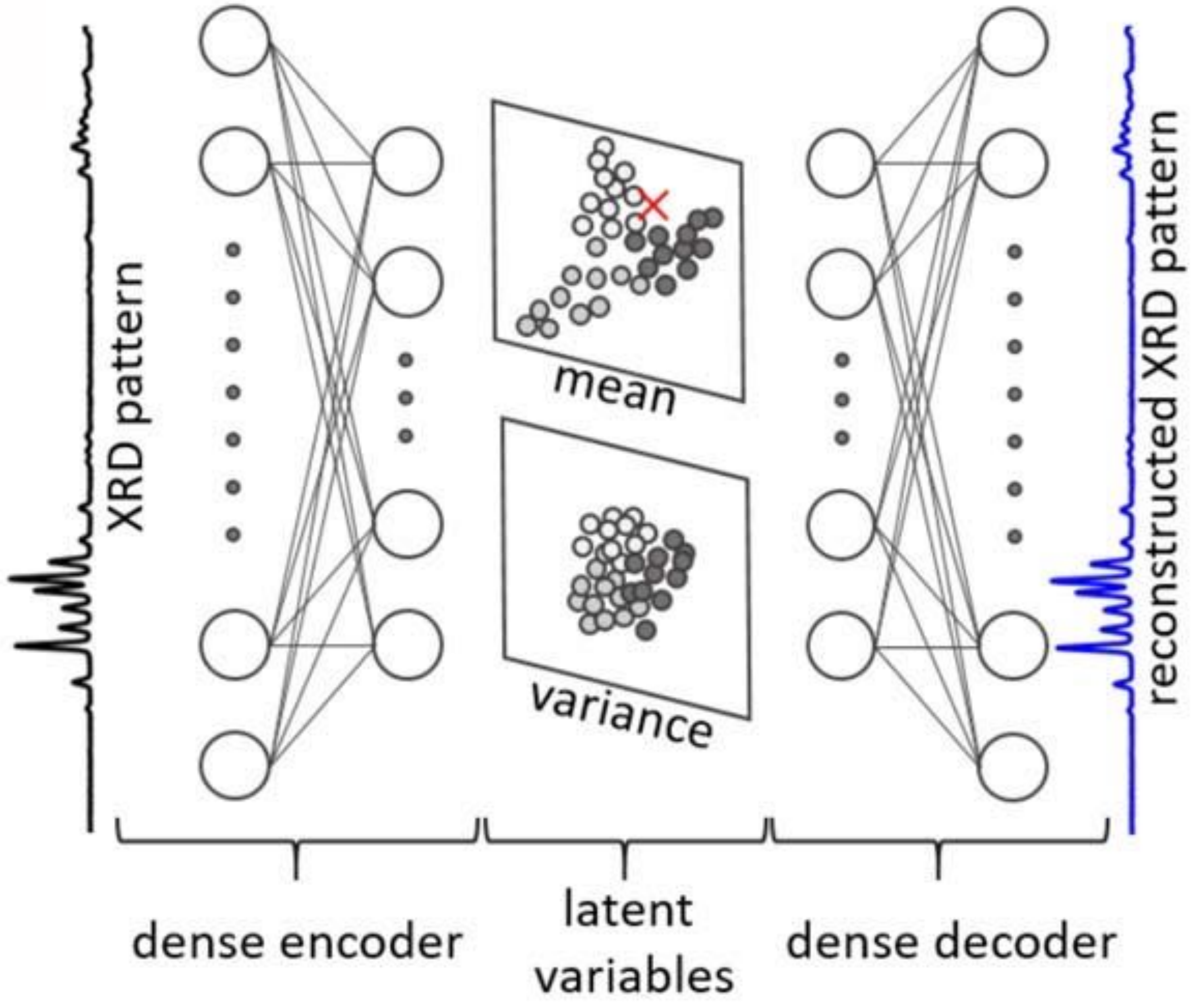
Uncertainty is a proxy for novelty.

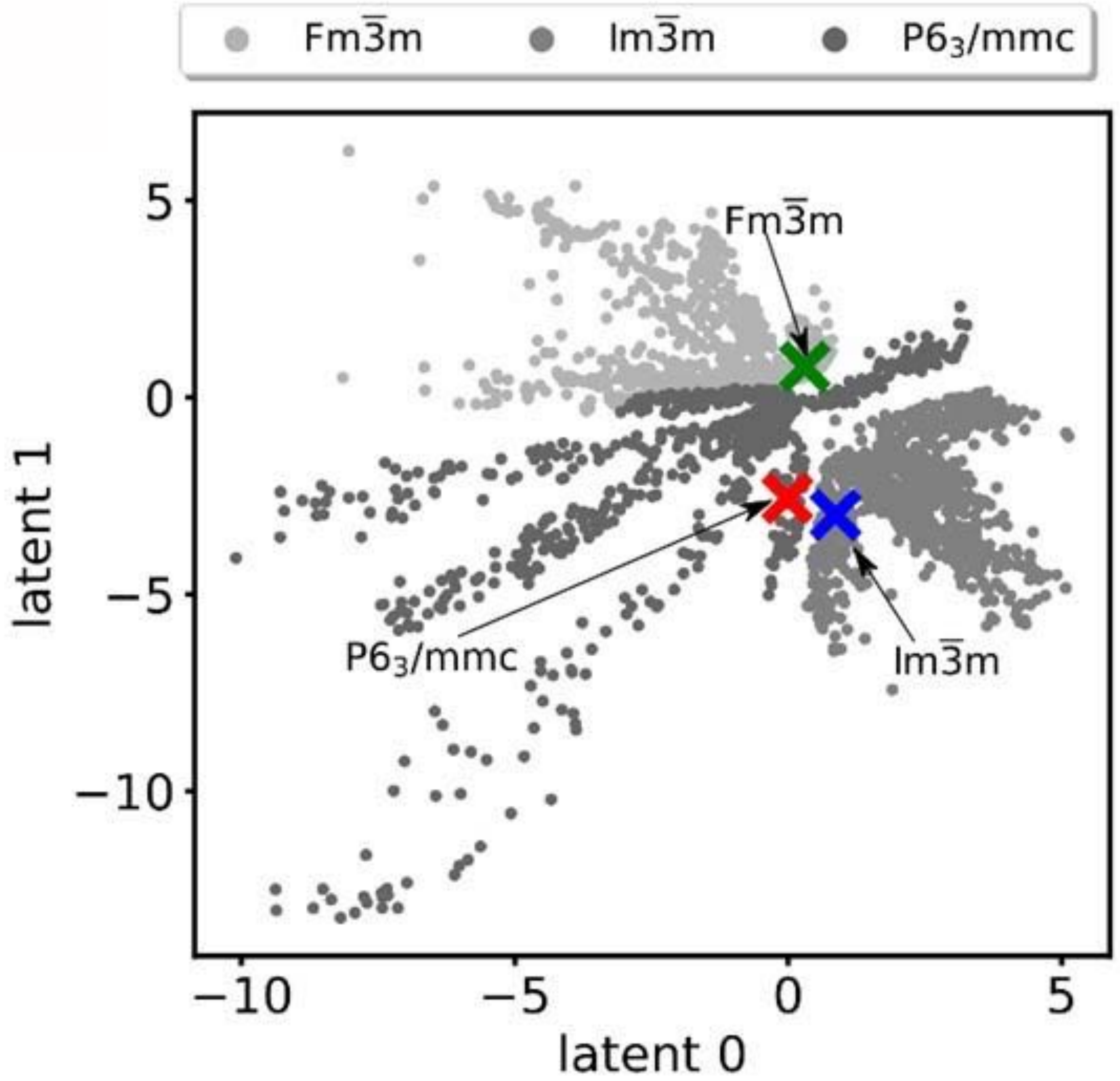




Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

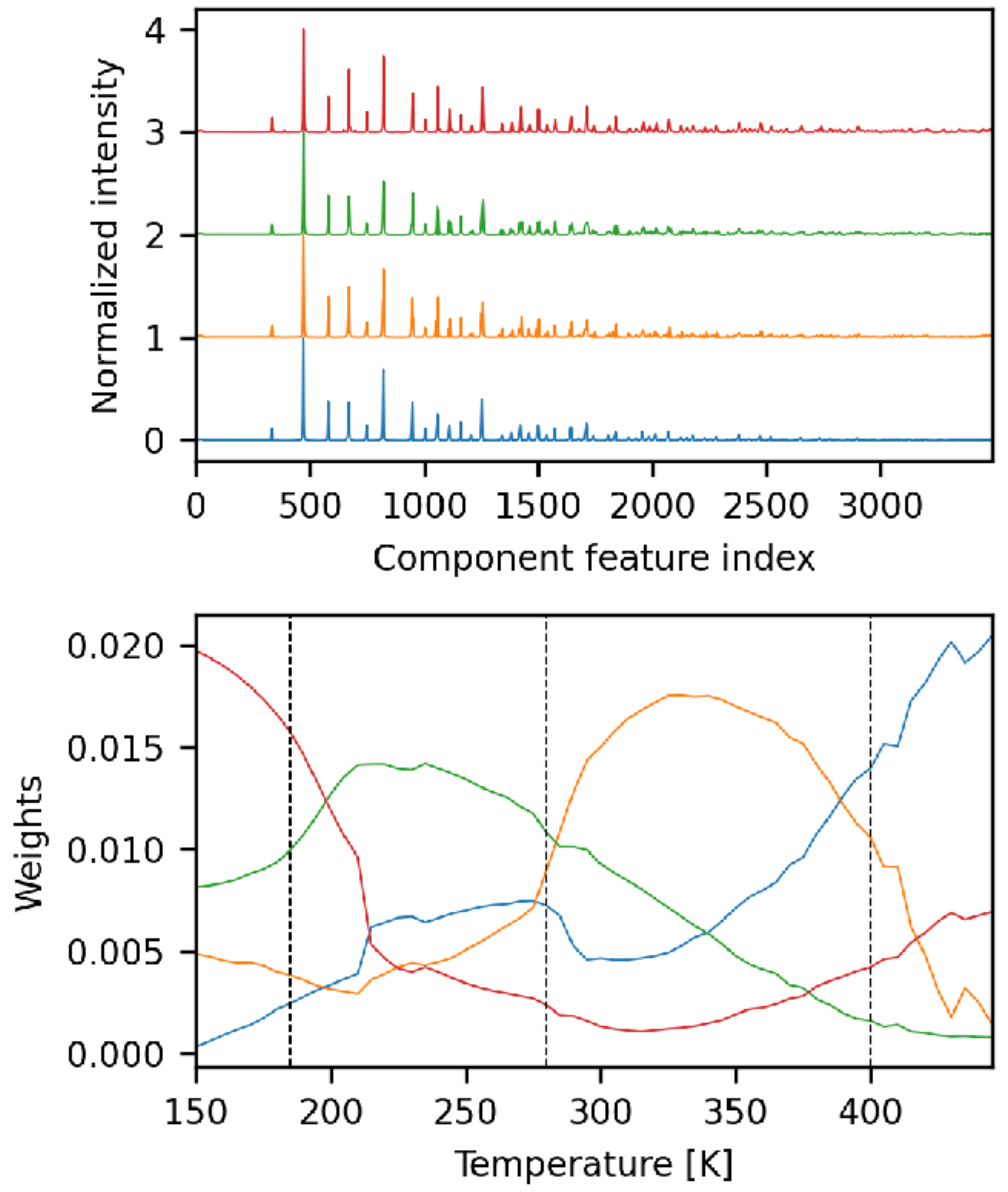
The latent space of variational auto encoders conditioned on the same synthetic dataset is a guide for novelty.





Banko, L. *et al.*, *npj. Comp. Matter.*, ***7***, 104 (2021)

Non-negative matrix factorization (NMF) for decomposing datasets without priors.



$\mathbf{X} \sim \mathbf{WH}$

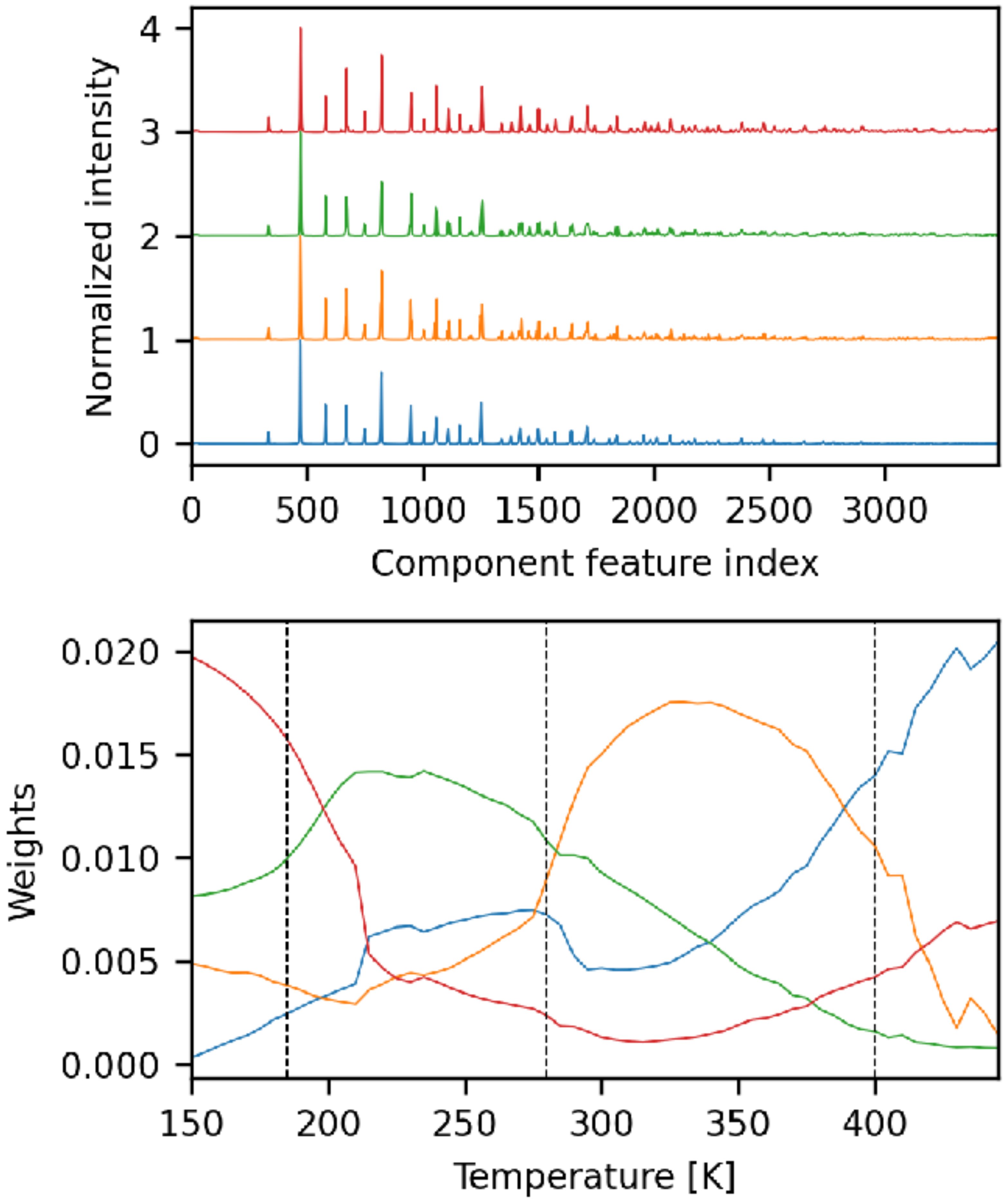
W (m patterns, k components)

H (k components, n features)

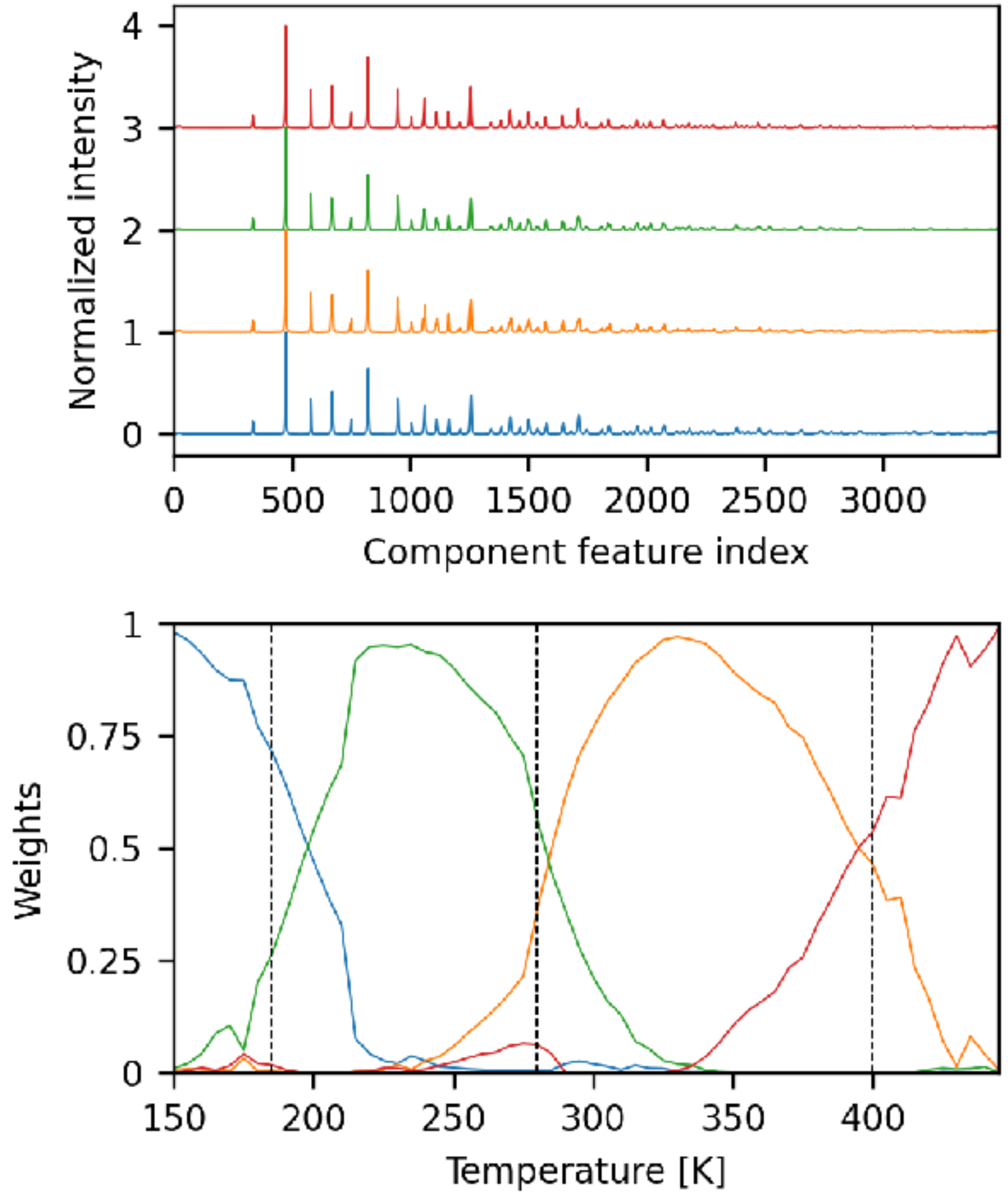
Appl. Phys. Rev. **Accepted** (arXiv:2104.00864)

Constrained NMF produces physically realistic components and weights.

