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

Dr. Phil Maffettone

13 Sept 2021

Acknowledgements

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 @ BNL

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

Columbia University

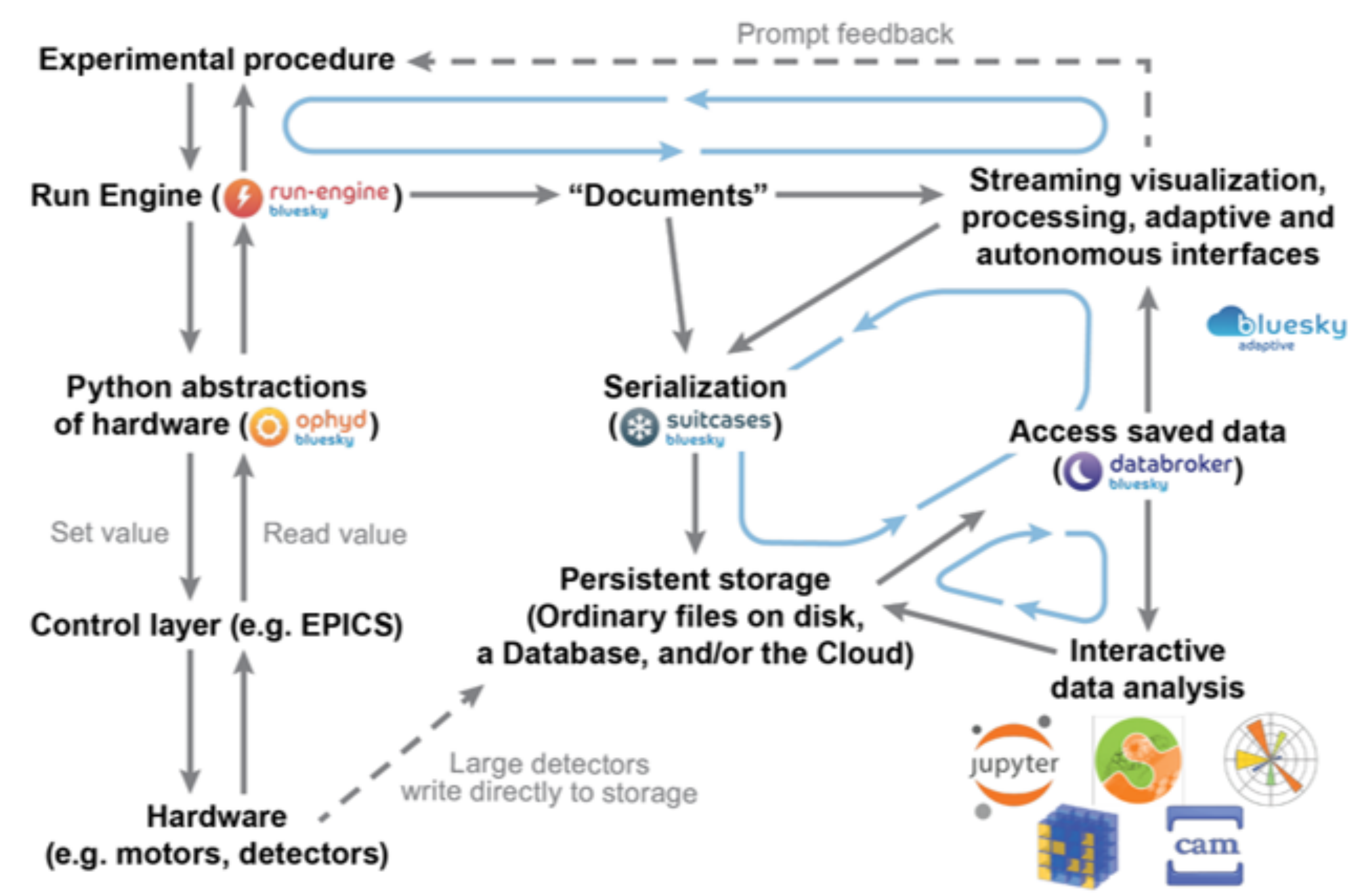
Dr. Boyan Penkov

Prof. Ken Shepard

Flatiron Institute

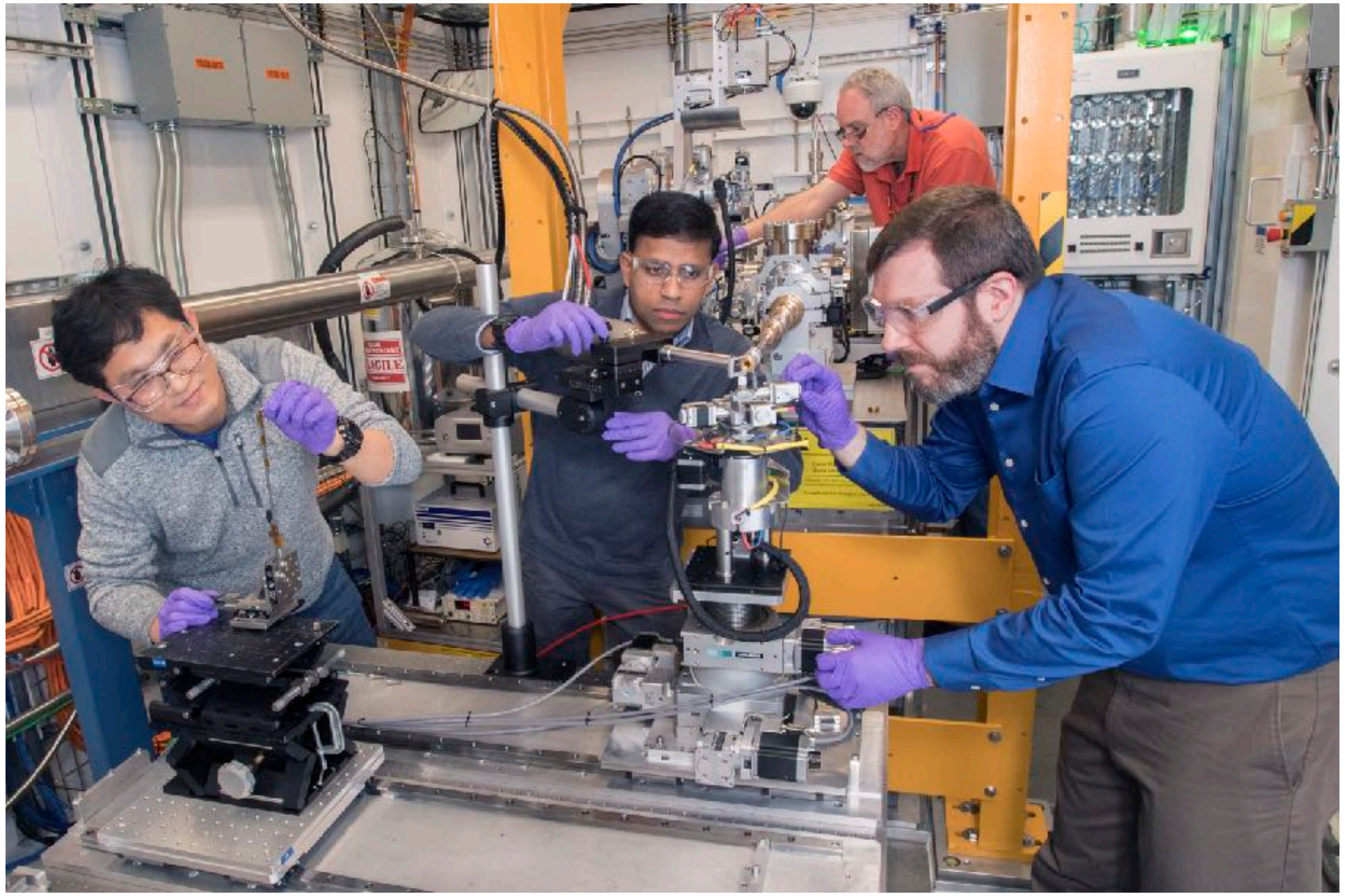
Dr. Aidan Daly

Artificial intelligence for beamline science

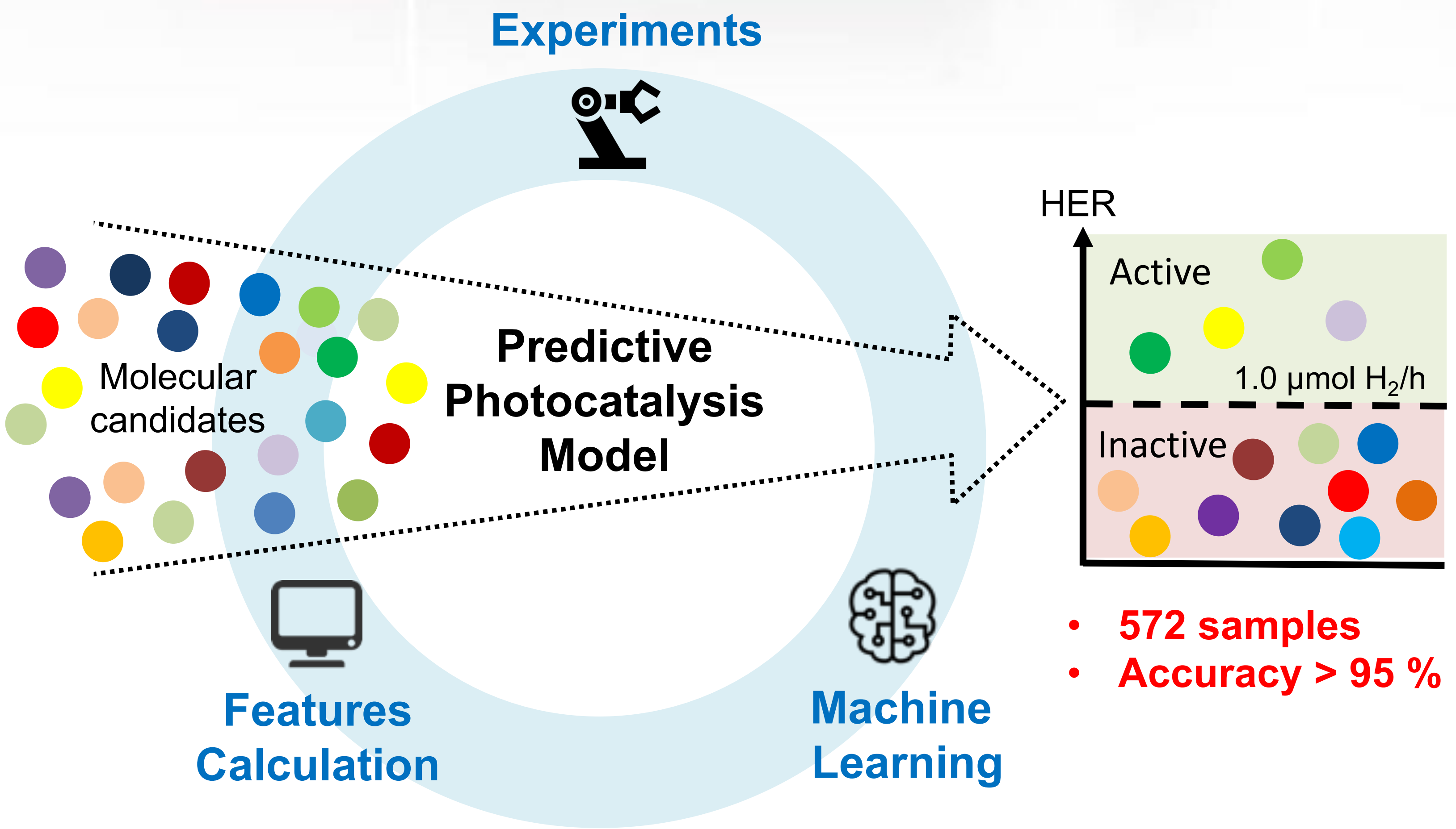


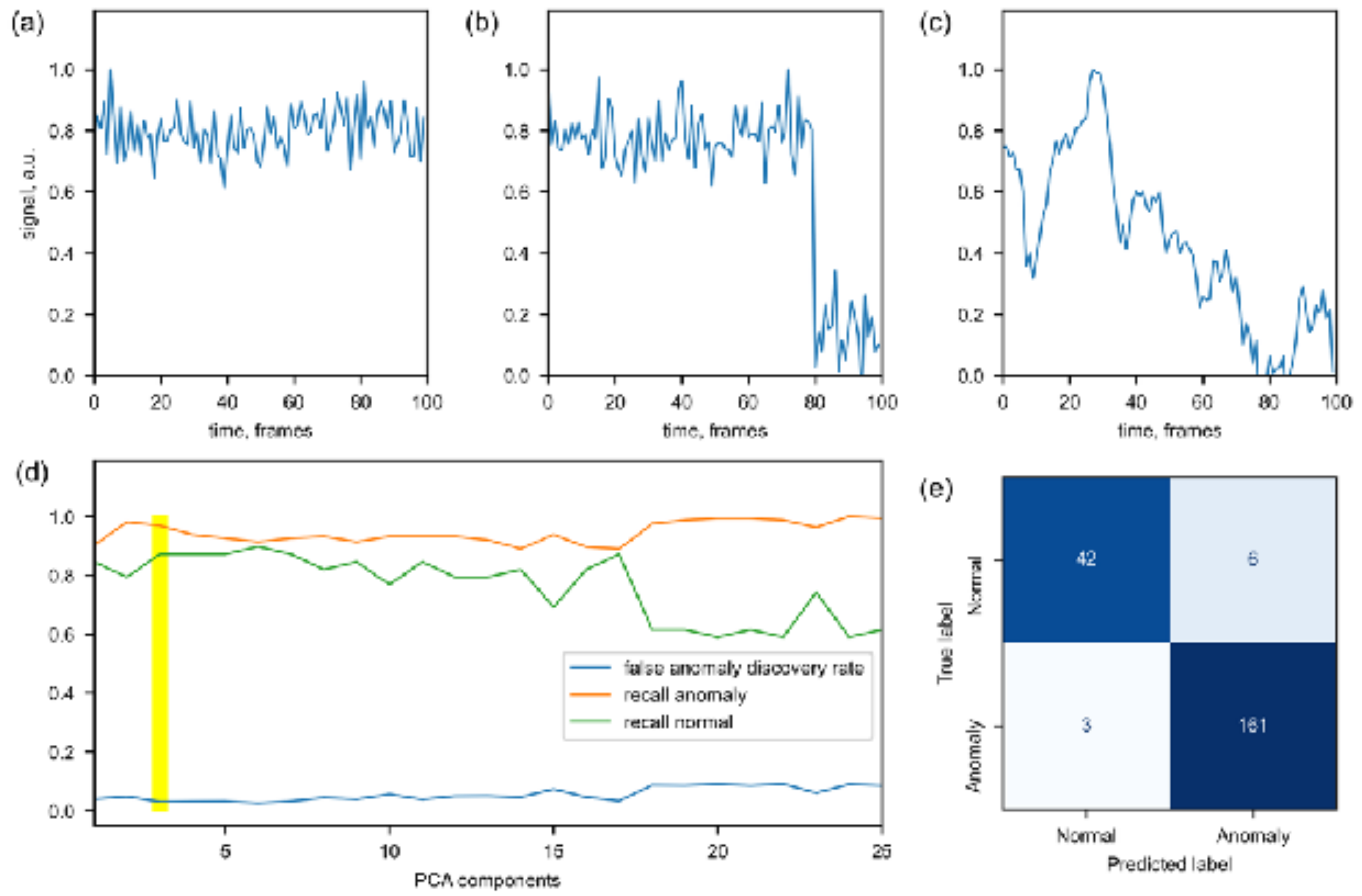
Federated AI, data streaming, and pragmatic engineering

- 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.
- Scalable automation is the frontier for high-throughput and autonomous experiments.

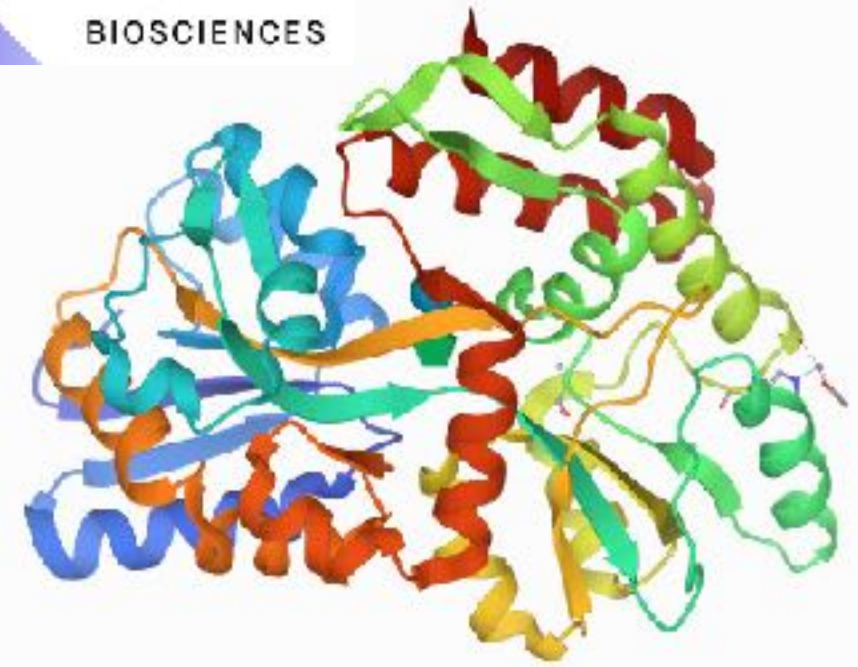


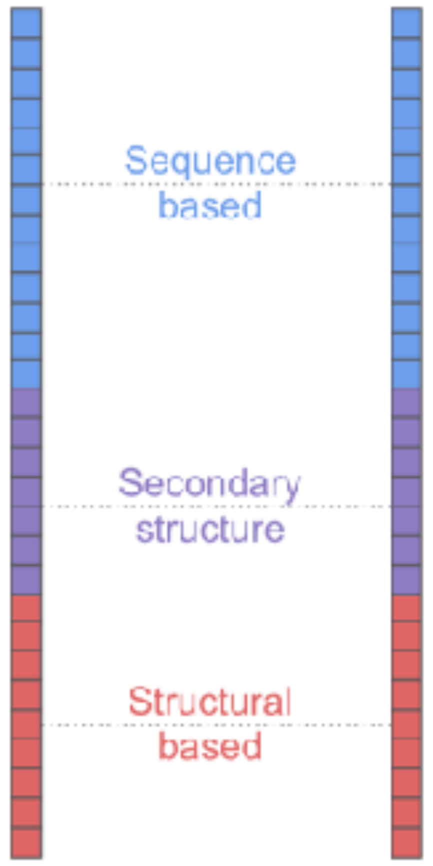
Feature engineering can be incredibly effective, and requires collaboration with domain experts.