Canonical



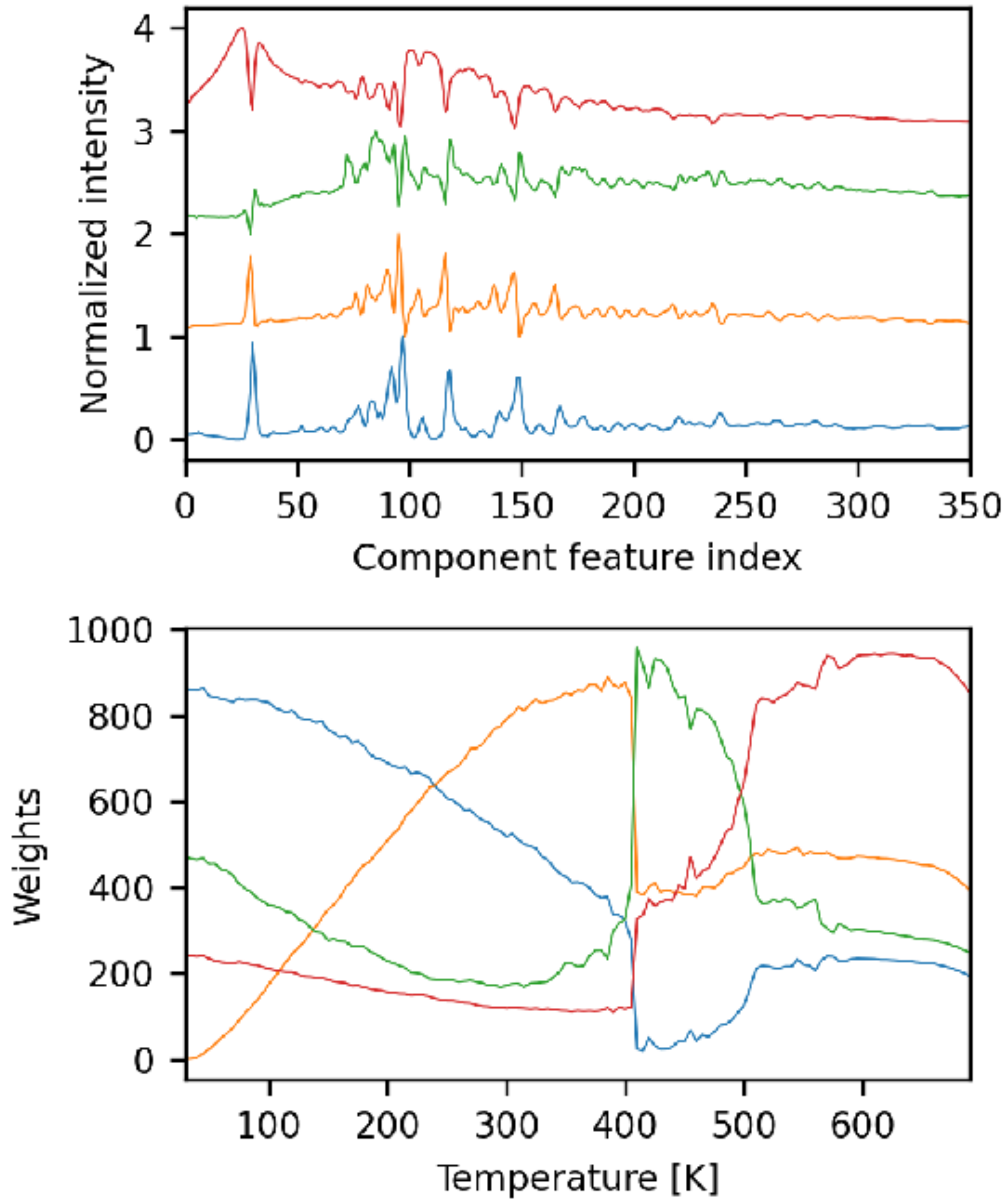
Constrained



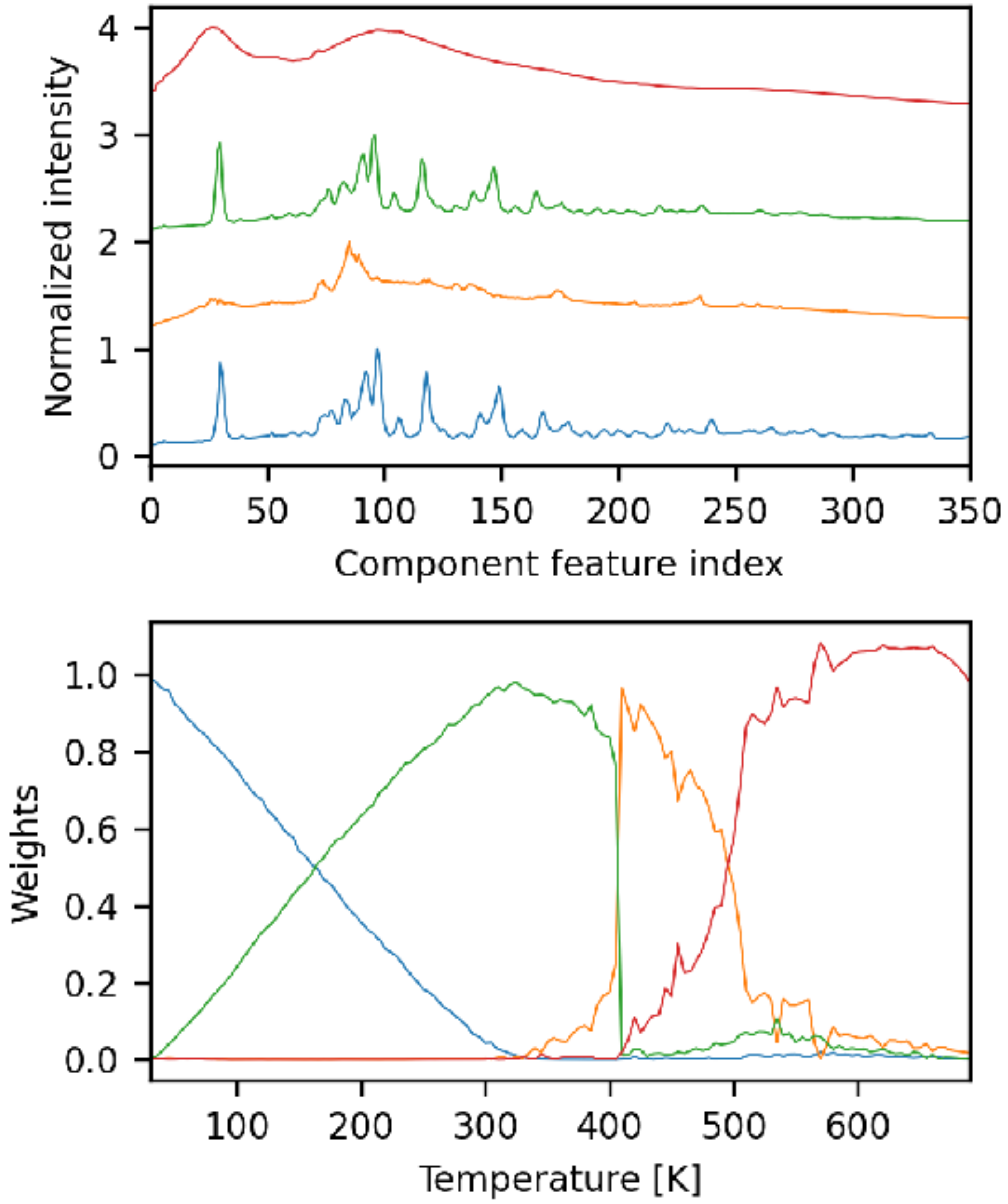
Appl. Phys. Rev. **Accepted** (arXiv:2104.00864)

Dynamically adjusted constraints leads to directly interpretable decompositions

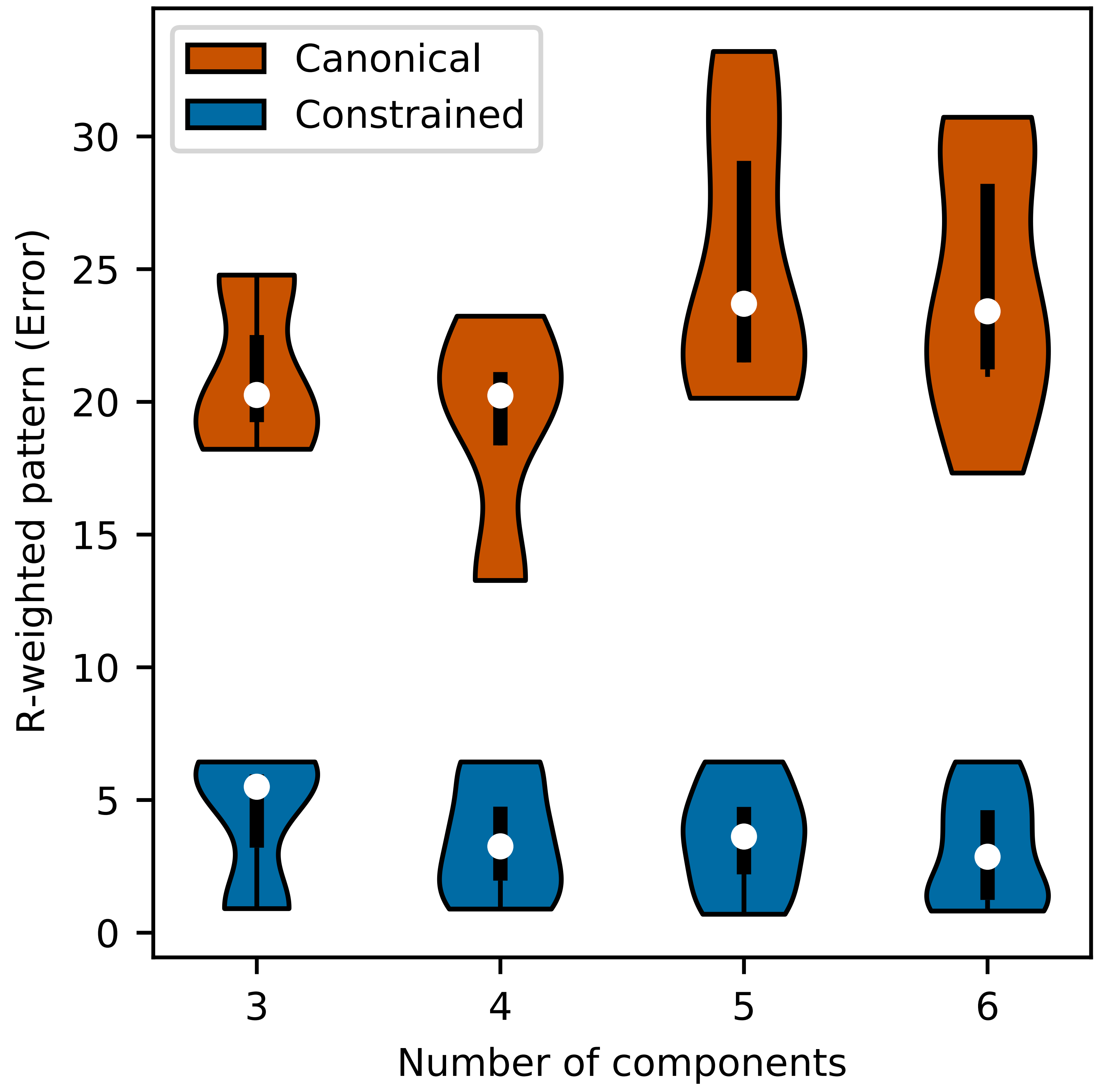
Canonical



Constrained



Refinement Results



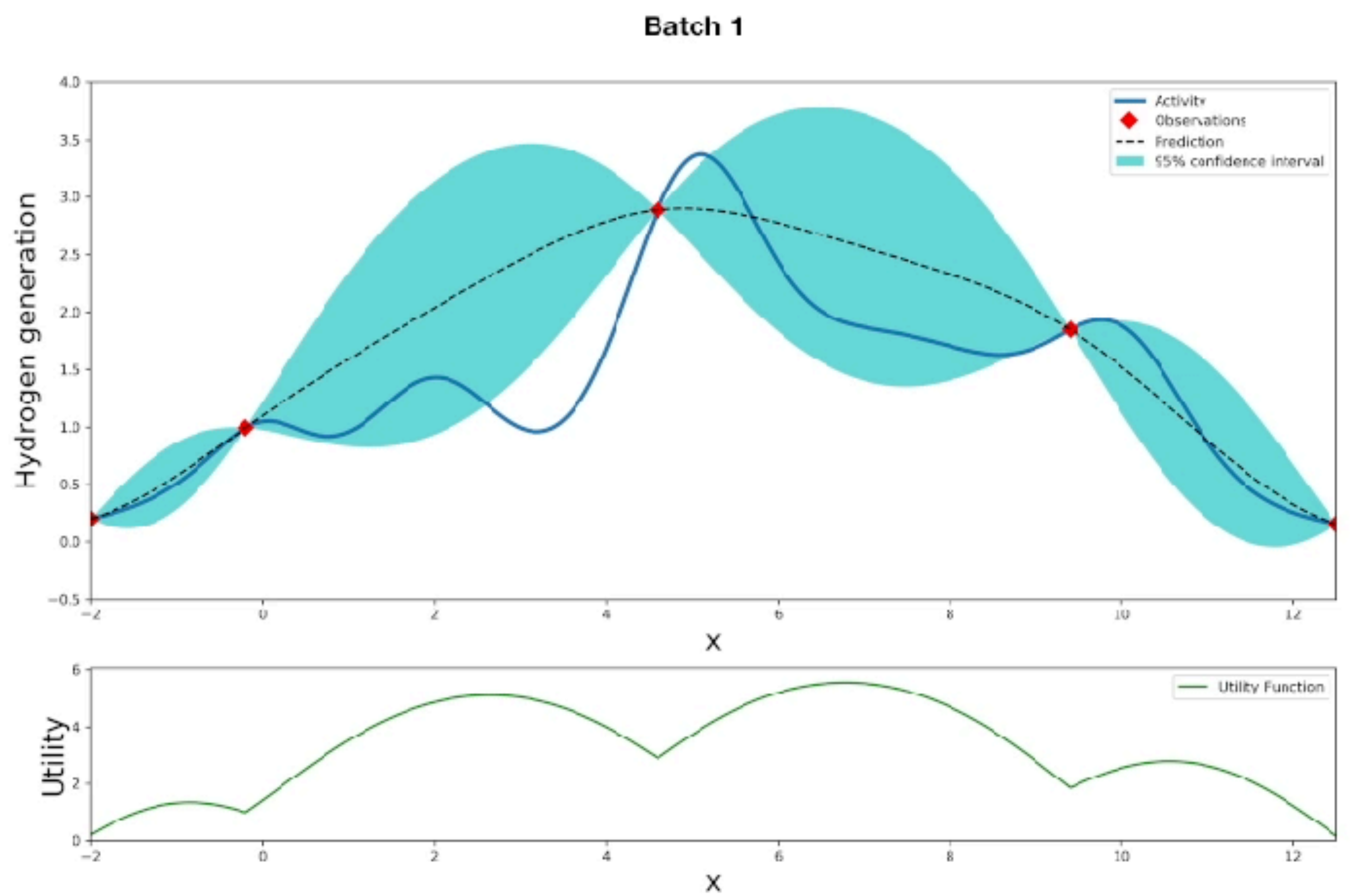
Appl. Phys. Rev. **Accepted** (arXiv:2104.00864)

Making decisions on what to measure next:

Active learning for exploring phase space

Reinforcement learning for operating under resource constraints

Bayesian optimization to guide experiments.



1. Prescribe a prior belief (Gaussian).
2. Calculate the posterior probability.
3. Use an acquisition function based on the posterior.
4. Sample the acquisition function according to the batch size and greed.

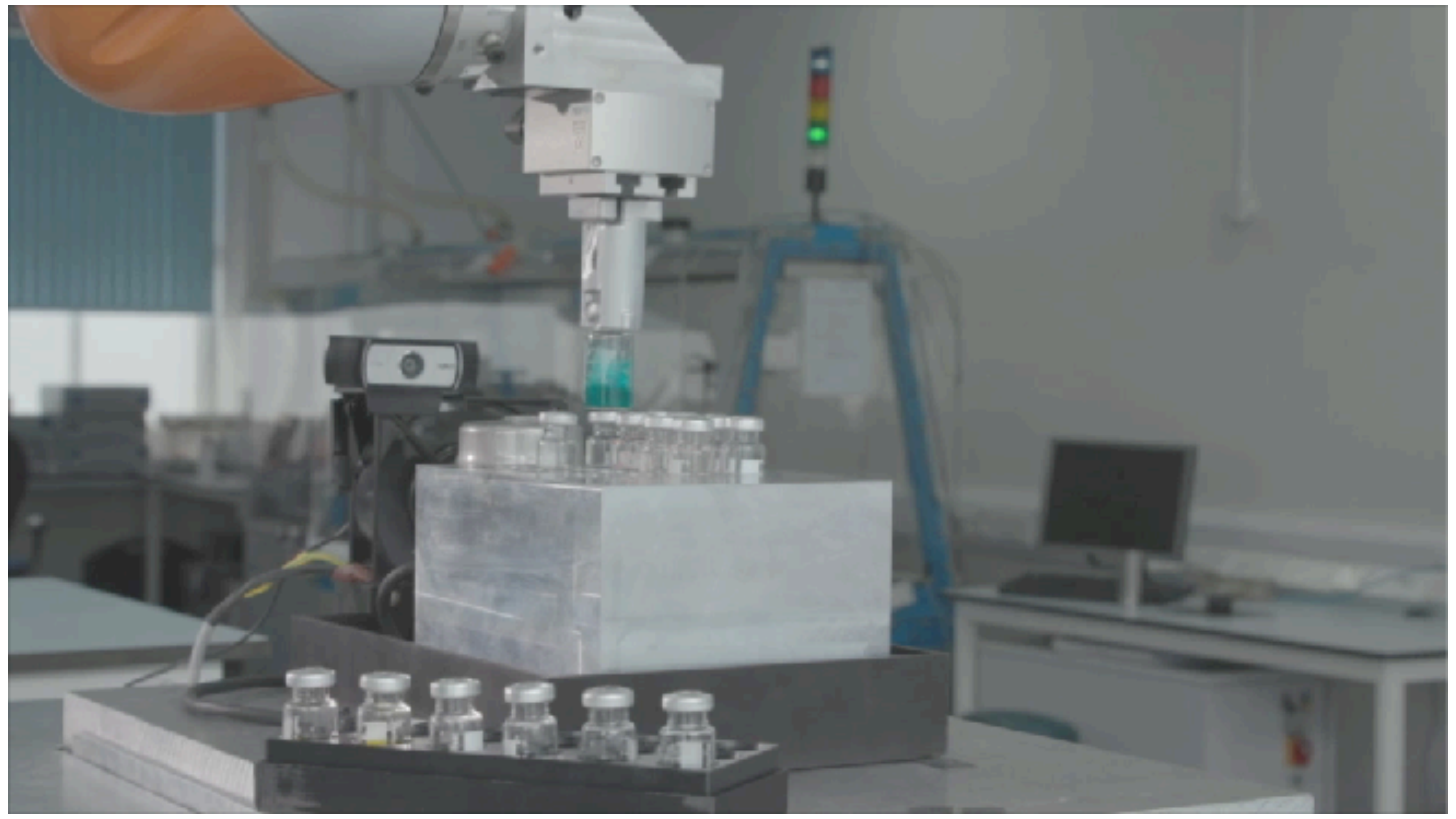
Nature **583**, 237-241 (2020)