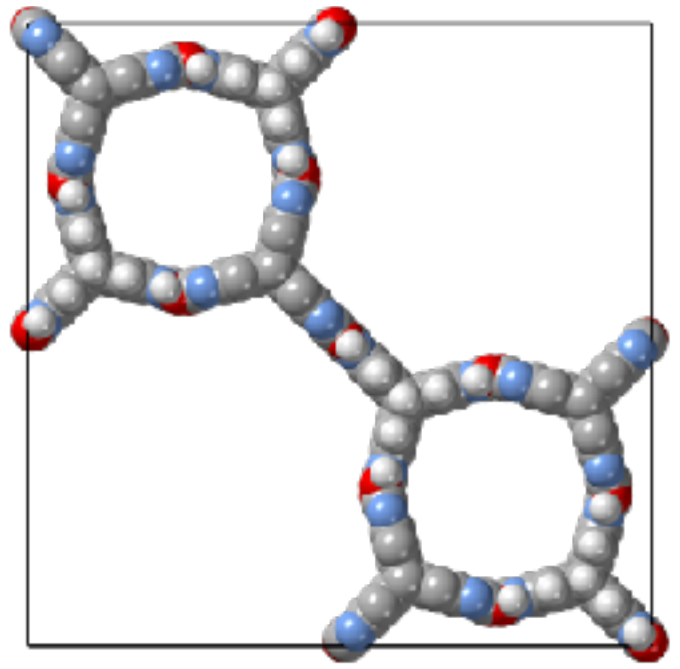


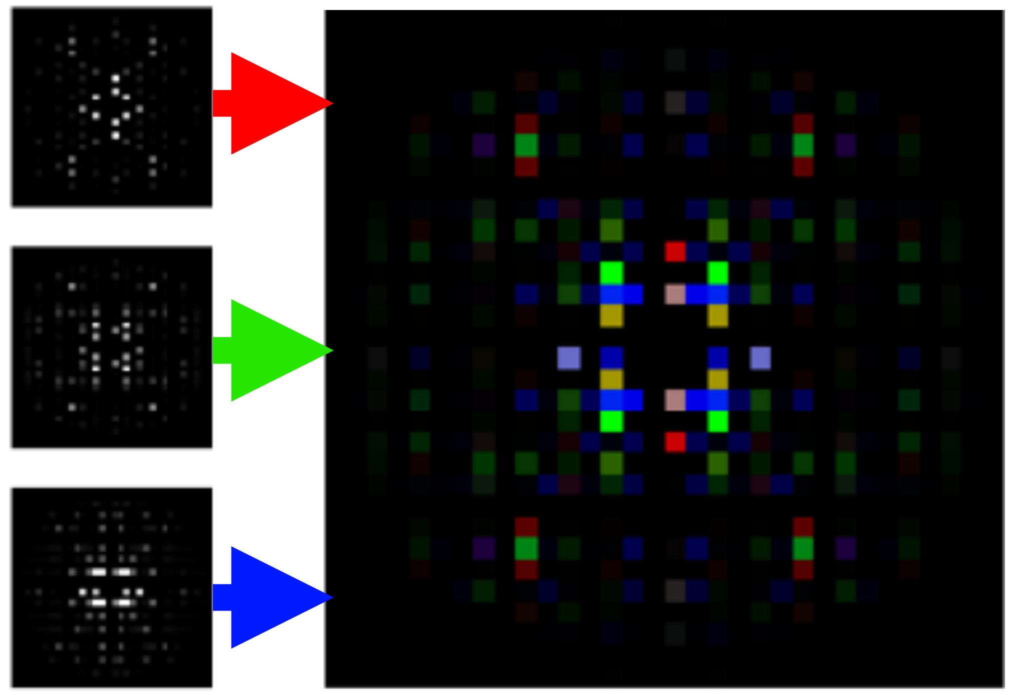


BigHat BIOSCIENCES









U.S. DEPARTMENT OF ENERGY | Office of Science

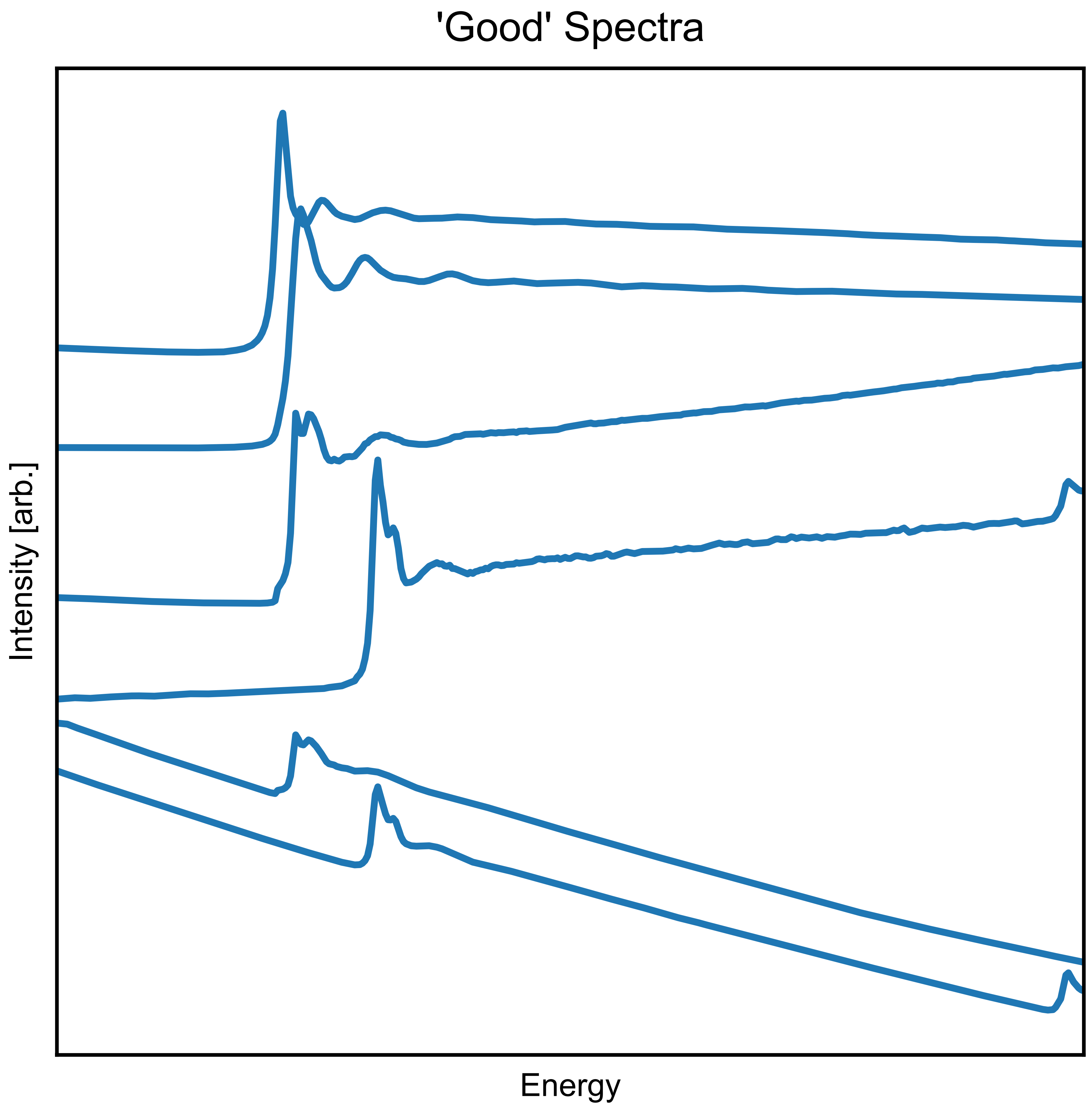
BROOKHAVEN NATIONAL LABORATORY

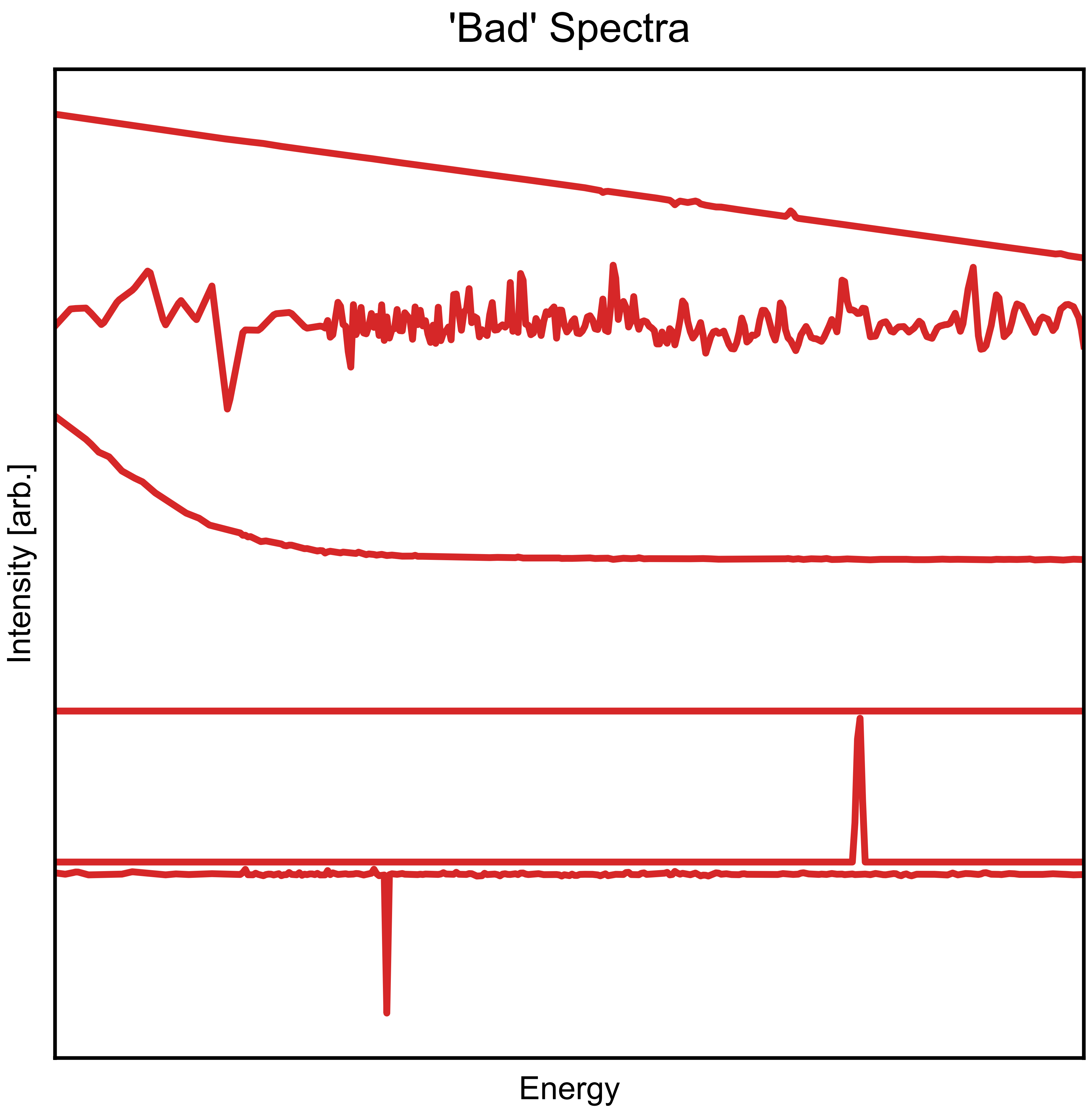
National Synchrotron Light Source II

Supervised learning:

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

Identifying experimental failures at BMM.



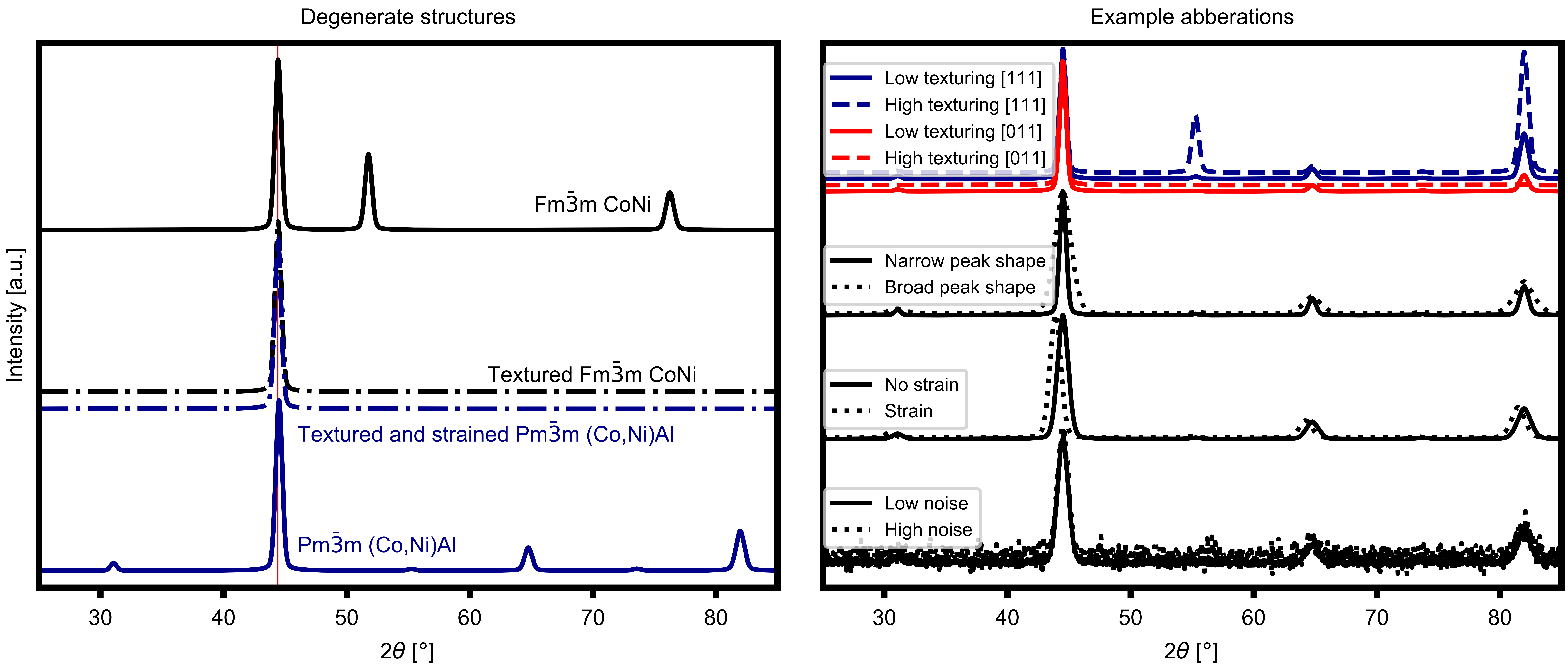