Autonomous discovery

Solid dispensing

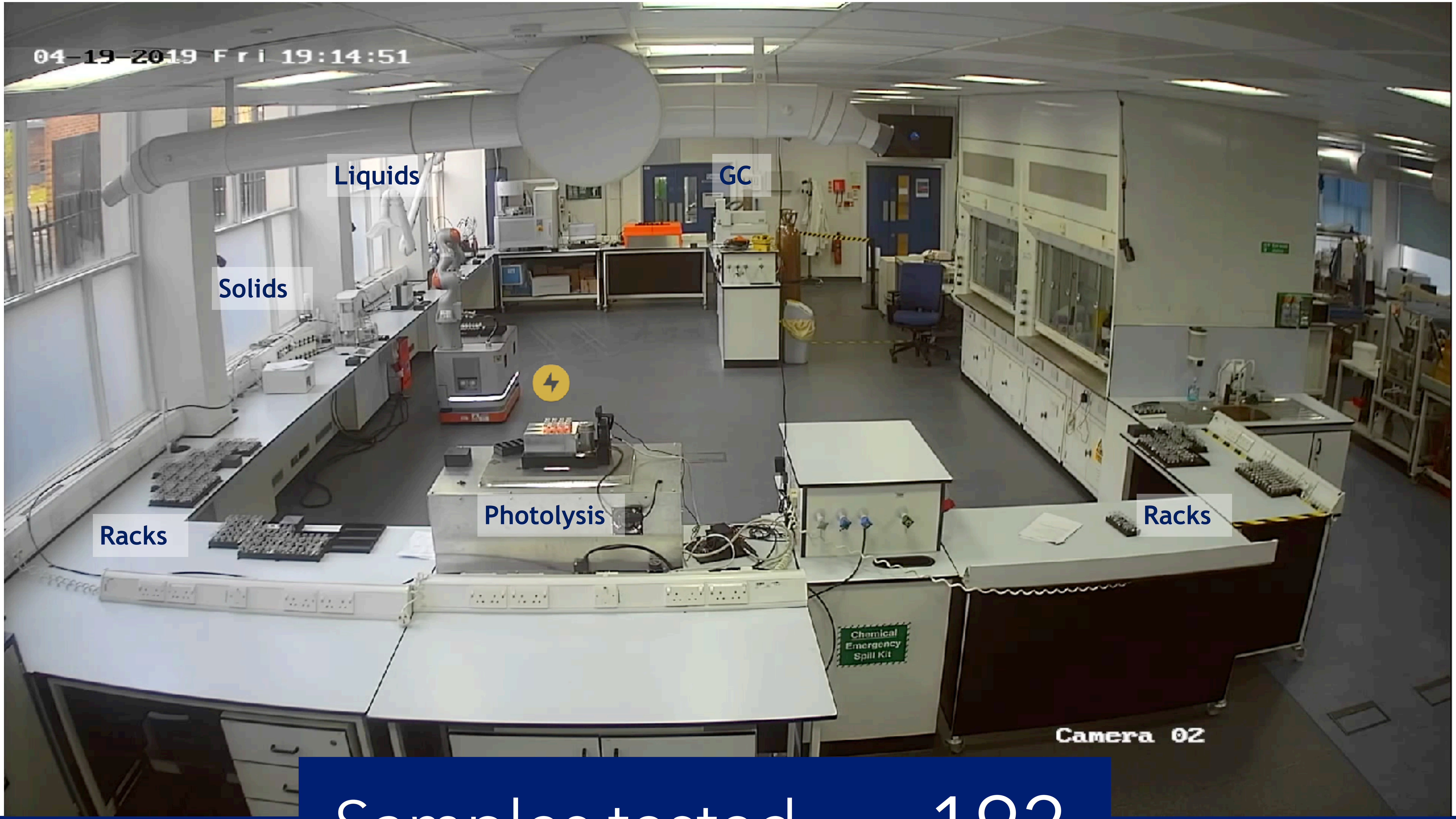
Liquid dispensing
Inertization

Photolysis



Measurement

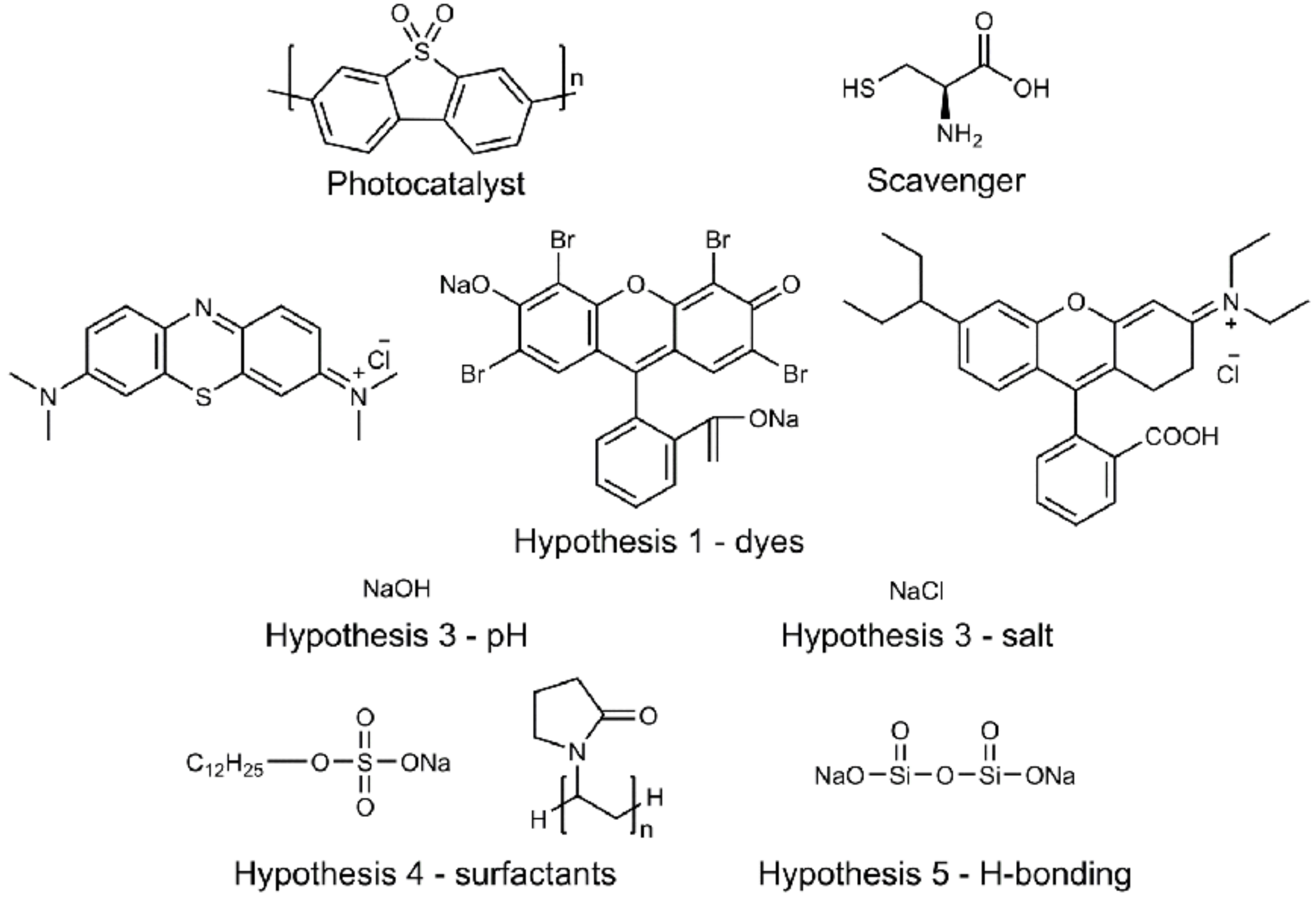
48 hours of research in 60 seconds

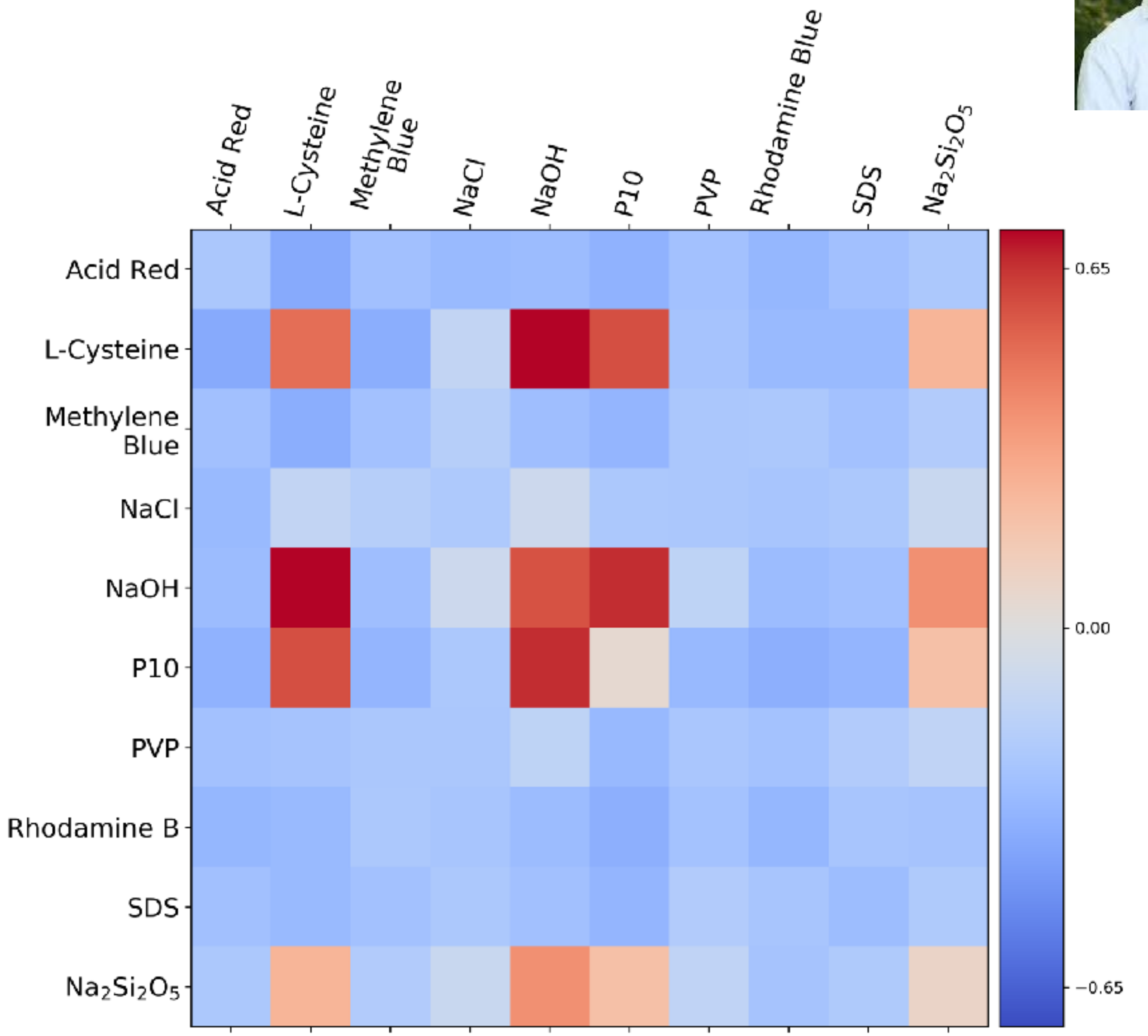


Samples tested = 192

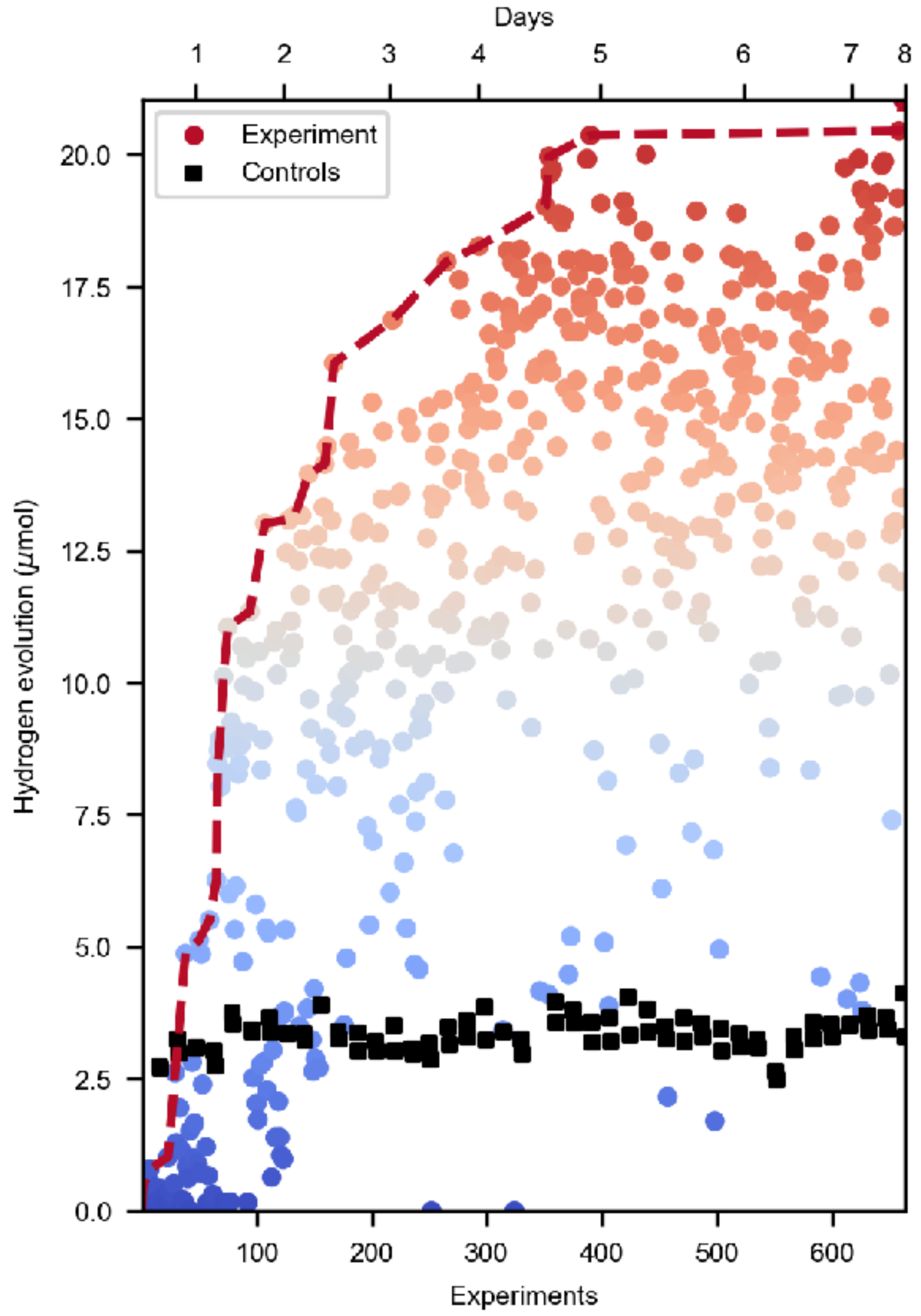
Nature **583**, 237-241 (2020).

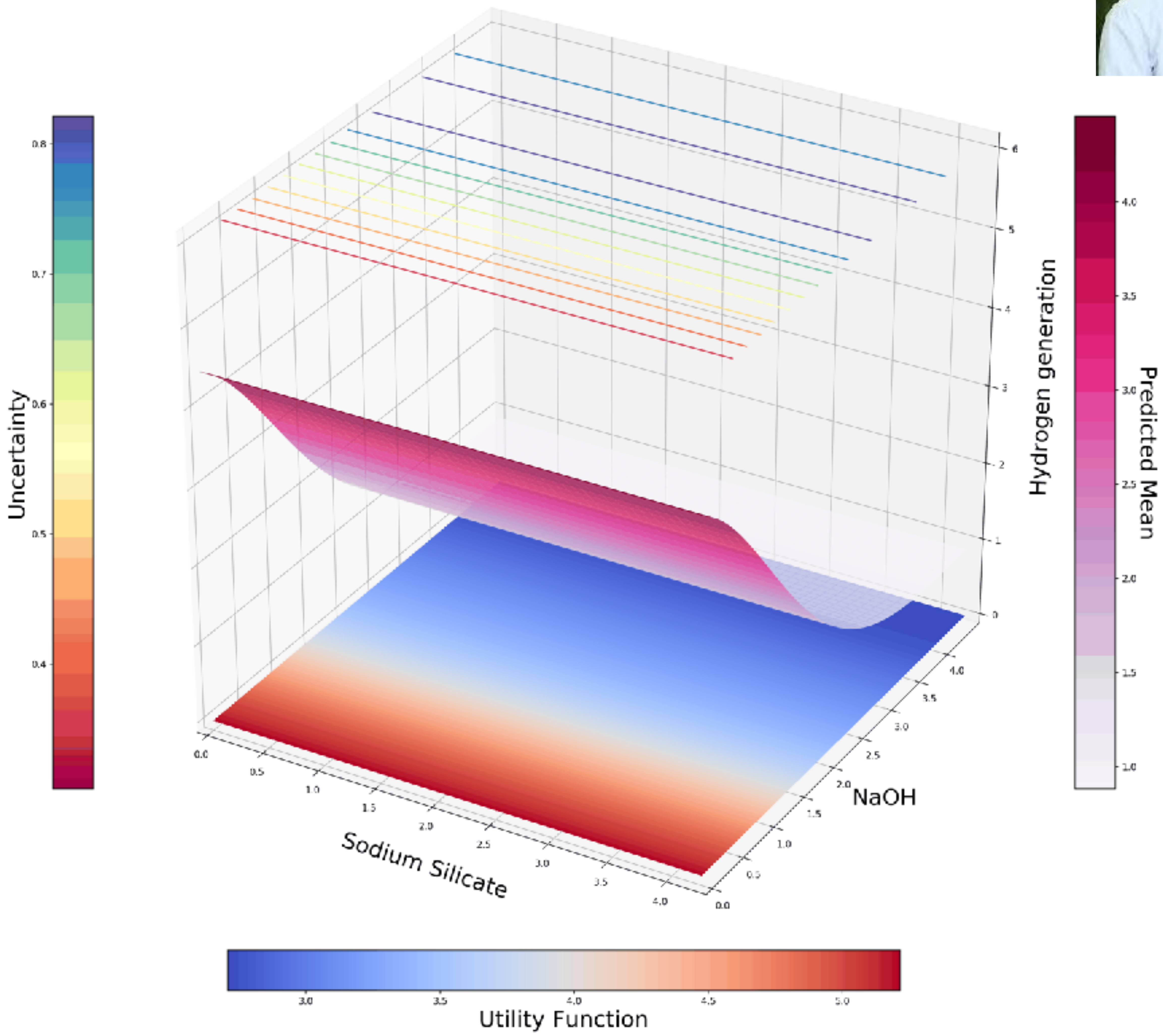
Human defined experiments ran by robot researchers



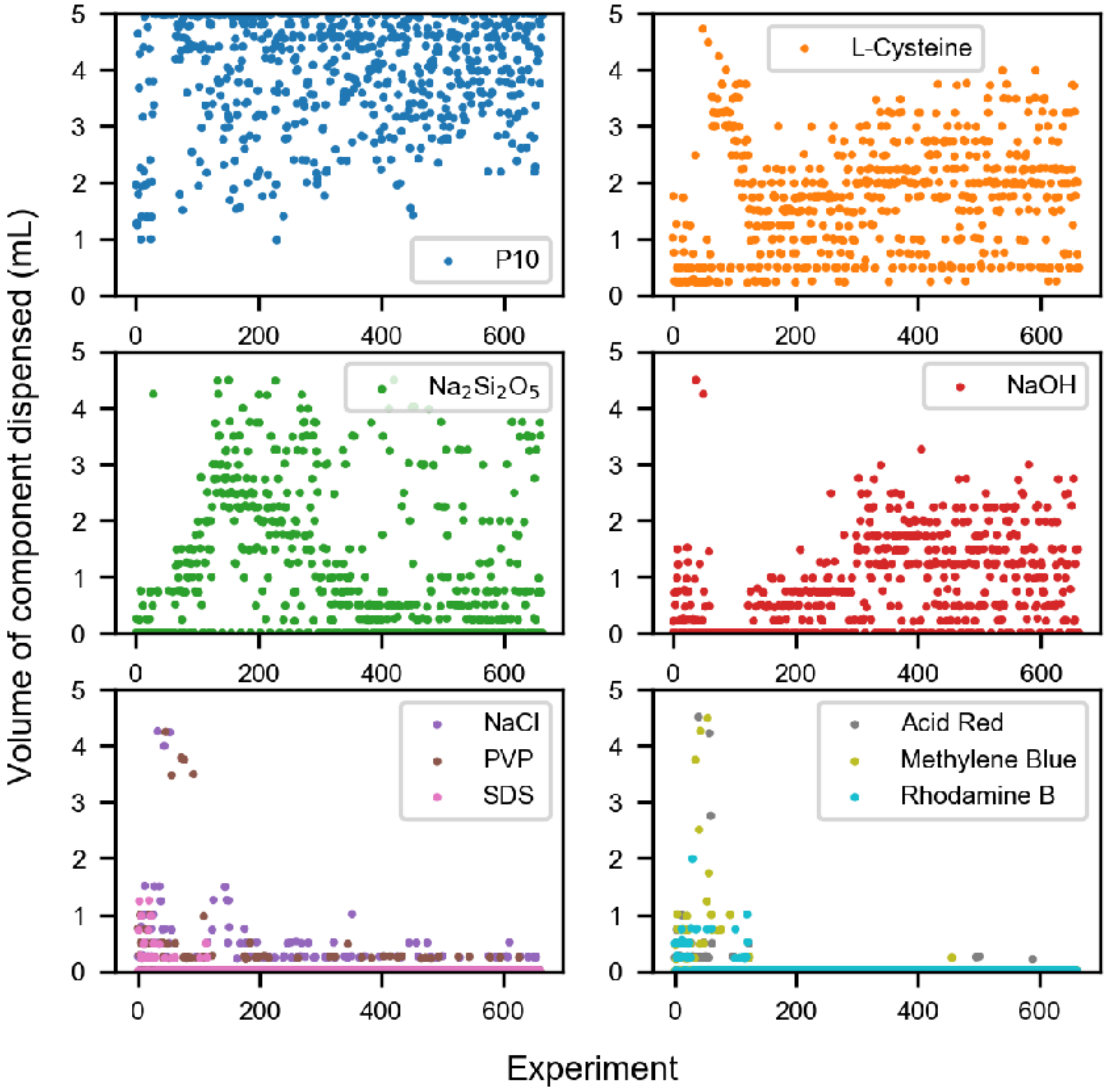


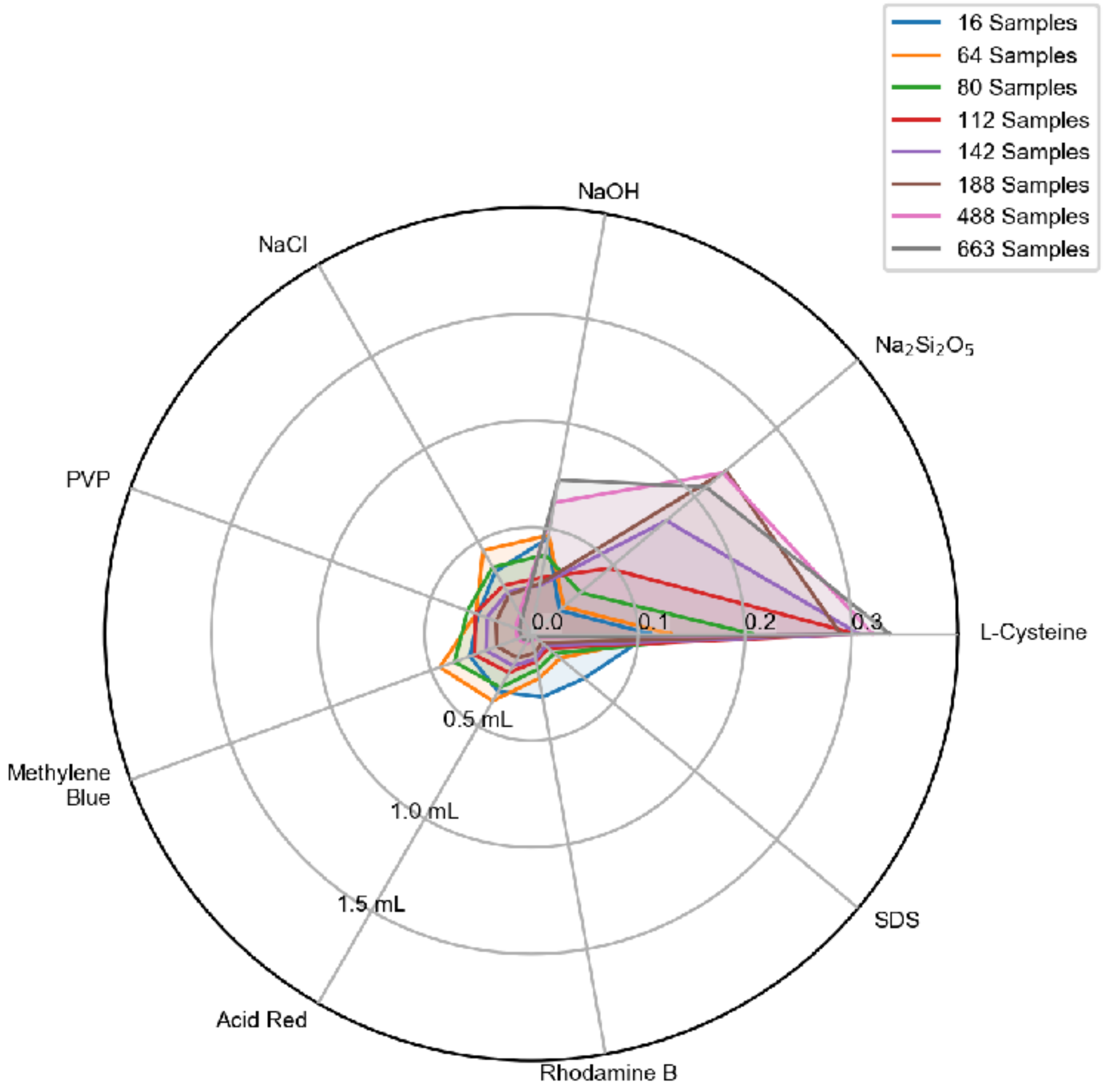
Models develop over time and balance exploration and exploitation





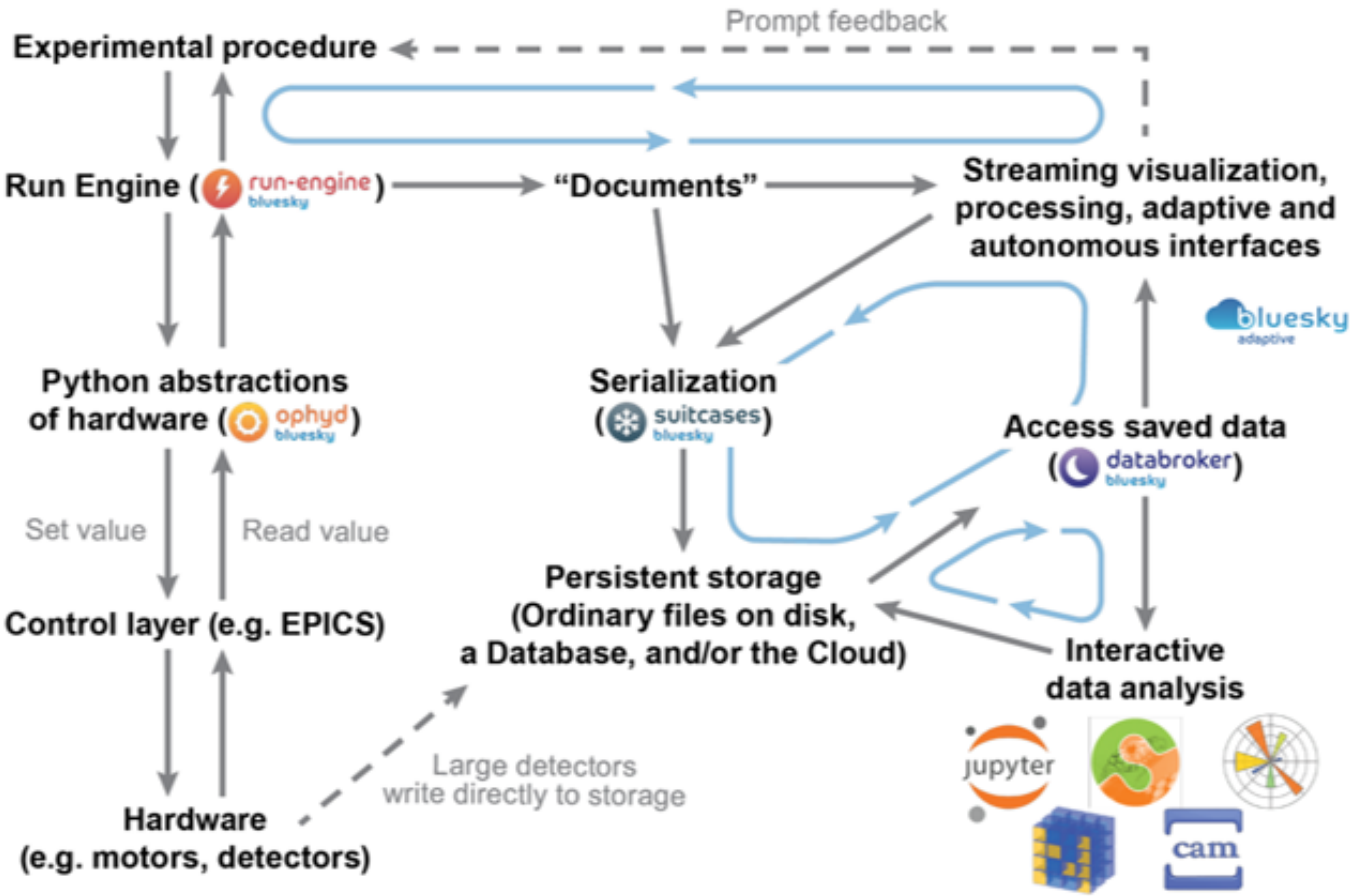
Important components are automatically selected

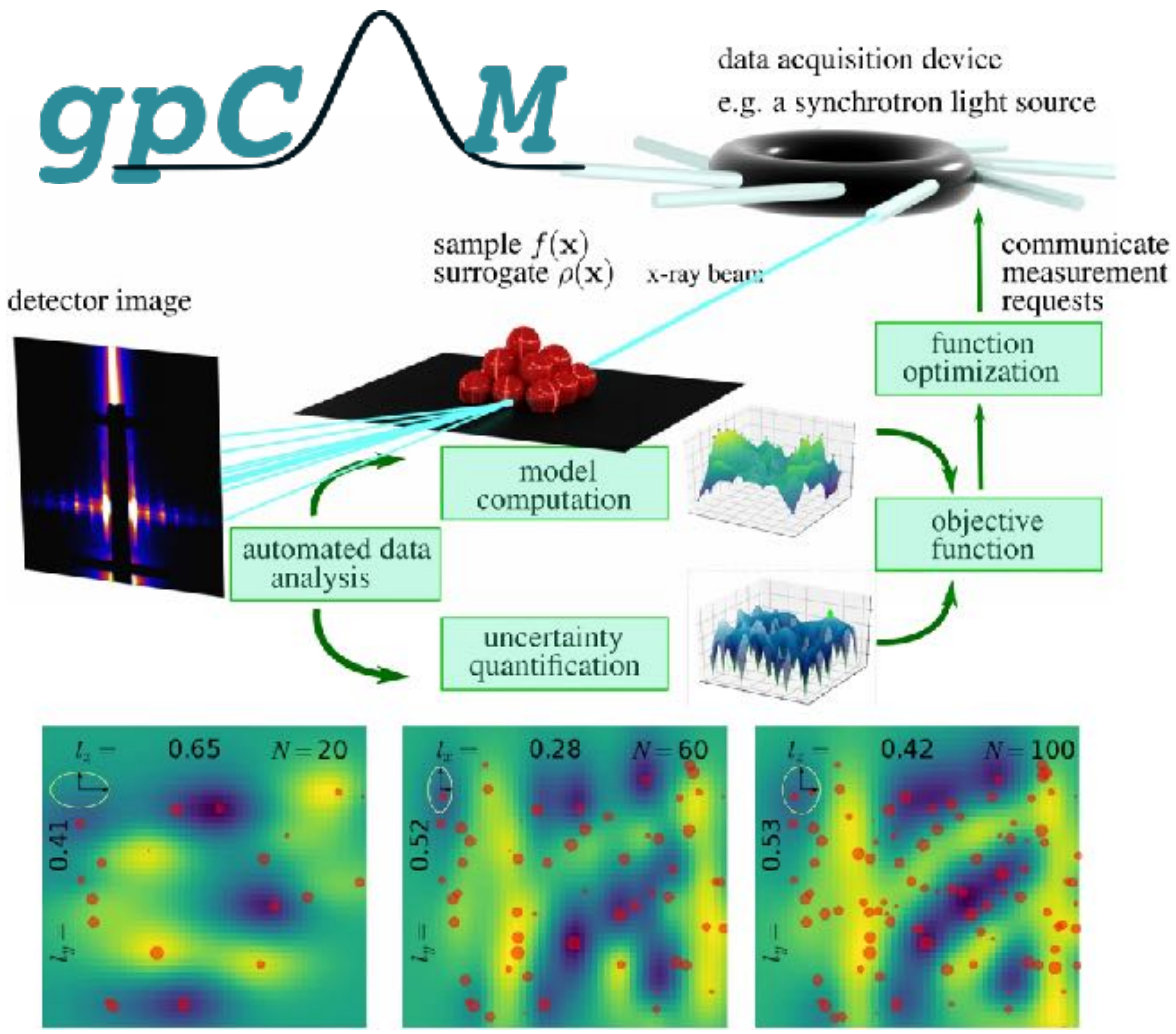




UNIVERSITY OF LIVERPOOL

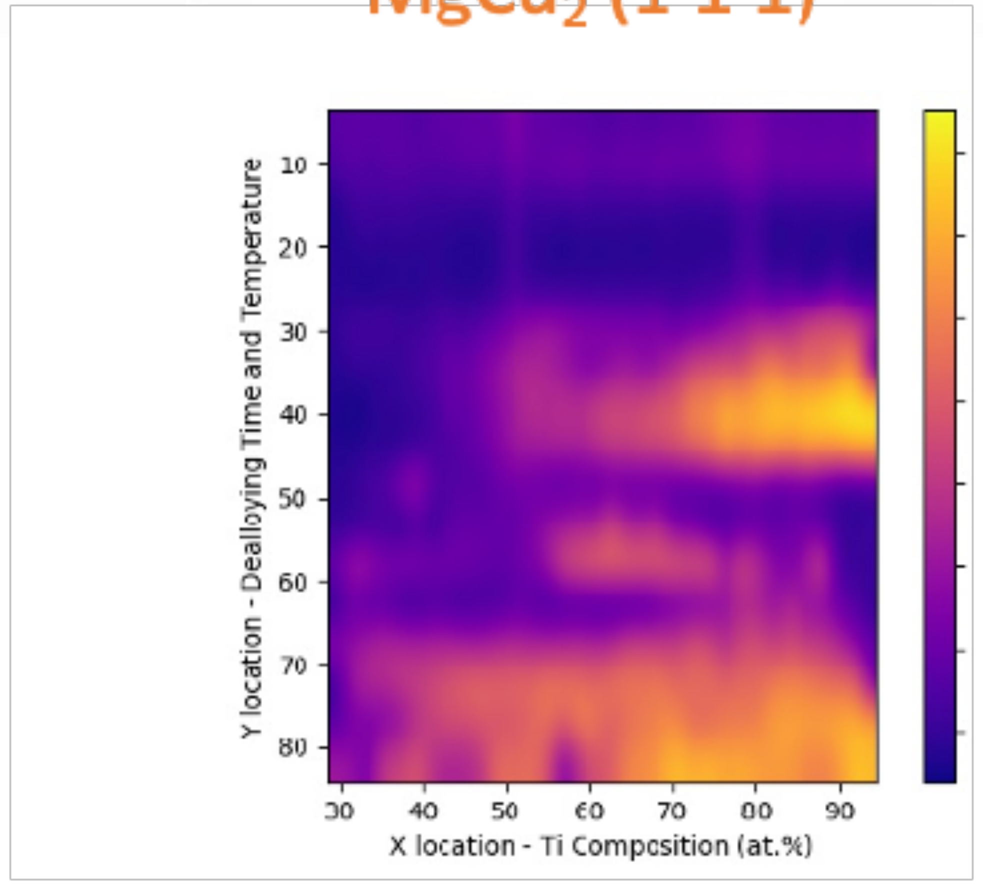
Bayesian optimization for autonomous characterization is enabled by bluesky-adaptive.



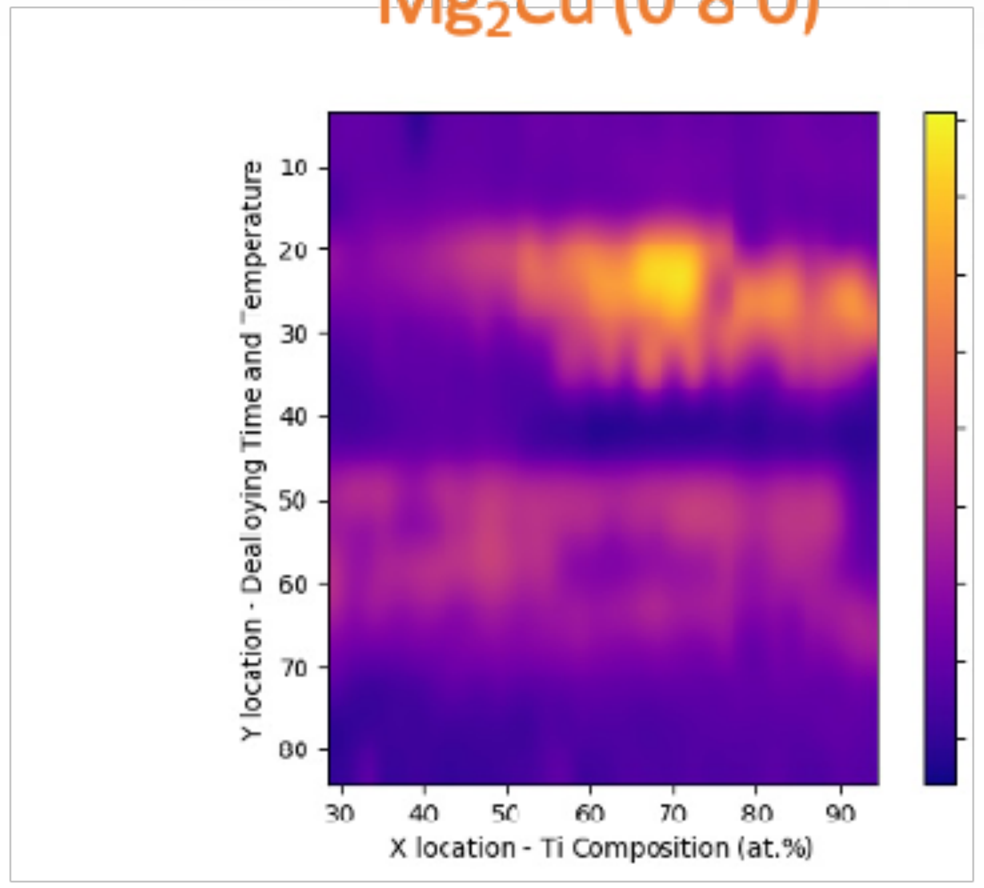


Probabilistic predictions from supervised models can also guide effective experimentation.

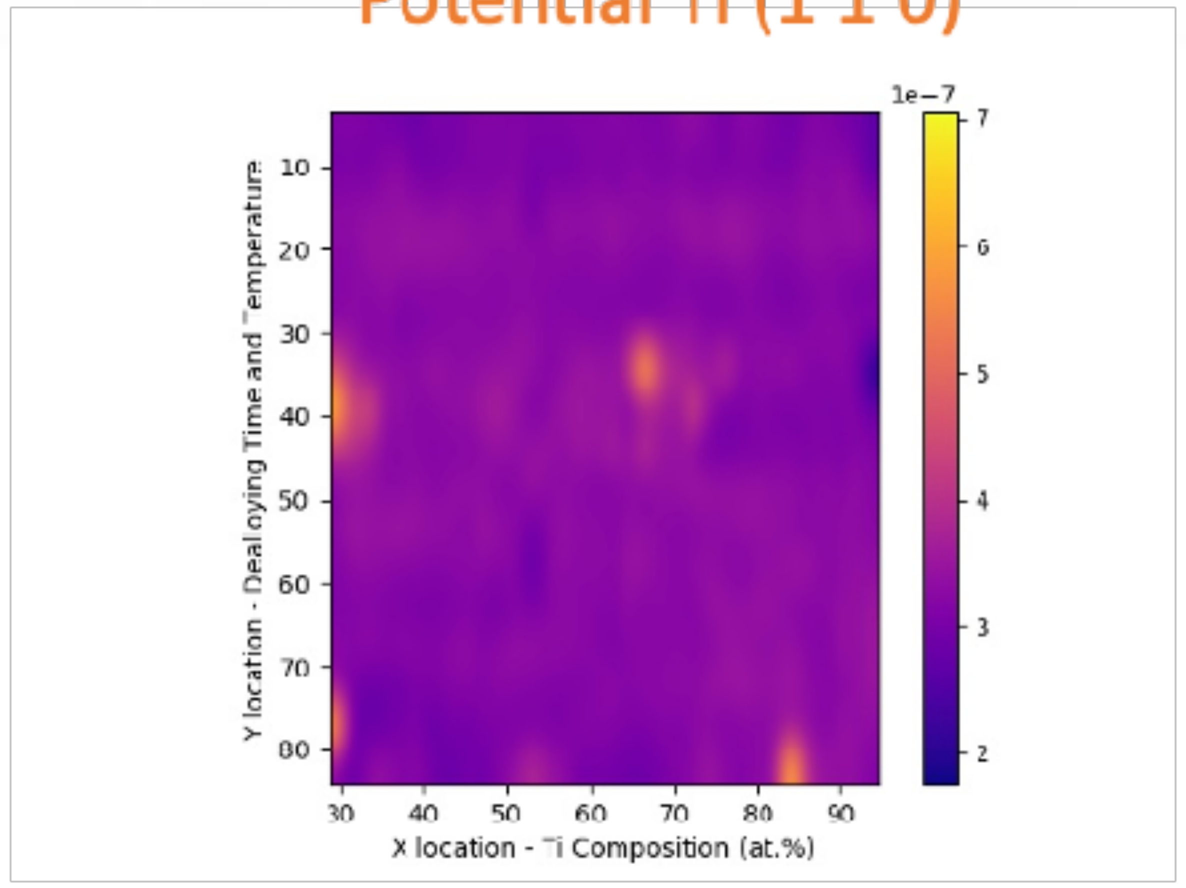
$MgCu_2$ (1 1 1)