Companion agents for classifying data streams:

Applications in phase hunting, mapping, and transitions.

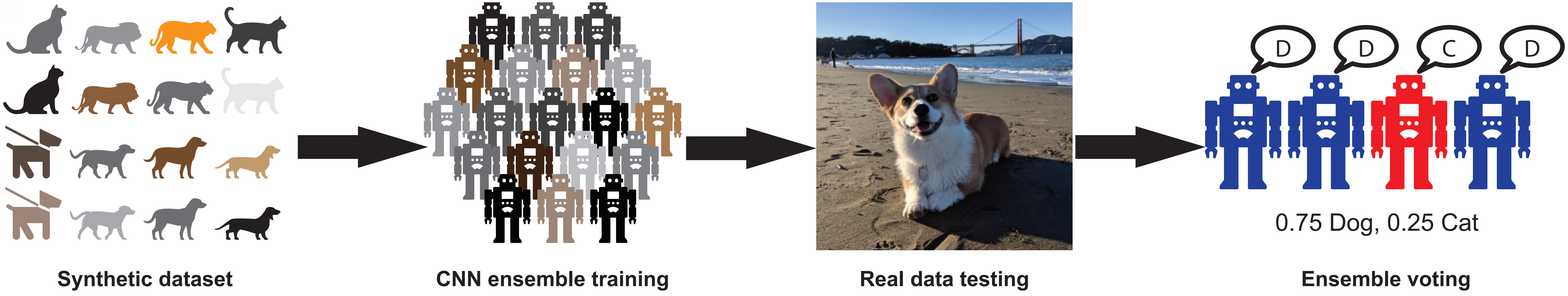
Maffettone, P.M. *et al., Nat. Comp. Sci.*, **1**, 290-297 (2021)

X-ray diffraction is an information poor measurement.



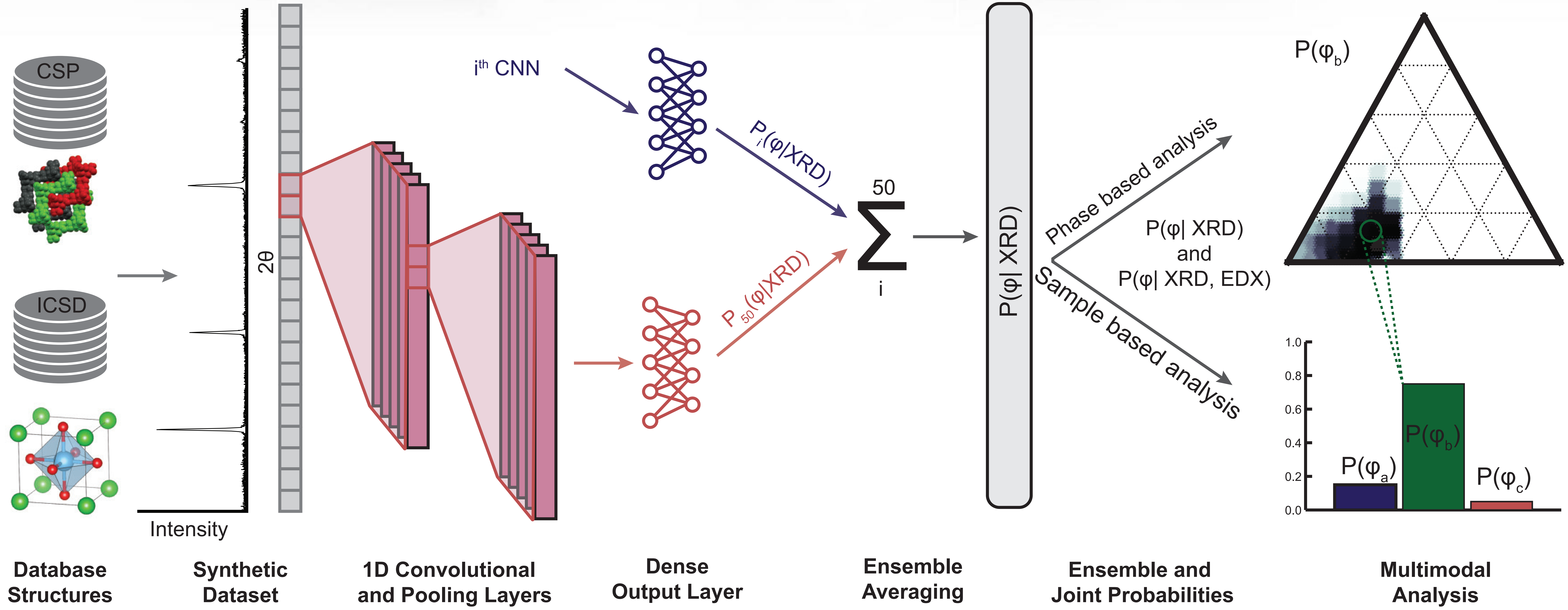
Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

We can synthesize realistic XRD datasets, and use ensembles to overcome the overconfidence of single neural nets.