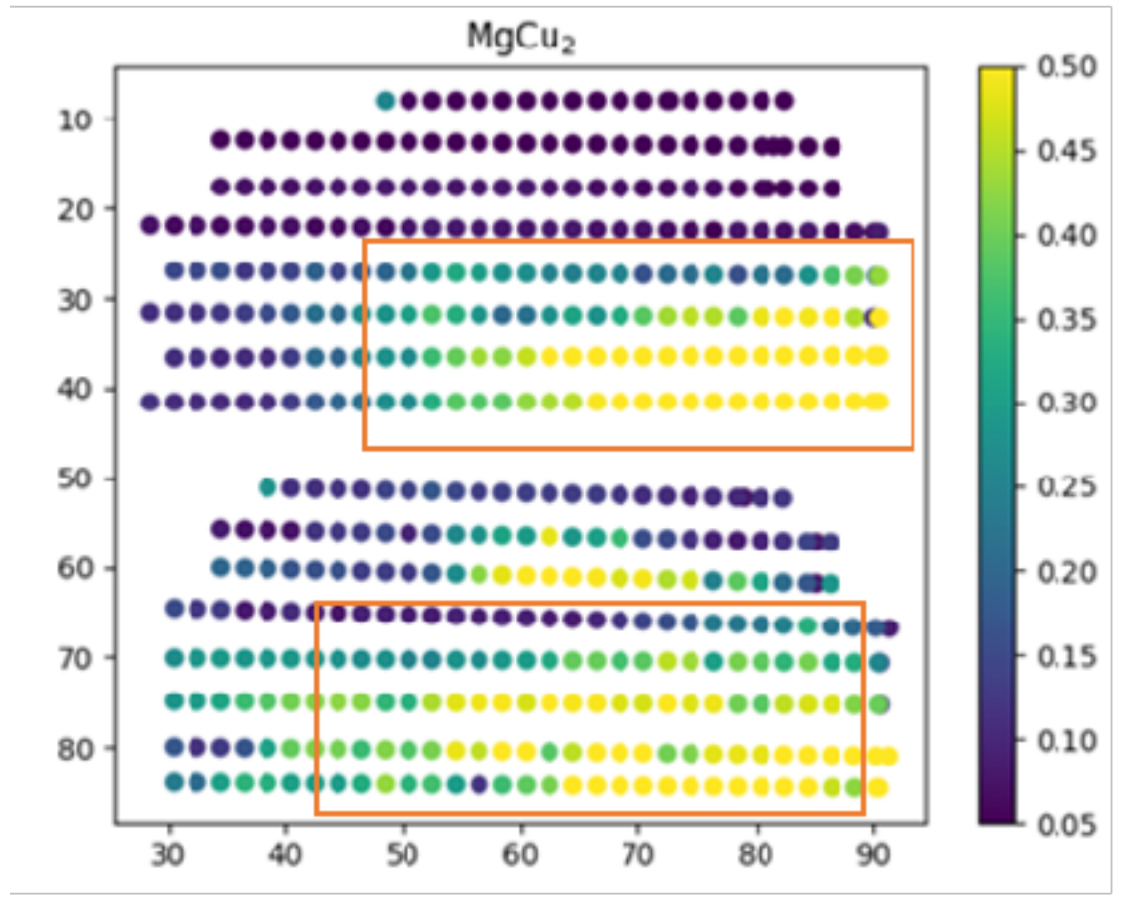


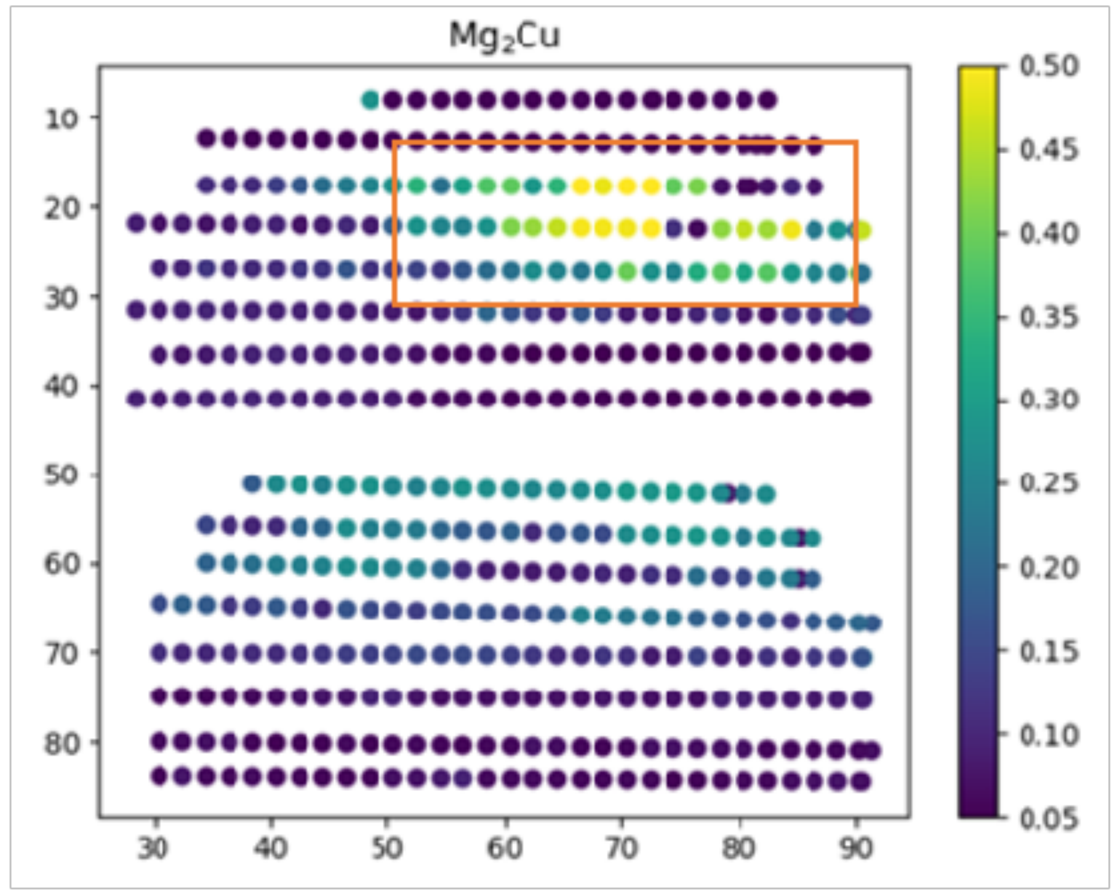
Mg_2Cu (0 8 0)

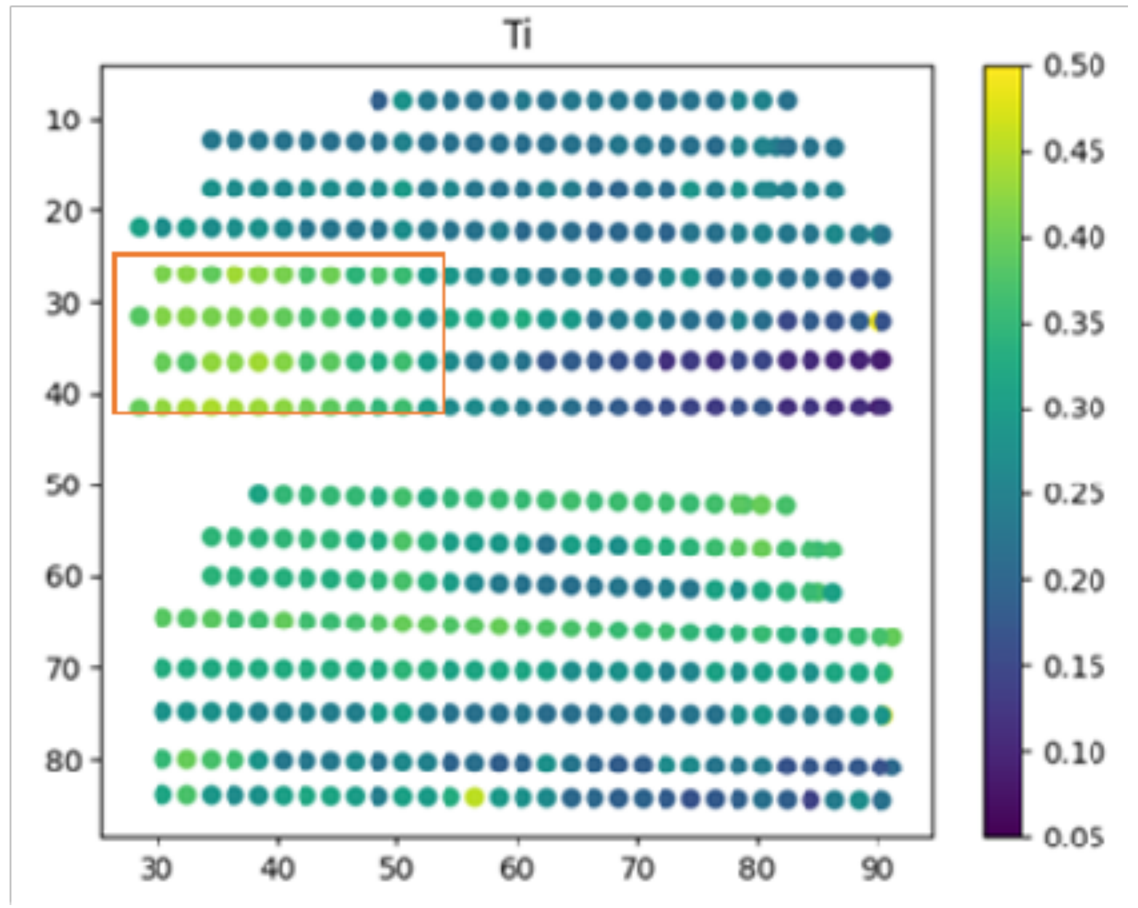


Potential Ti (1 1 0)









U.S. DEPARTMENT OF ENERGY | Office of Science

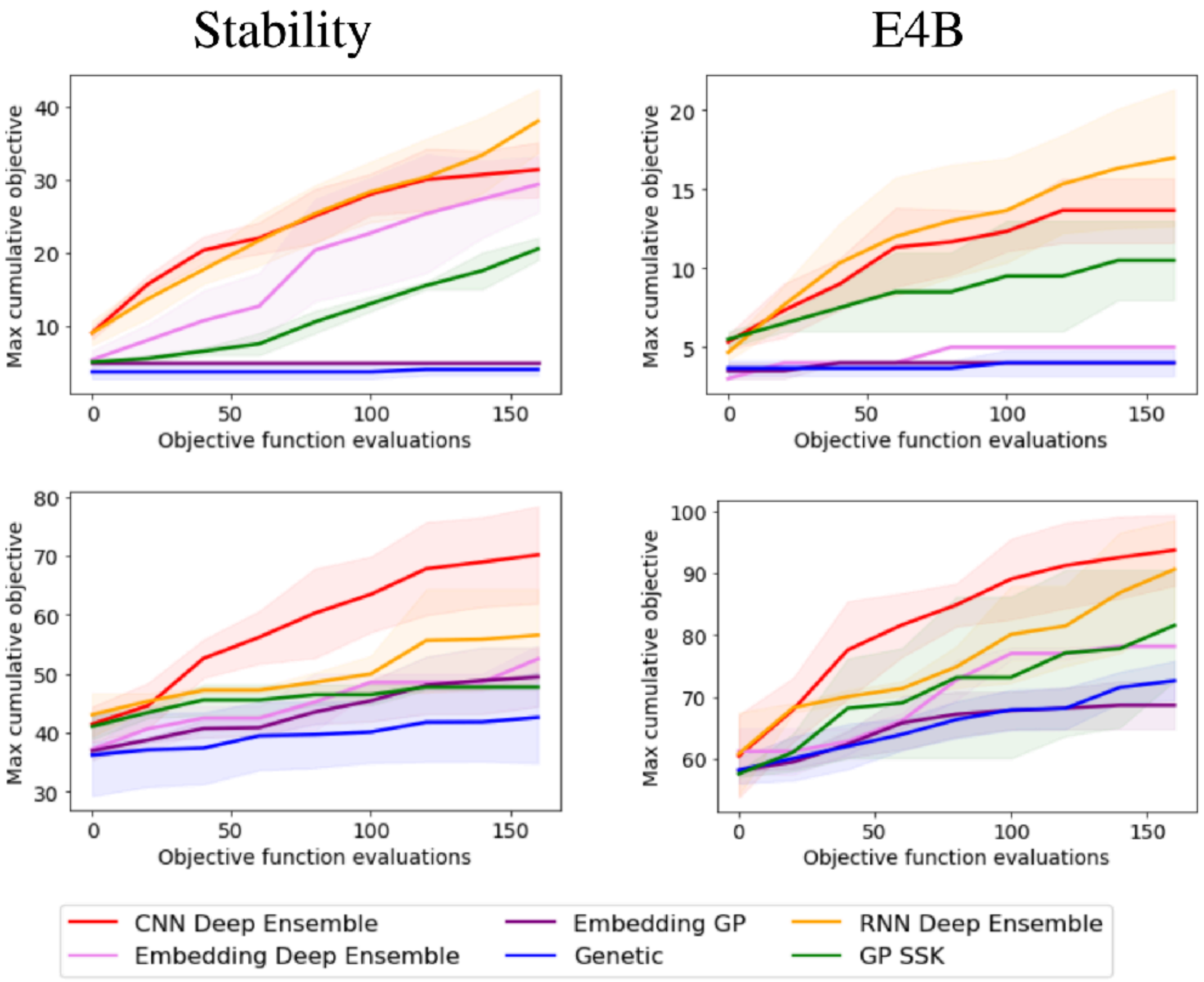
BROOKHAVEN NATIONAL LABORATORY

Stony Brook University

Under Review

National Synchrotron Light Source II

String based optimization tasks can be accomplished using a suite of network architectures.

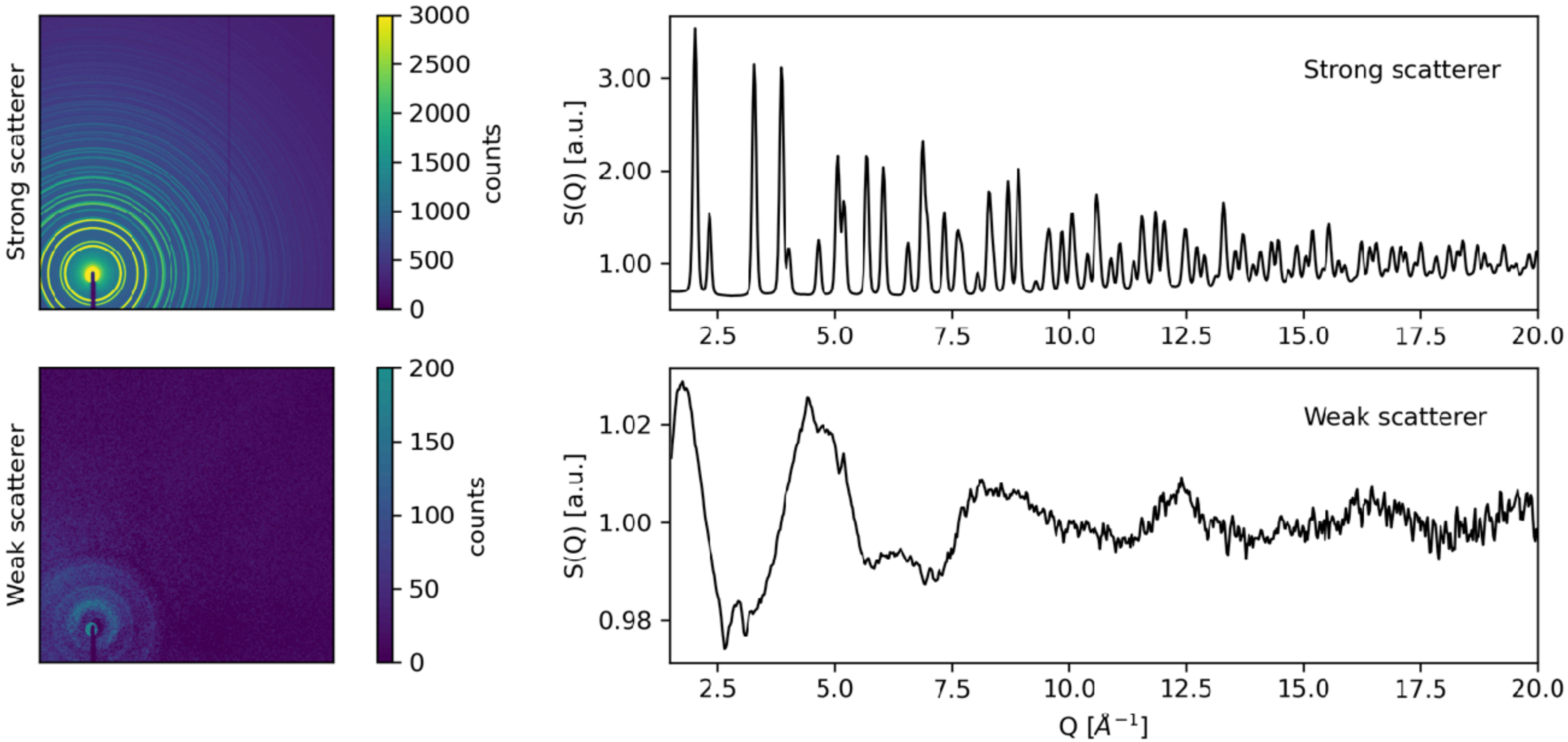


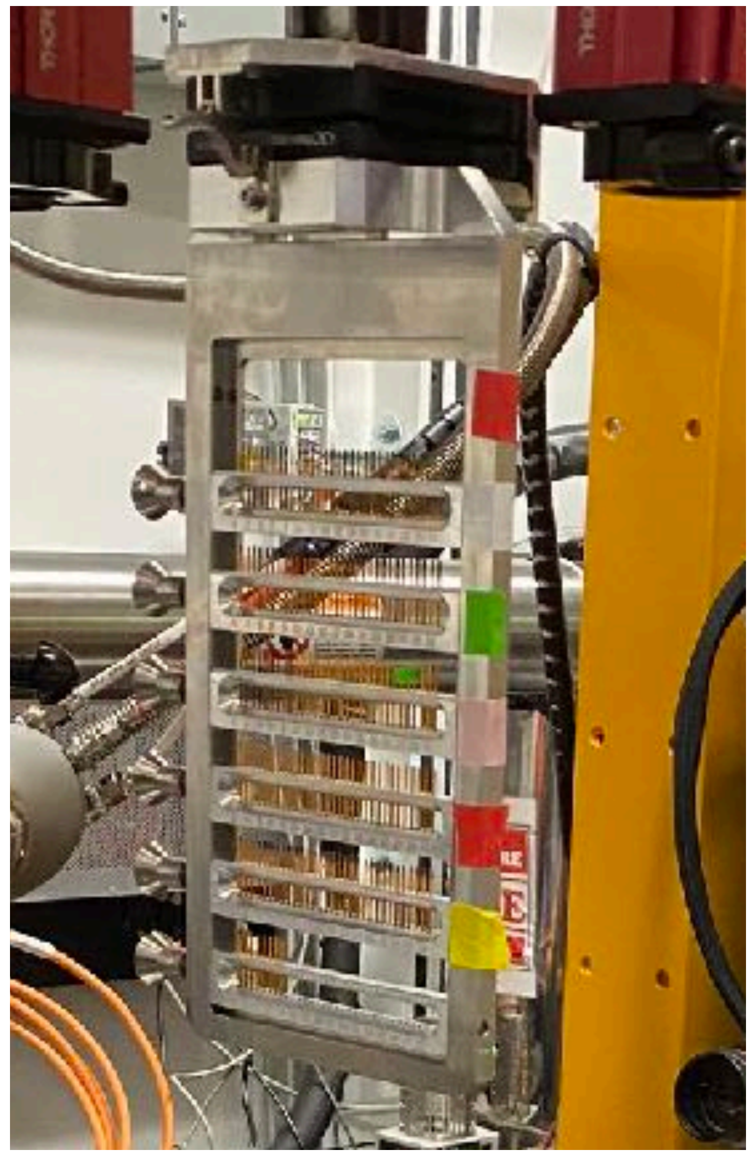
Gruver, N. *et al., ICML 2020*

Reinforcement learning:

For when model training is more costly than an experimental step.