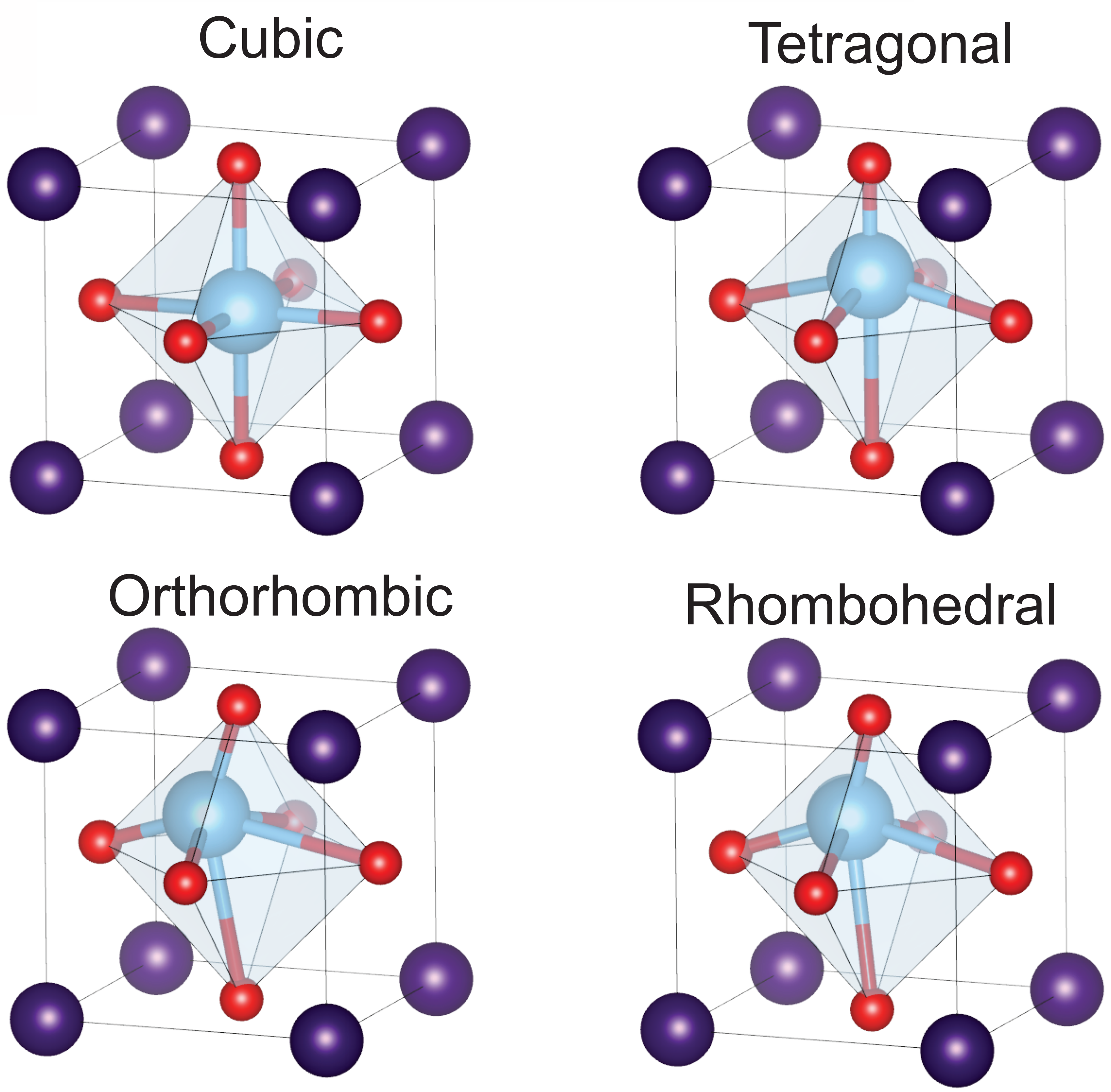
Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

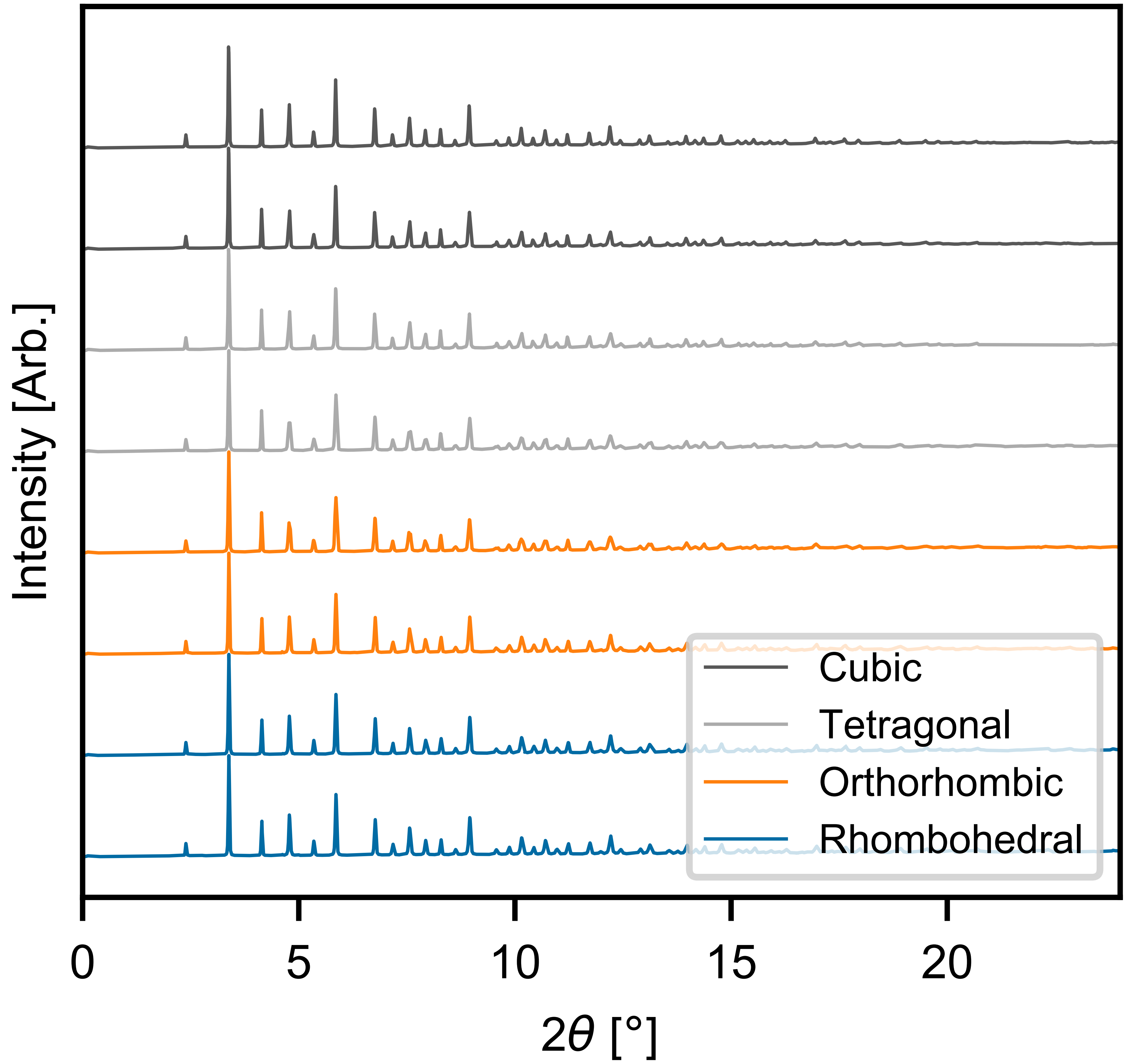
A fully synthetic data pipeline can train an accurate probabilistic model prior to the experiment.



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

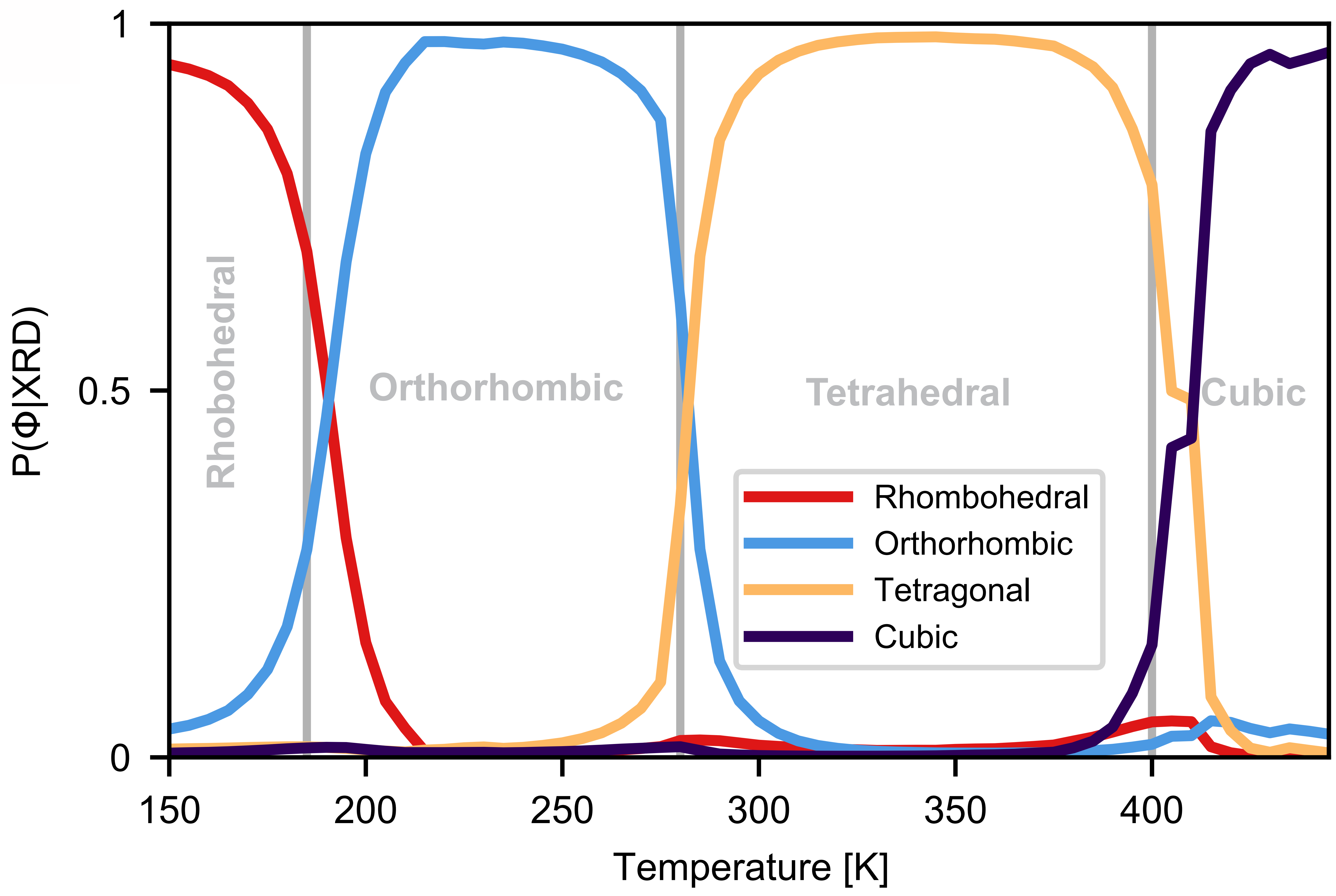
Classifying subtle phase transitions in $BaTiO_3$.