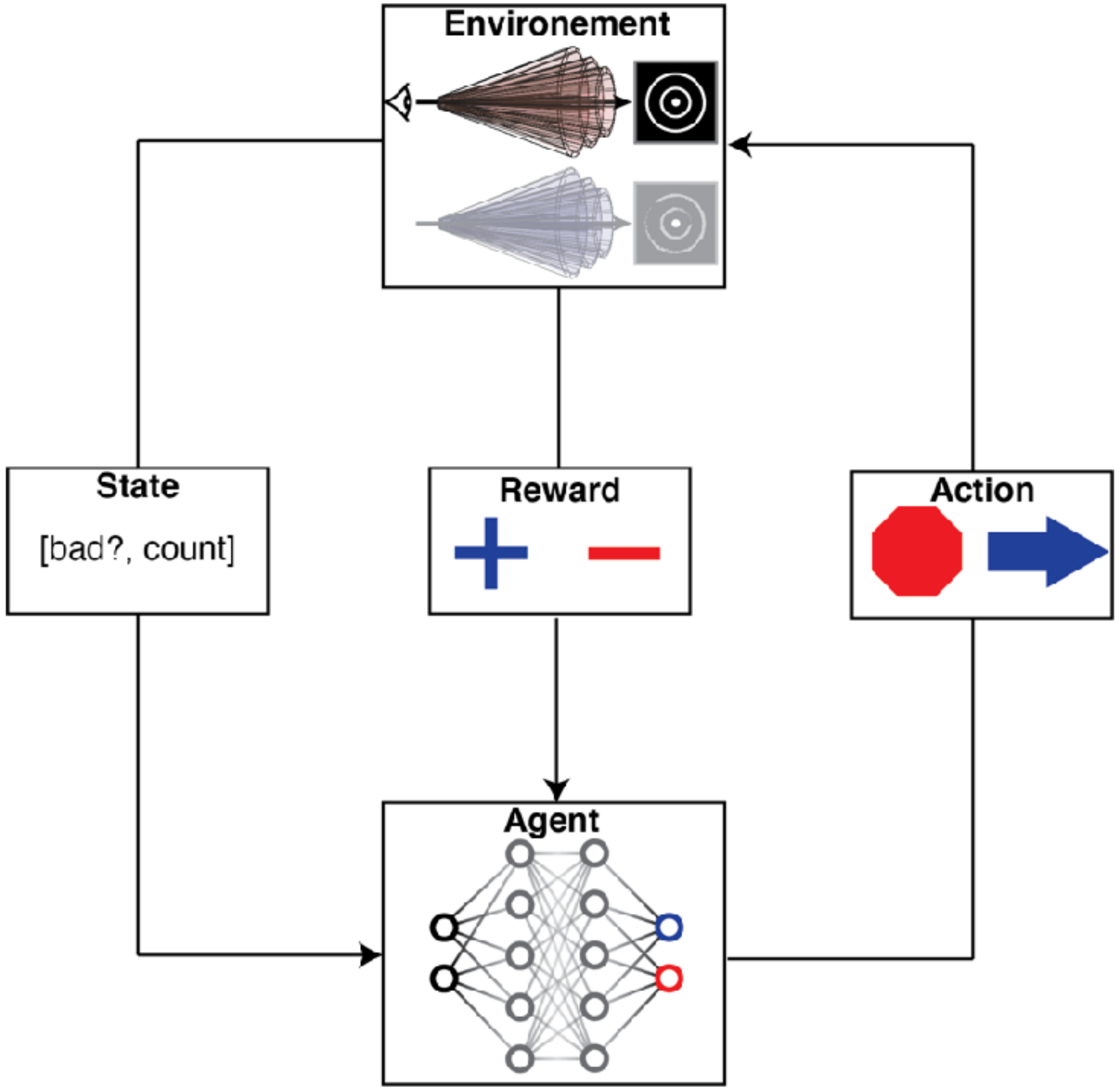
With hundreds of samples to run remotely, how do we best utilize our resources?

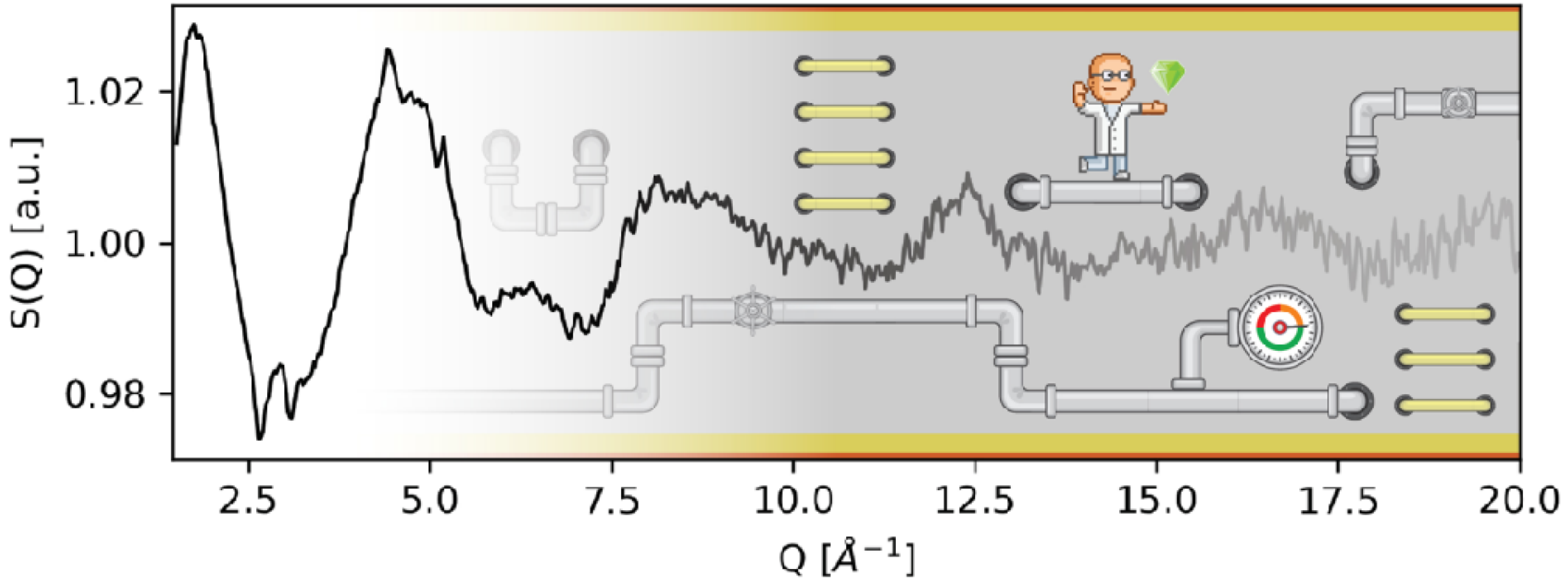




Mach. Learn.: Sci. Technol. **2**, 025025 (2021)

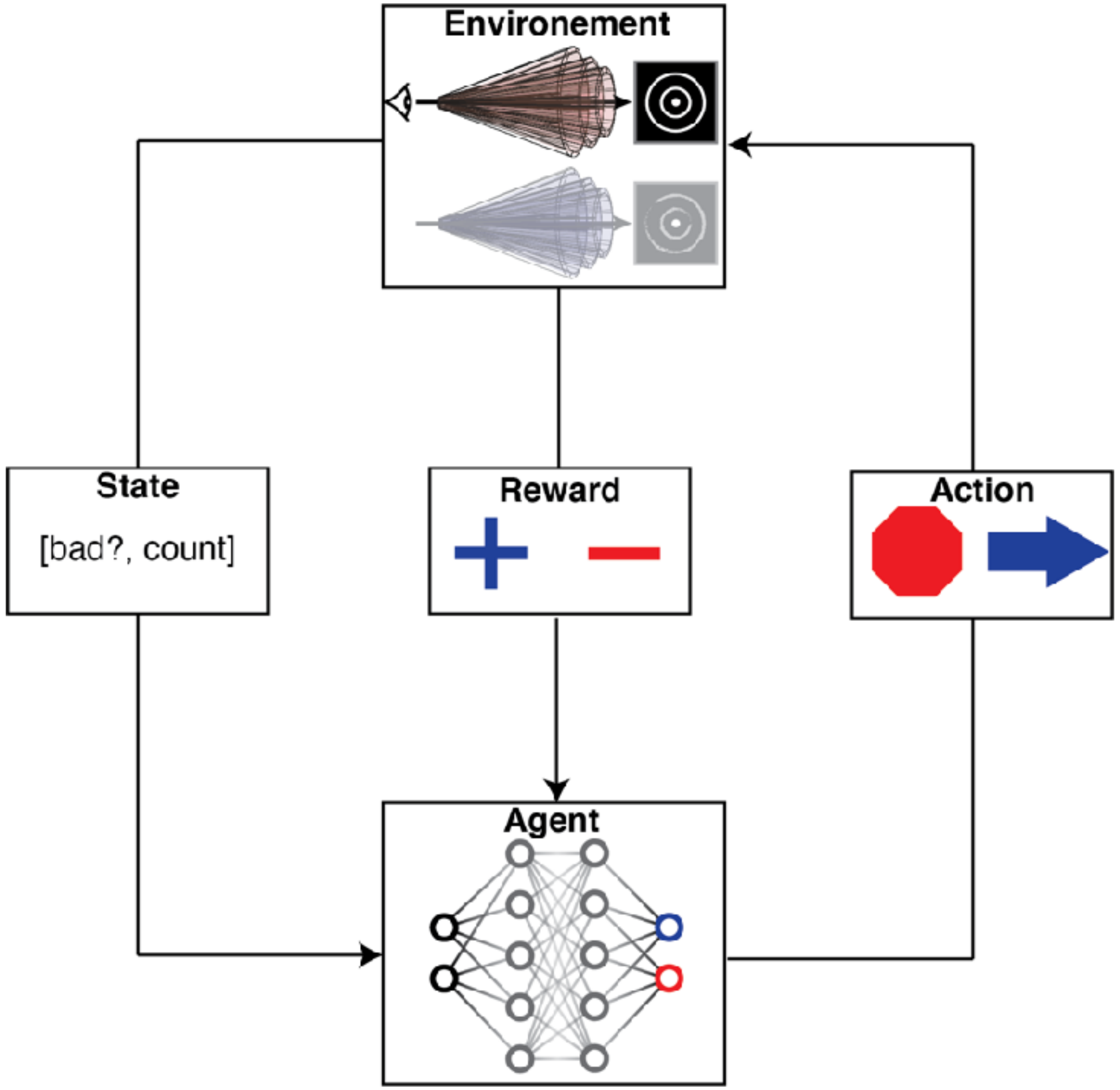
Reinforcement learning develops policies for optimal measurement strategies.

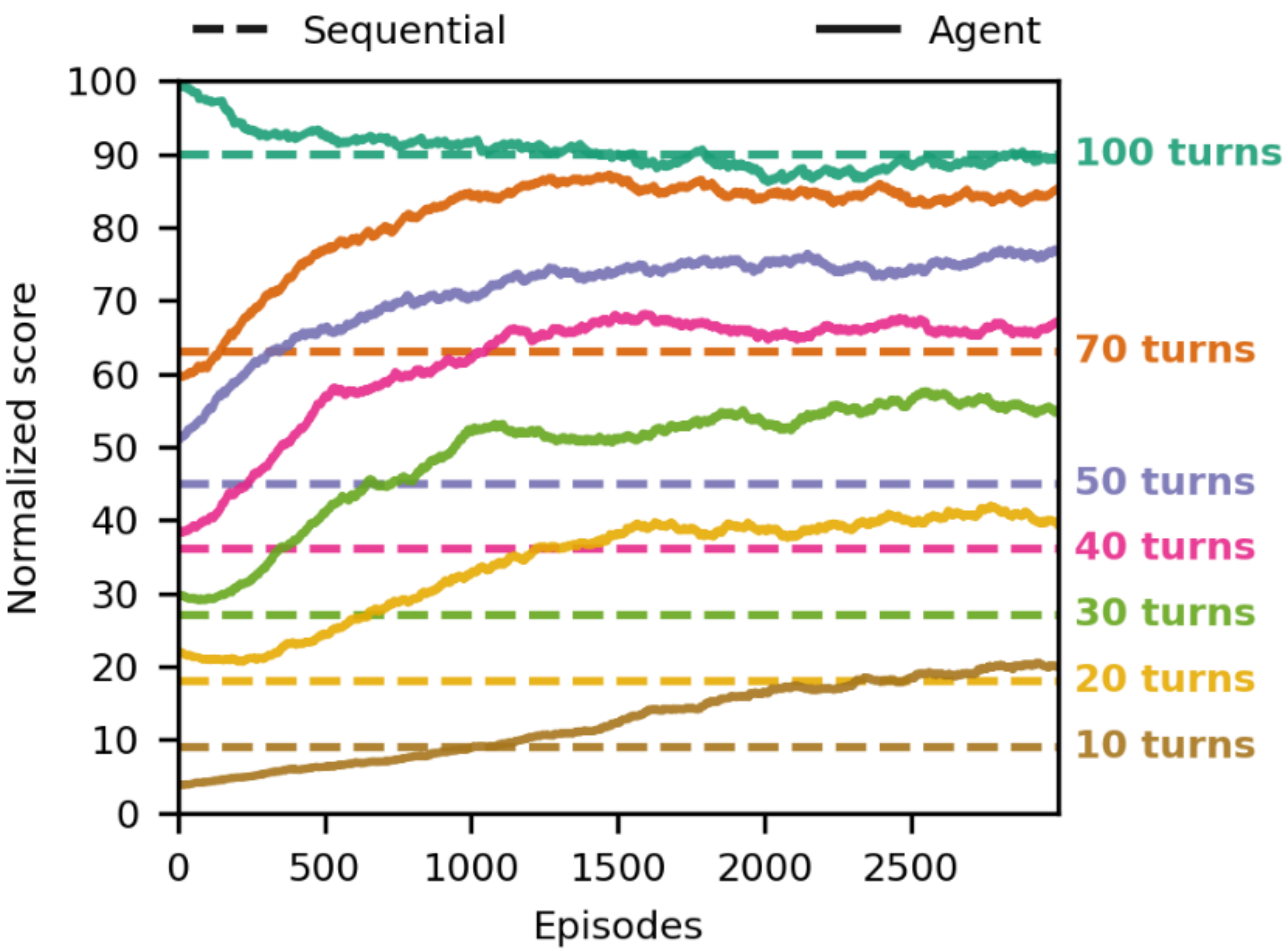




Mach. Learn.: Sci. Technol. **2**, 025025 (2021)

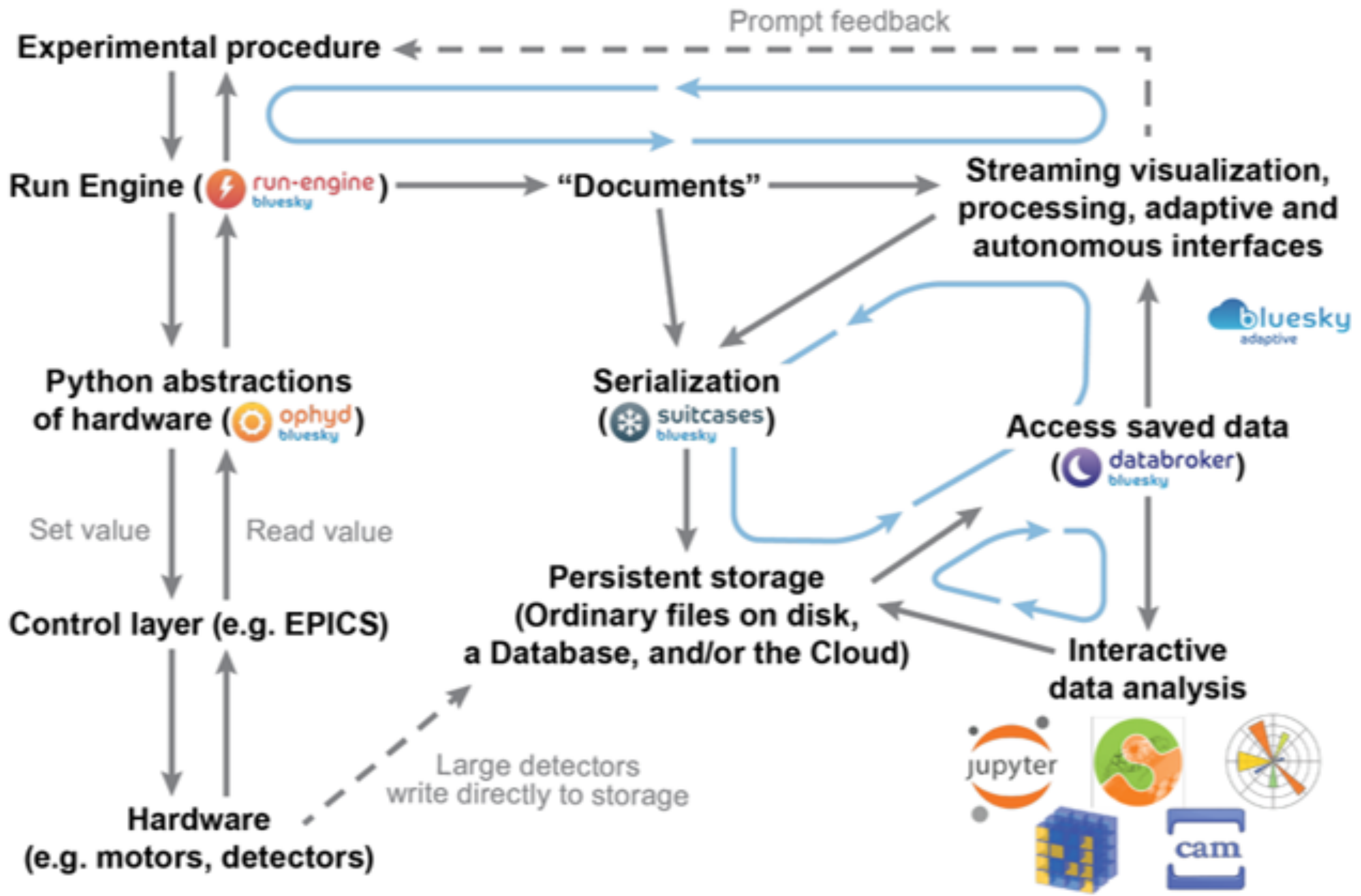
Reinforcement learning develops policies for optimal measurement strategies.