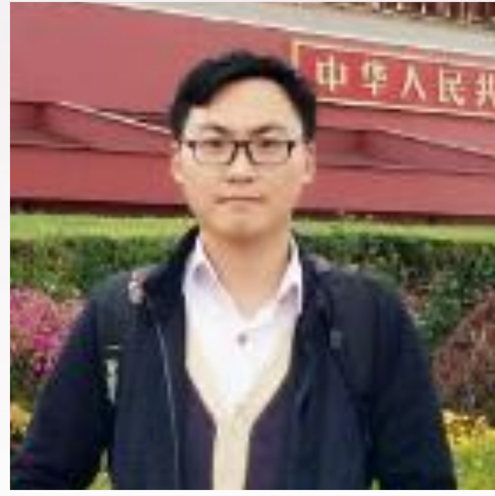
Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

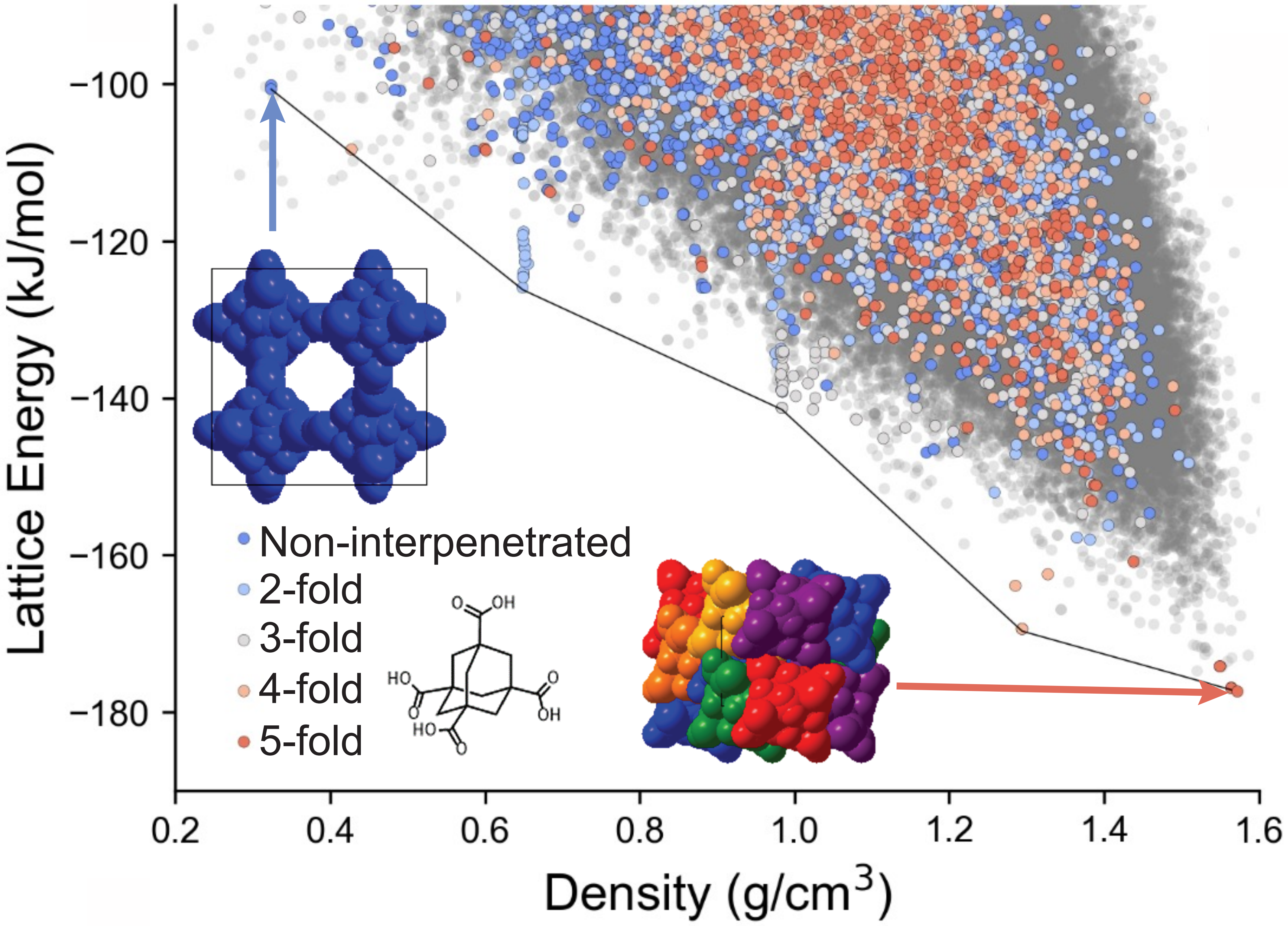
Classifying subtle phase transitions in $BaTiO_3$.

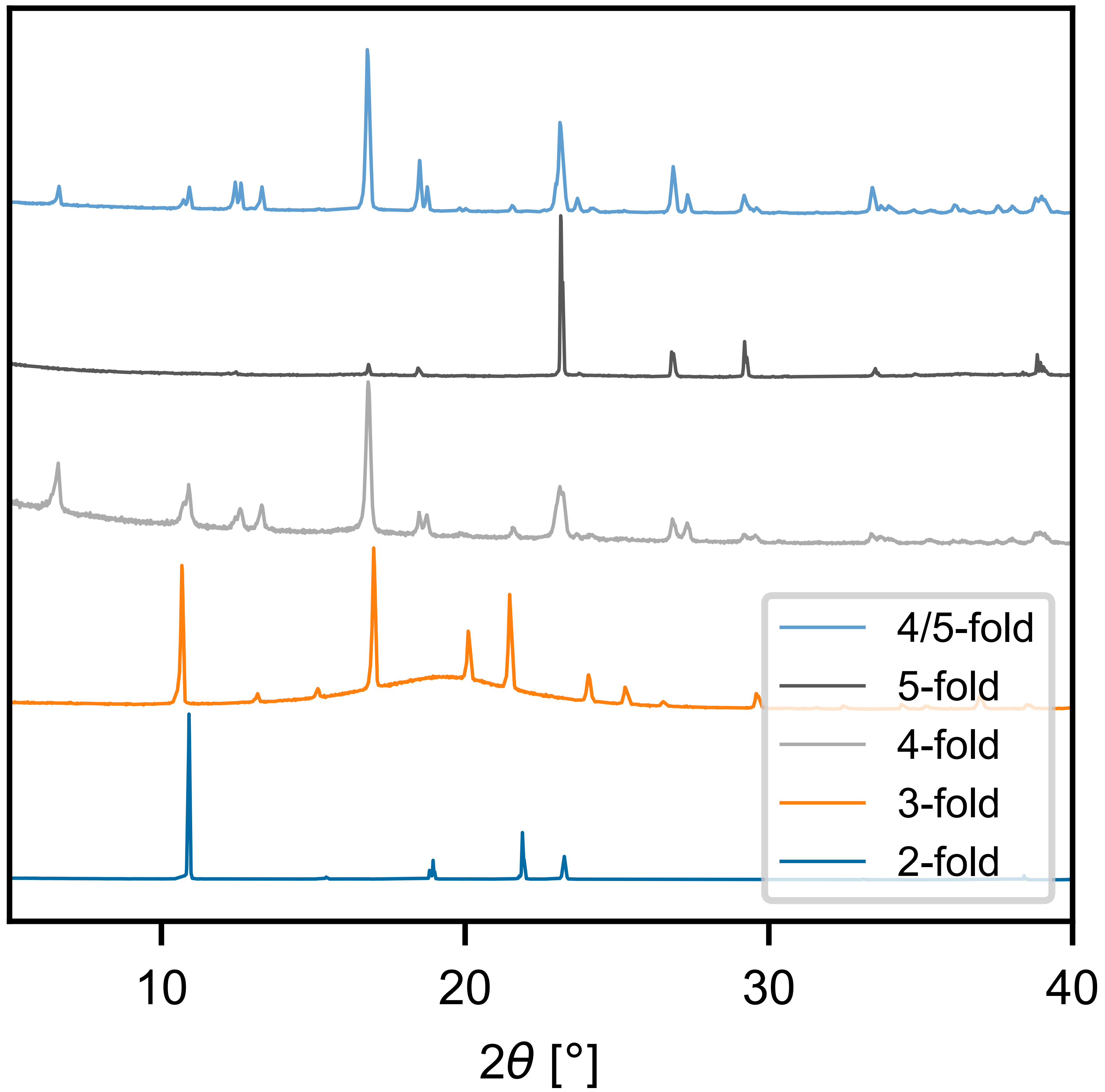


Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

Searching for an elusive porous polymorph.

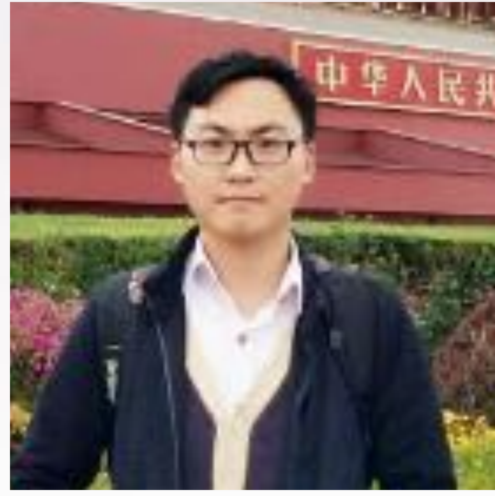


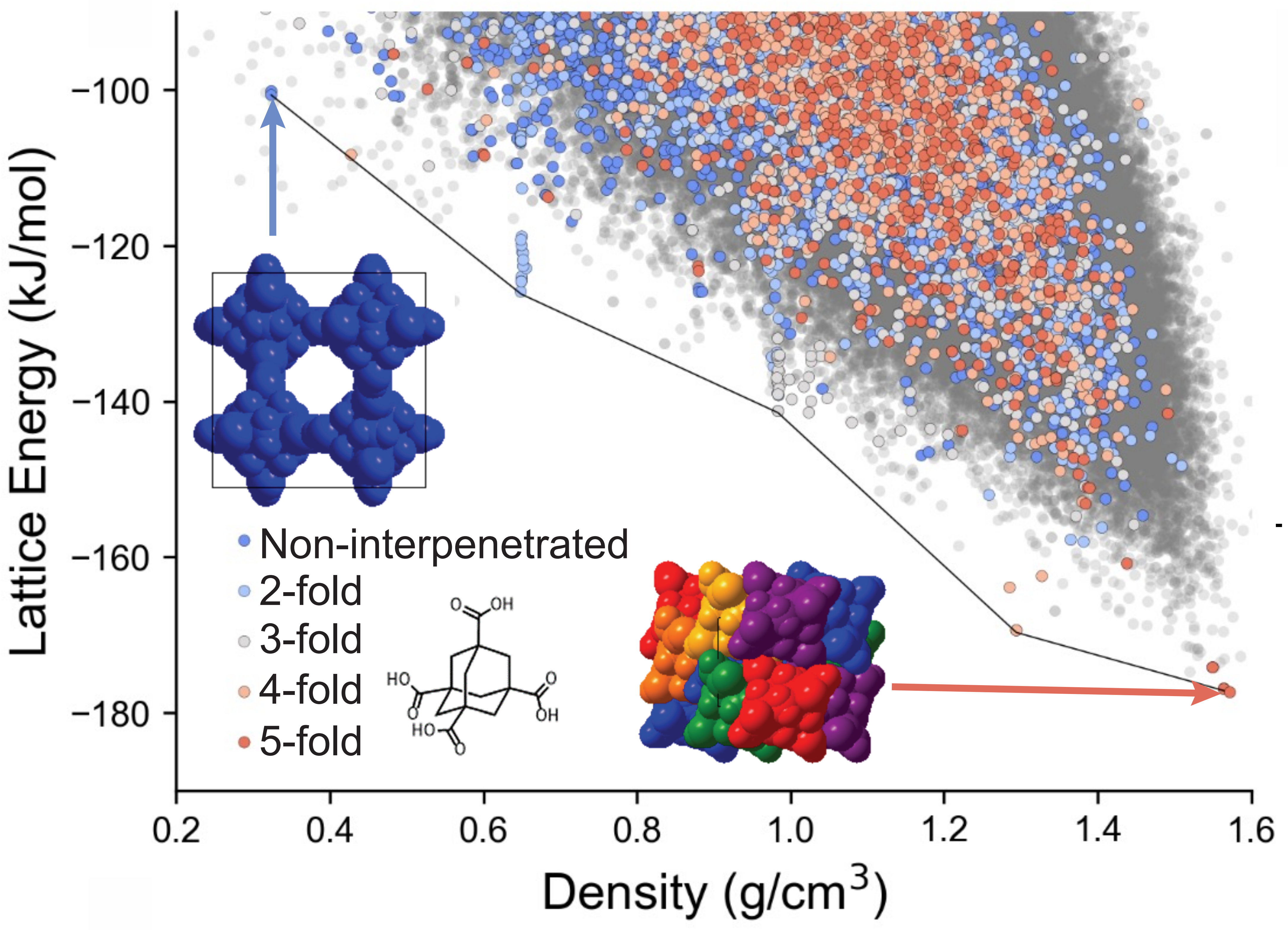


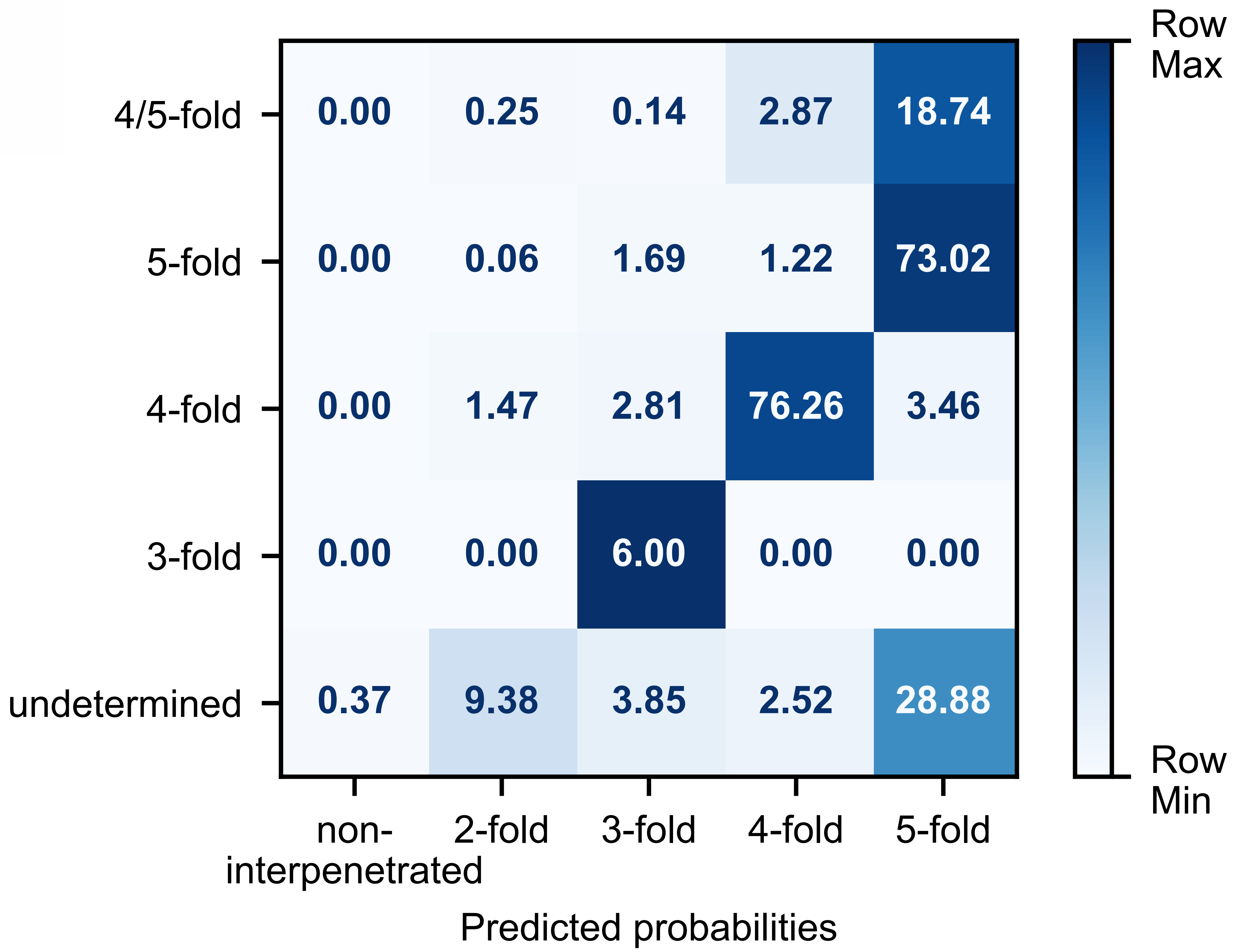


Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

Searching for an elusive porous polymorph.

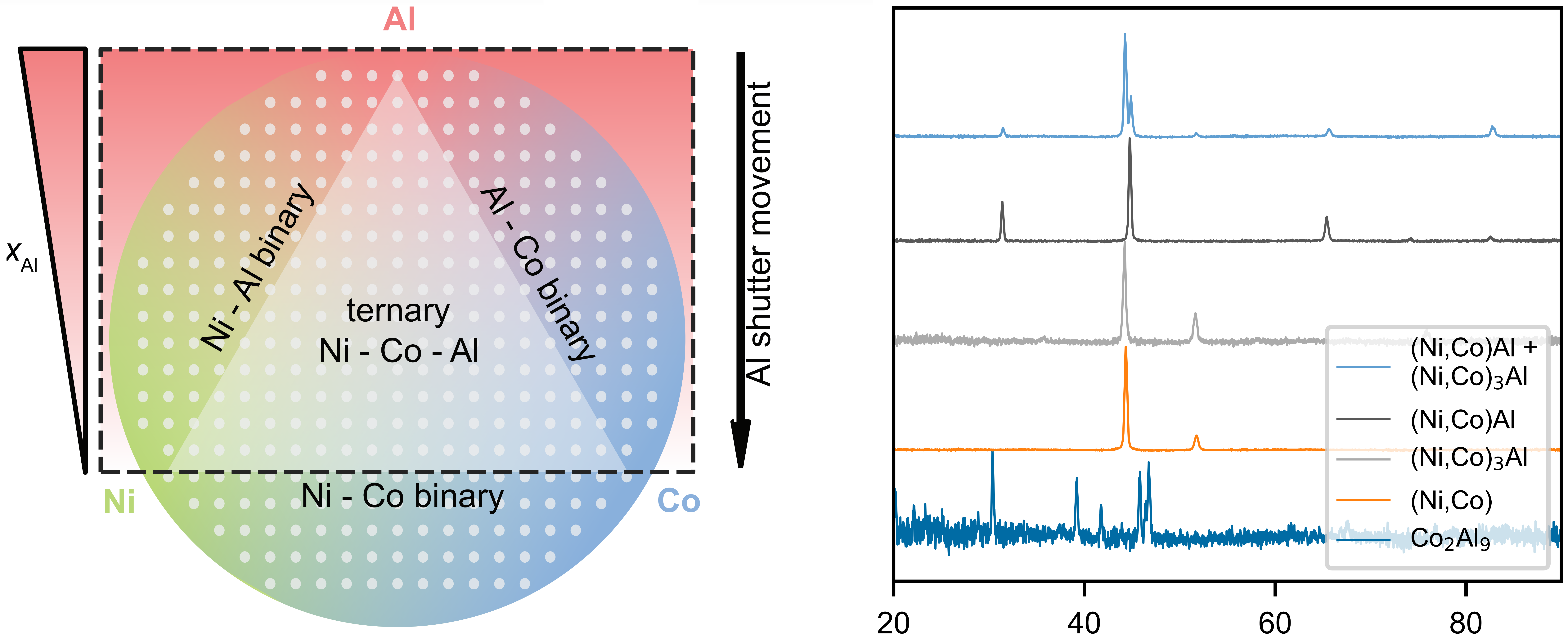






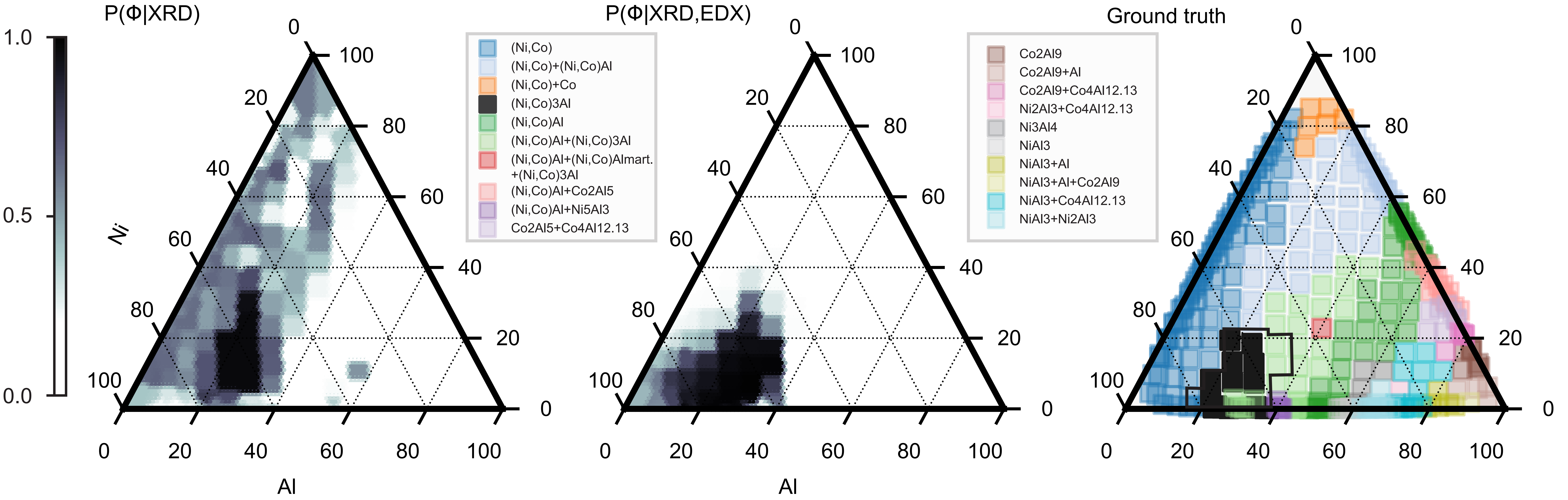
Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

XCA aids phase mapping of a ternary alloy.



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

XCA aids phase mapping of a ternary alloy.

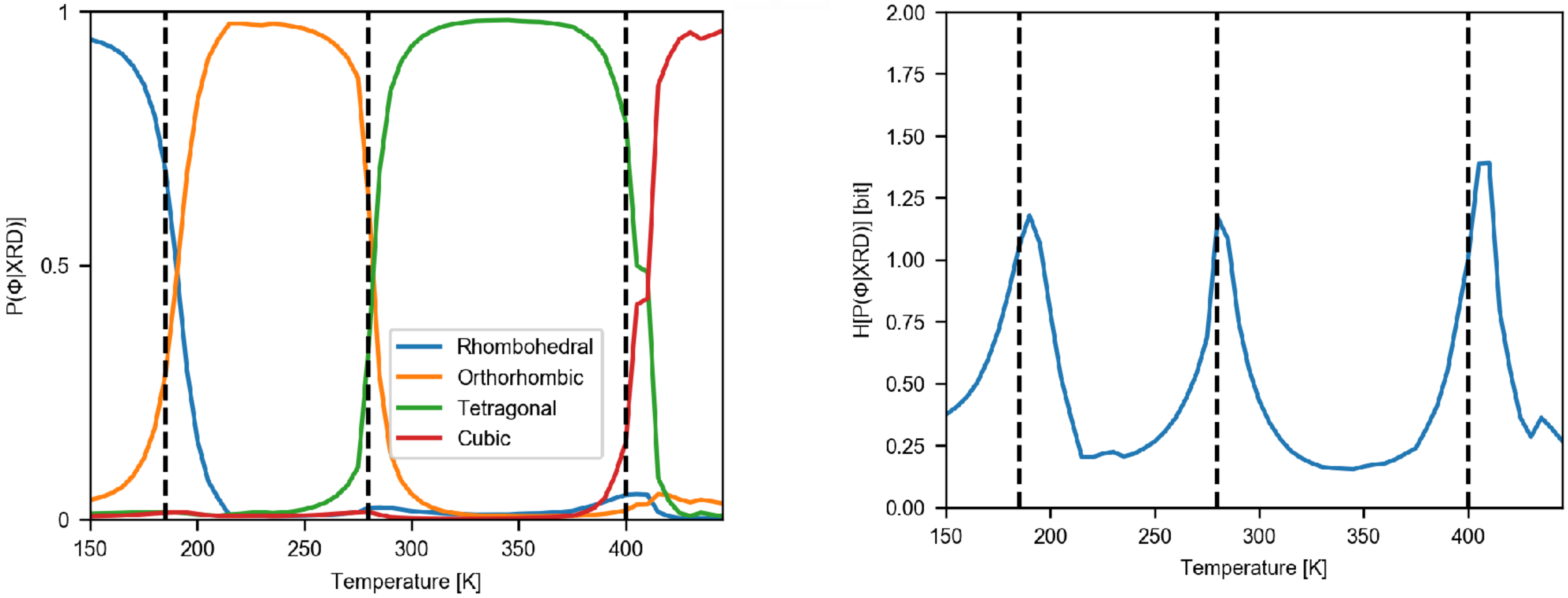


Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

Unsupervised learning:

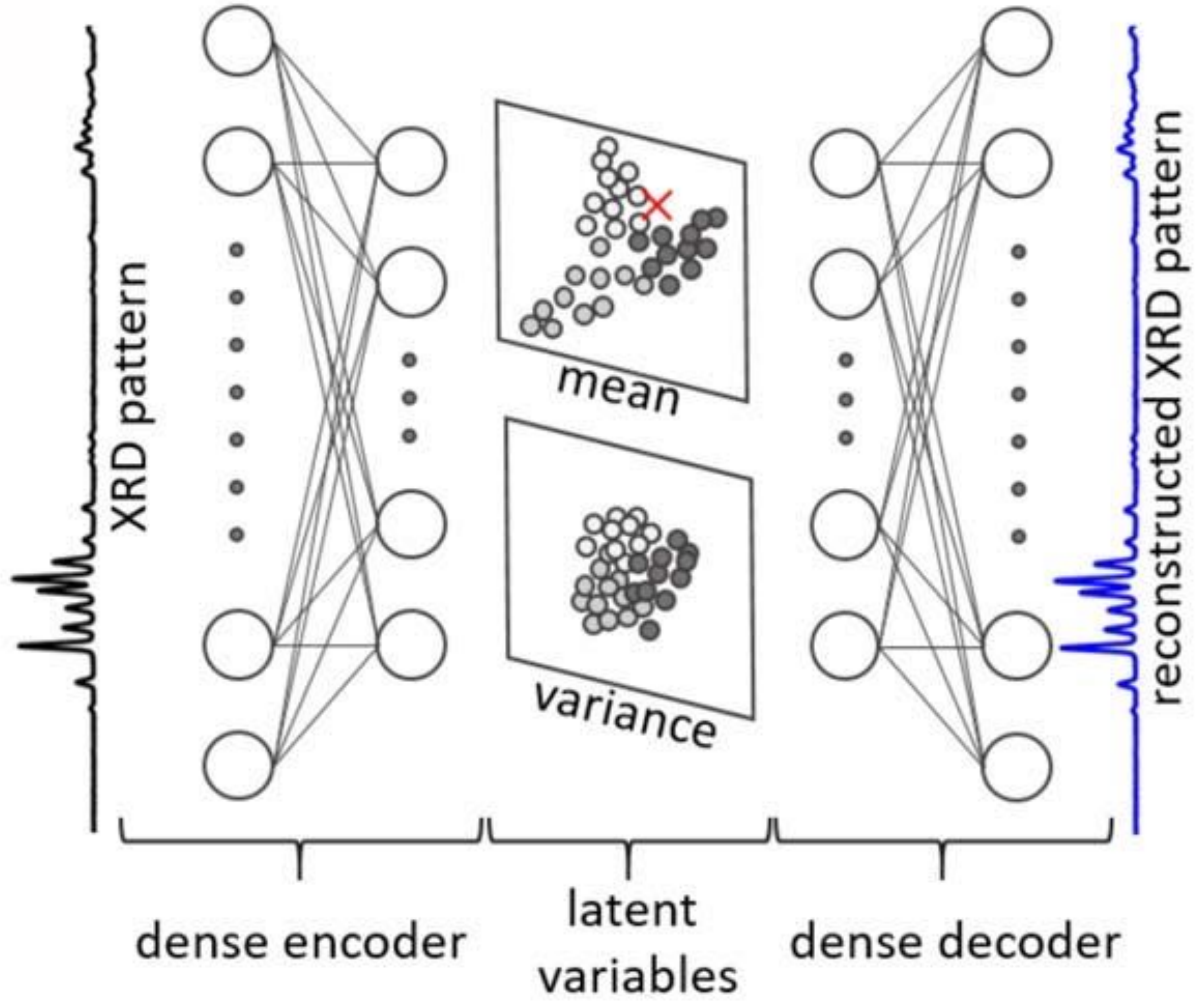
How do we approach situations when we are exploring the unknown?

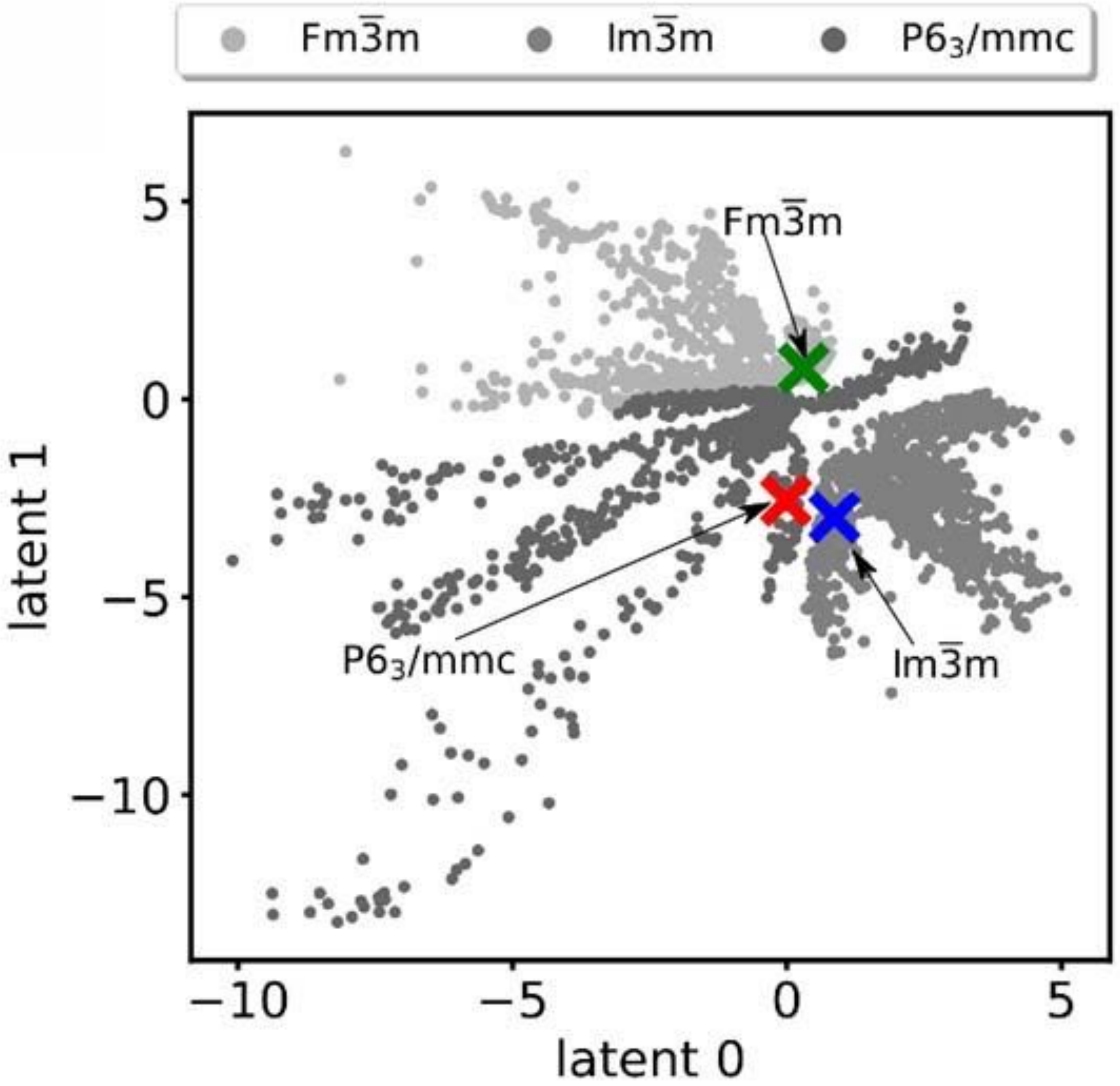
Uncertainty is a proxy for novelty.



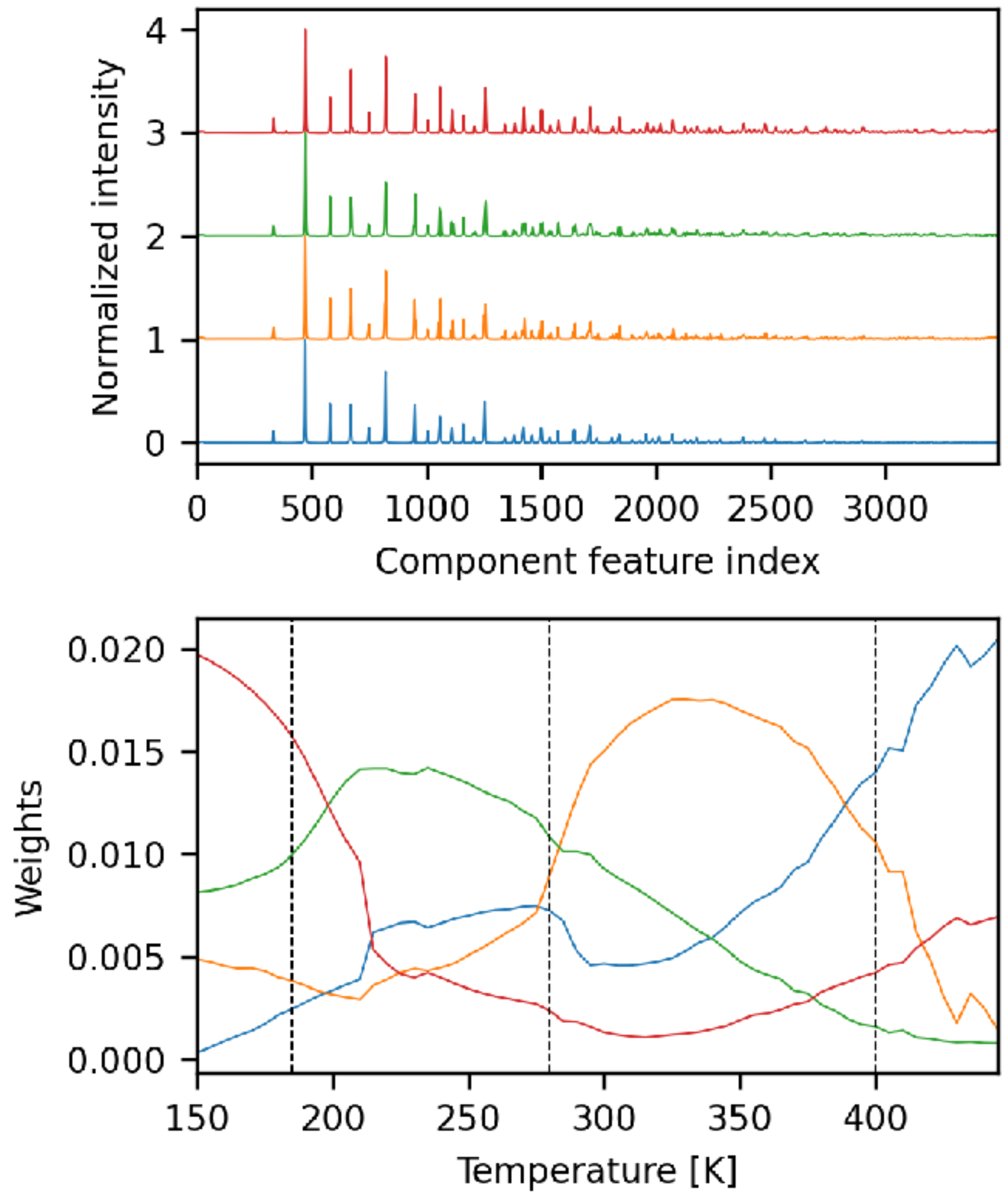
Maffettone, P.M. *et al.*, *Nat. Comp. Sci.*, **1**, 290-297 (2021)