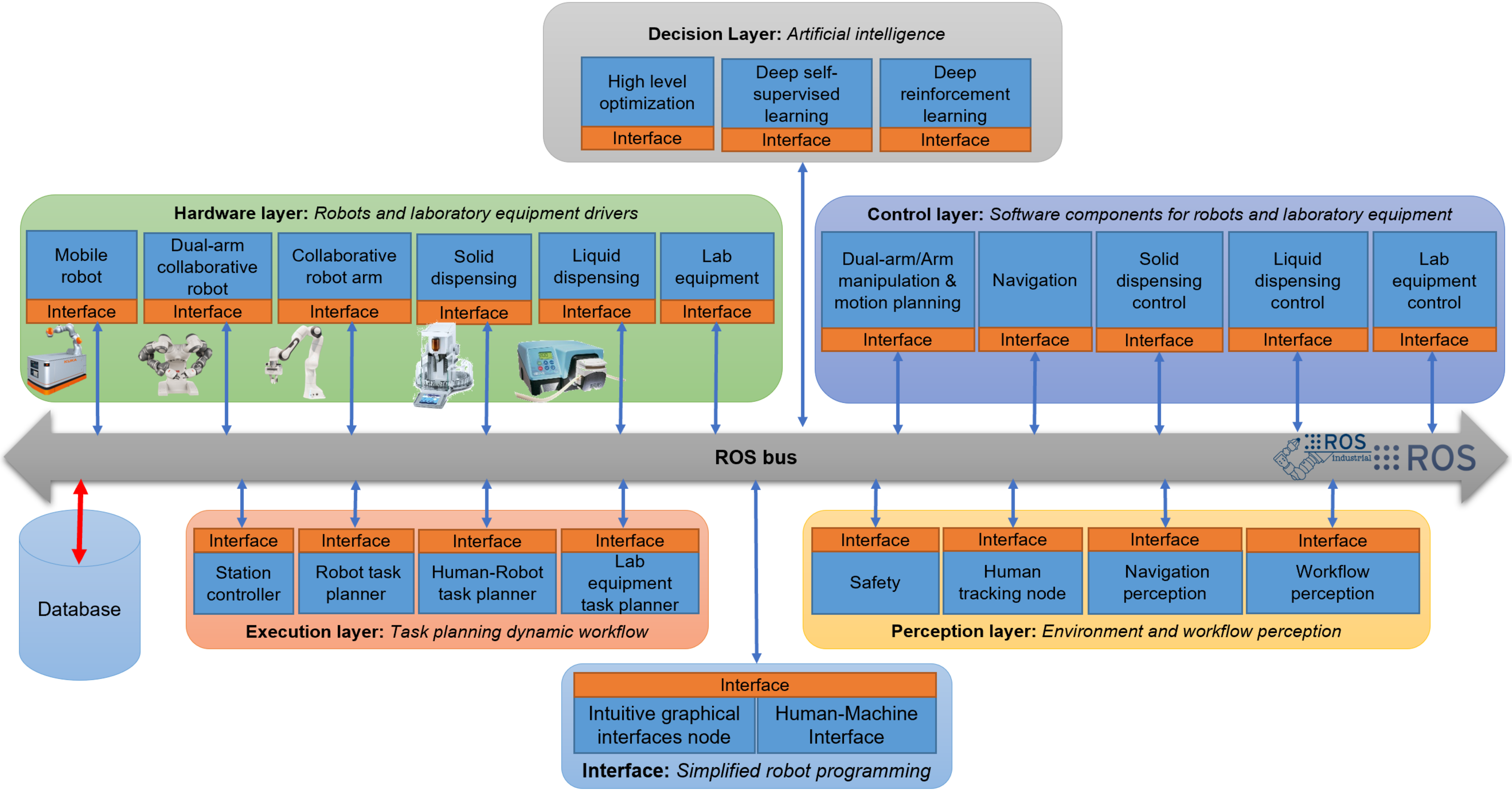


Mach. Learn.: Sci. Technol. **2**, 025025 (2021)

Artificial intelligence for beamline science



ROS-Laboratory



Diverse laboratory tasks connected via message bus, with real time bus for robotics.

Sample Preparation

Solid Dispensing:
/Dosing_status
/Door_position
/Sampler_position
/Dosing_target

Liquid Dispensing:
/Pump_status
/Pump_speed
/Pump_direction
/Timer

Mixing:
/mixer_status
/Speed_target
/Timer

Analysis

KUKA

ROS industrial

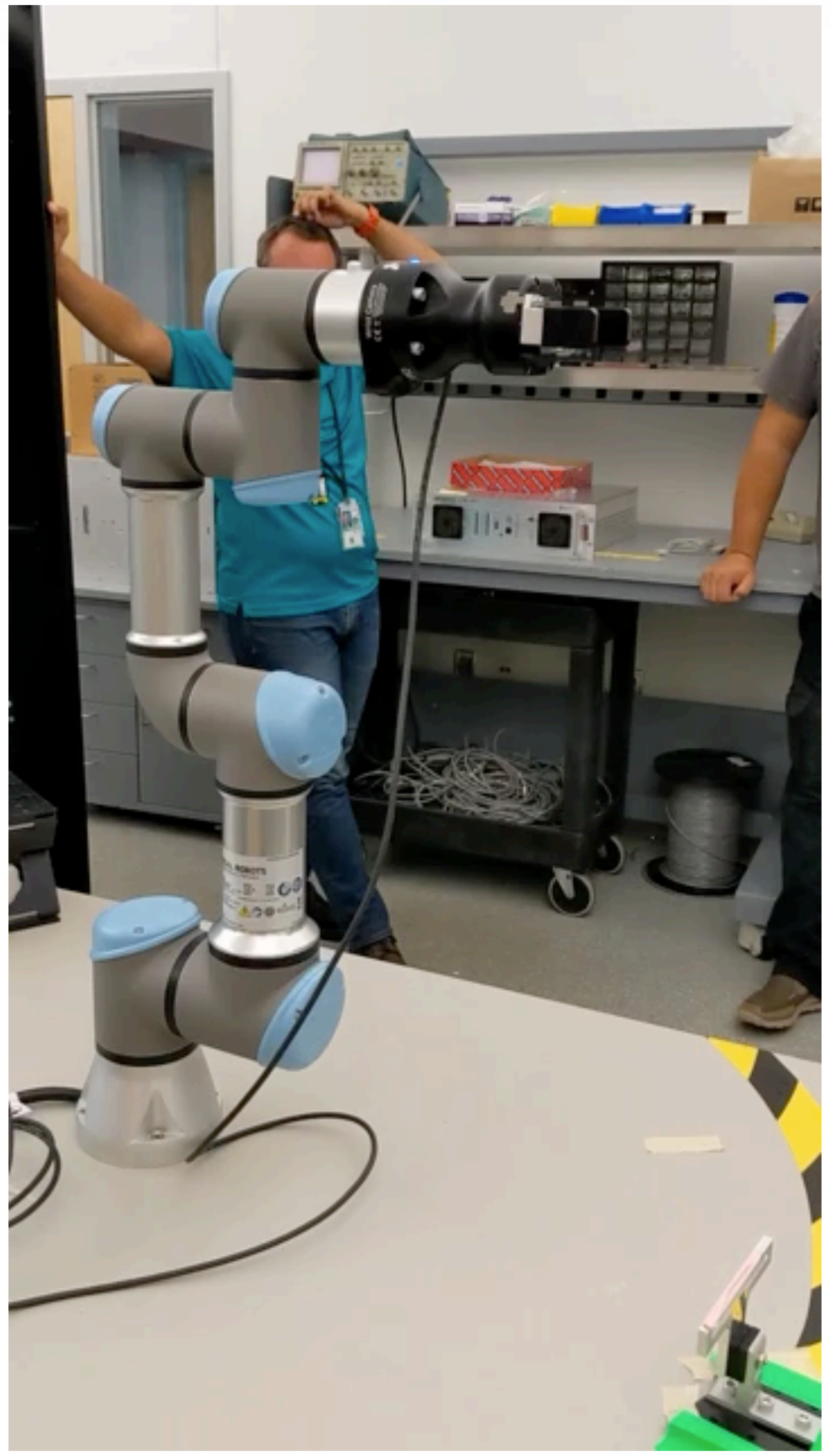
abb_driver

ROS industrial

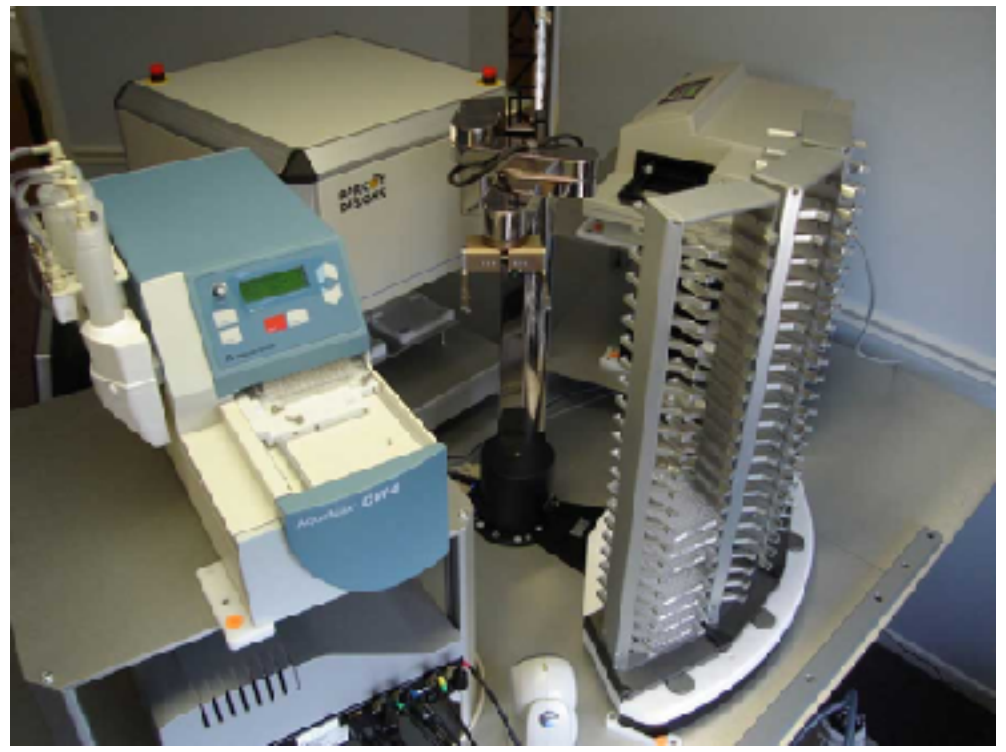
Mettler-Toledo
QS30 Autosampler

Watson Marlow
323DU Pump

IKA
ULTRA-TURRAX



Workcells are efficient self-contained unit operations.



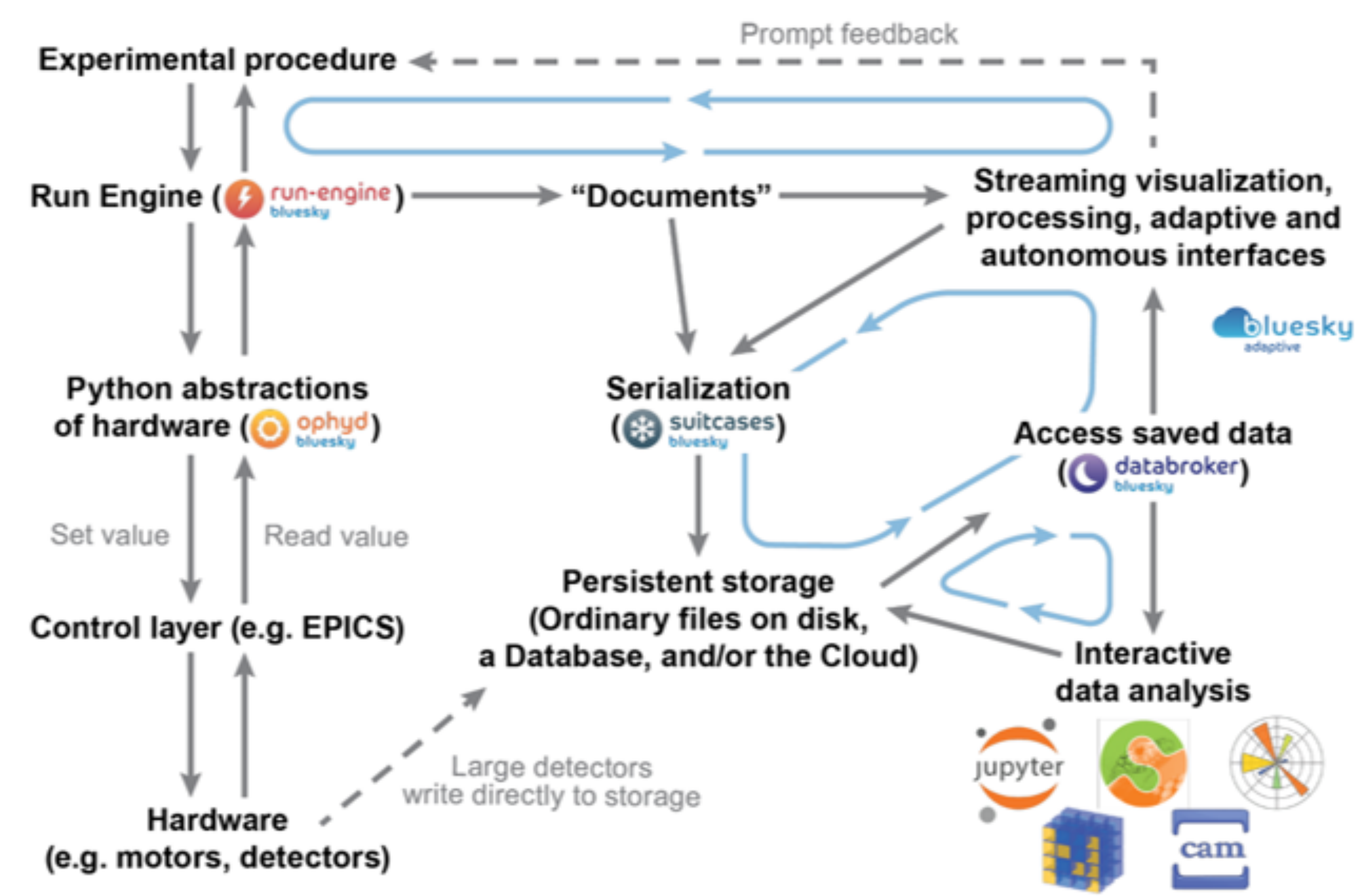
Automated Pipetting | Limited-function robot | Heavy robots (some integration) | Integrated Workcell

BigHat
BIOSCIENCES

Acknowledgements

Thank You!

- Using the right tool for the right job.
- Scalable infrastructure for linking models to experiments.
- Keeping the human in the loop.
- Open source automation is the frontier for autonomous experiments.



GitHub /Maffettone

@PhilMaffettone