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





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



X ~ **WH**

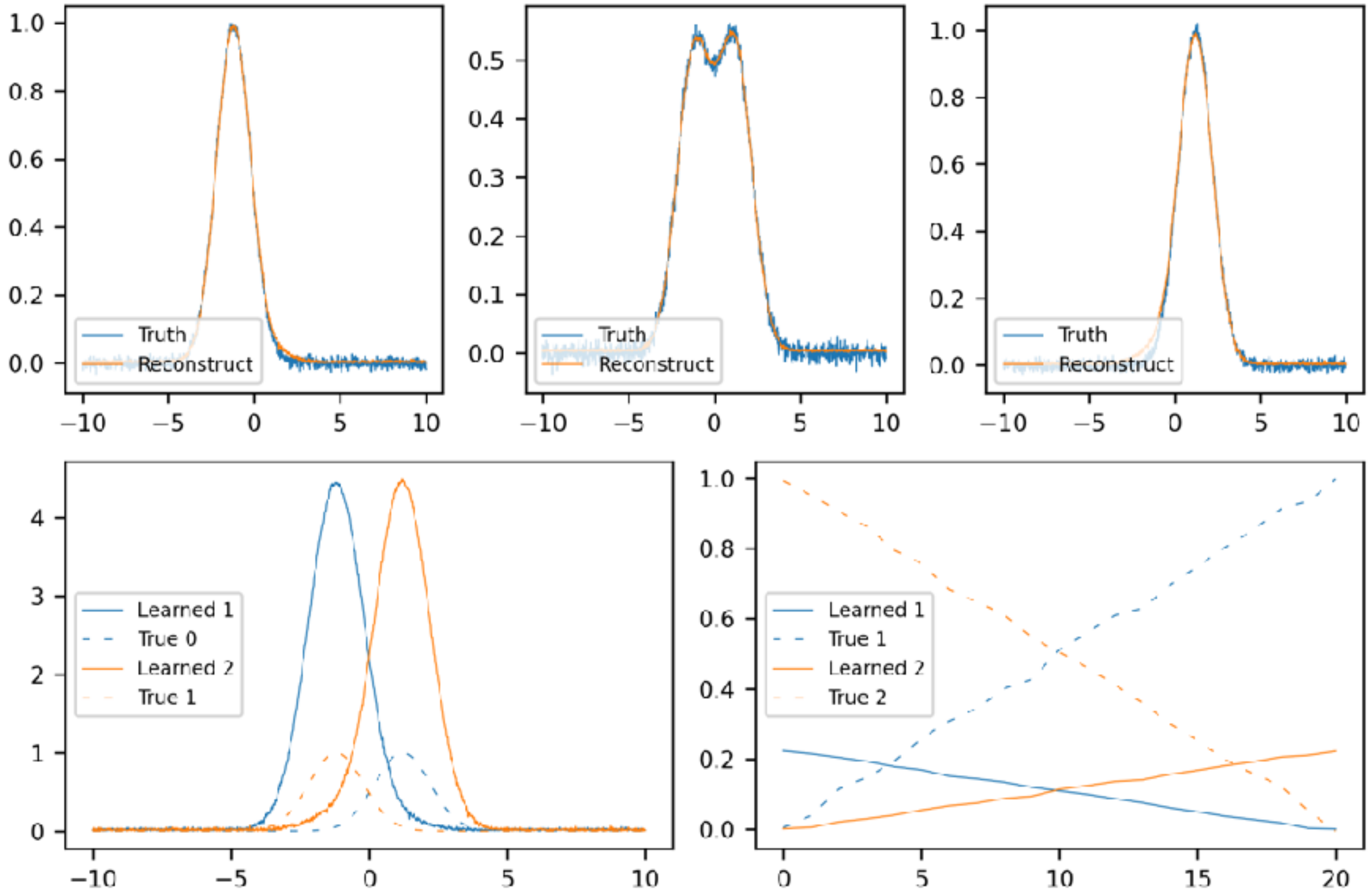
W (m patterns, k components)

H (k components, n features)

Appl. Phys. Rev. **Accepted**

We don't want the model that best fits the data, but the *most likely* model that best fits the data.

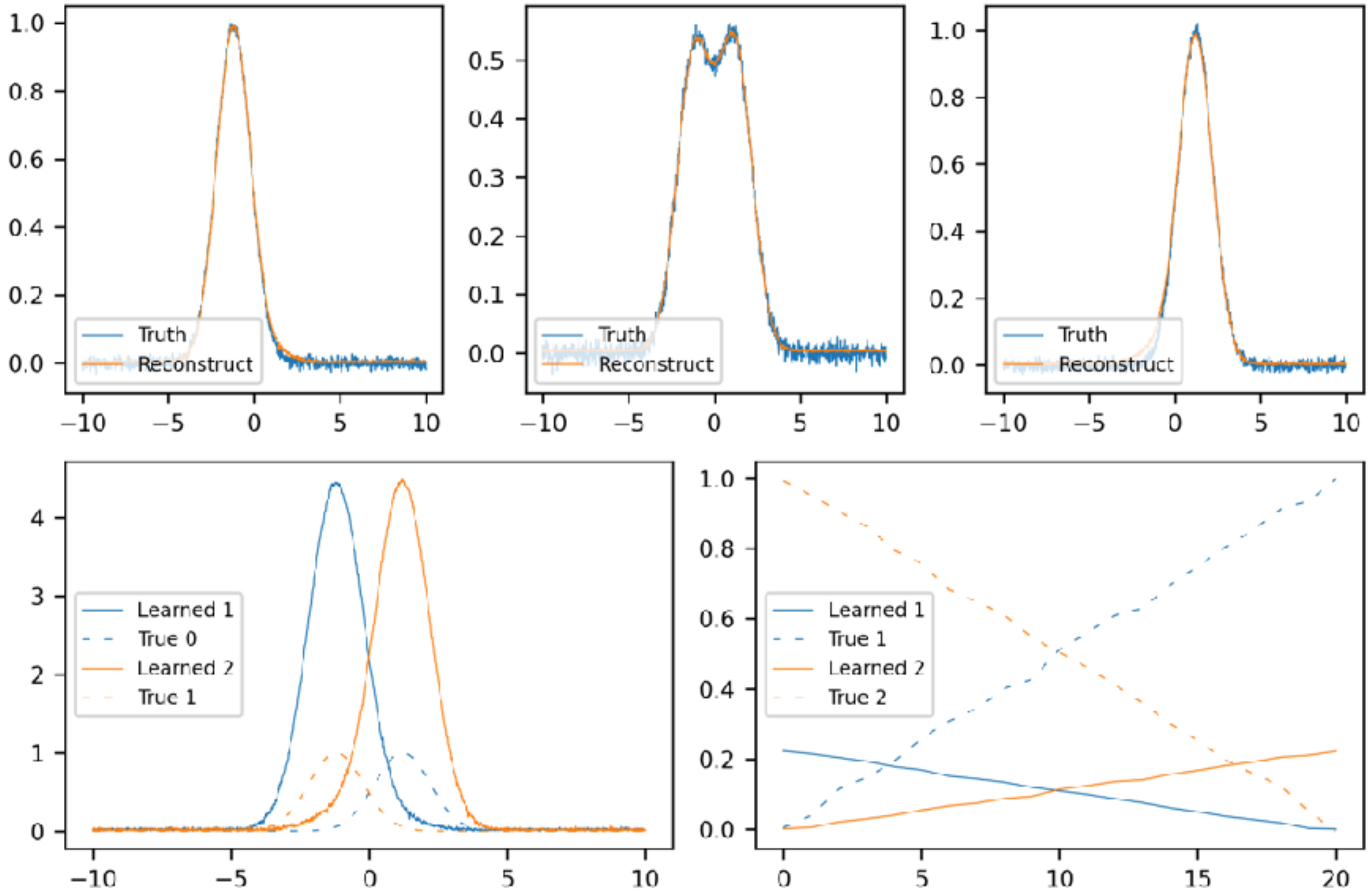
Canonical



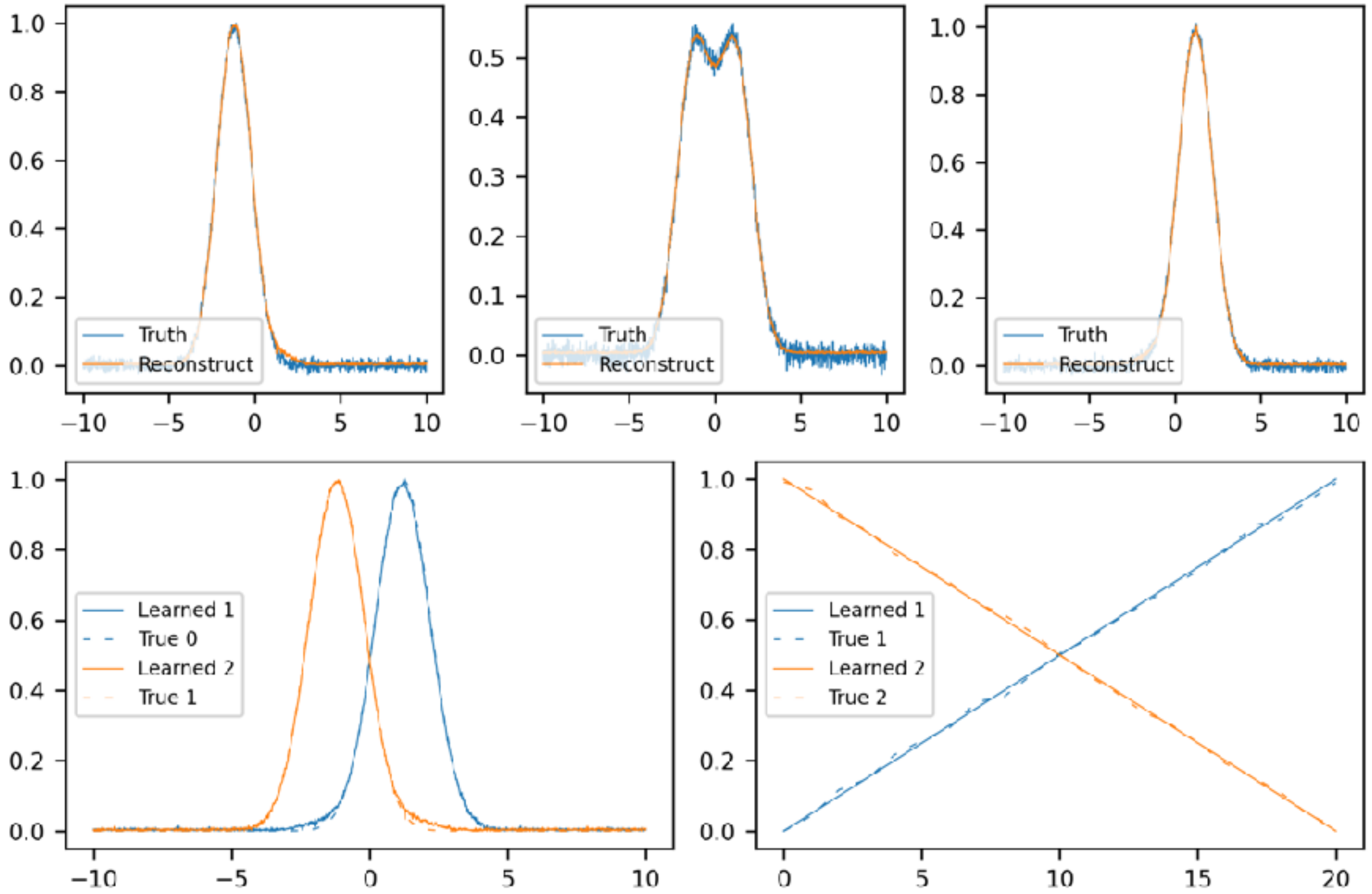
Appl. Phys. Rev. **Accepted**

We don't want the model that best fits the data, but the *most likely* model that best fits the data.

Canonical

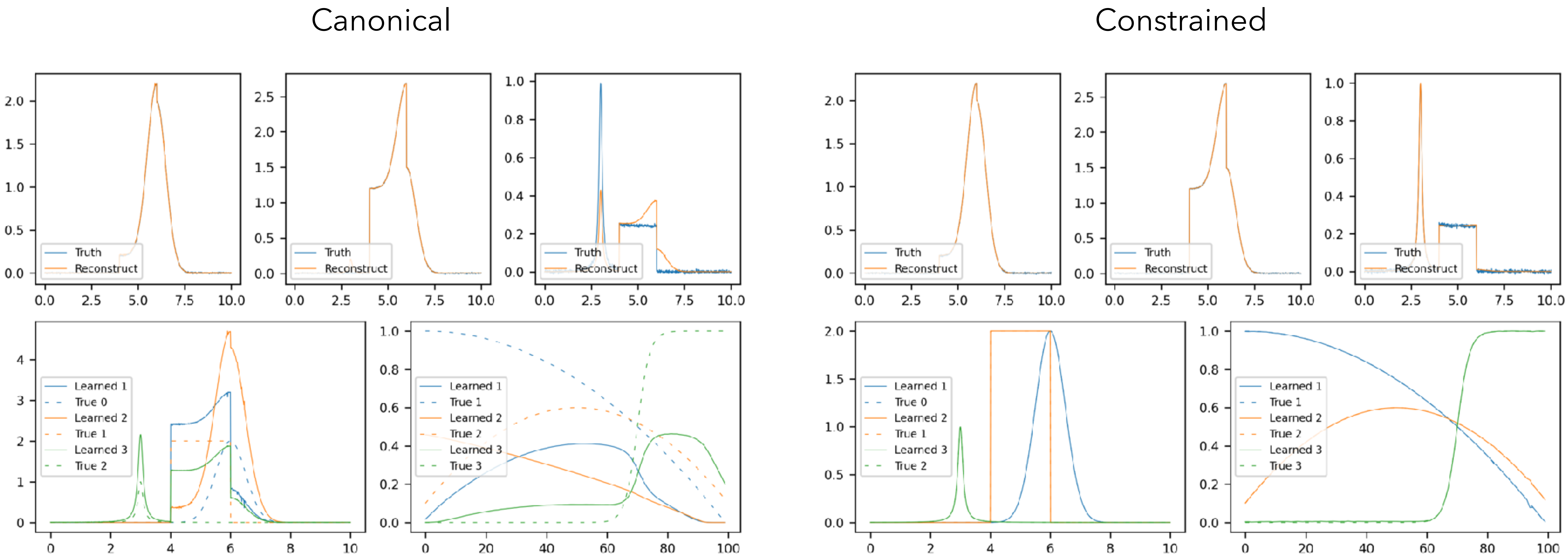


Constrained



Appl. Phys. Rev. **Accepted**

We don’t want the model that best fits the data, but the *most likely* model that best fits the data.

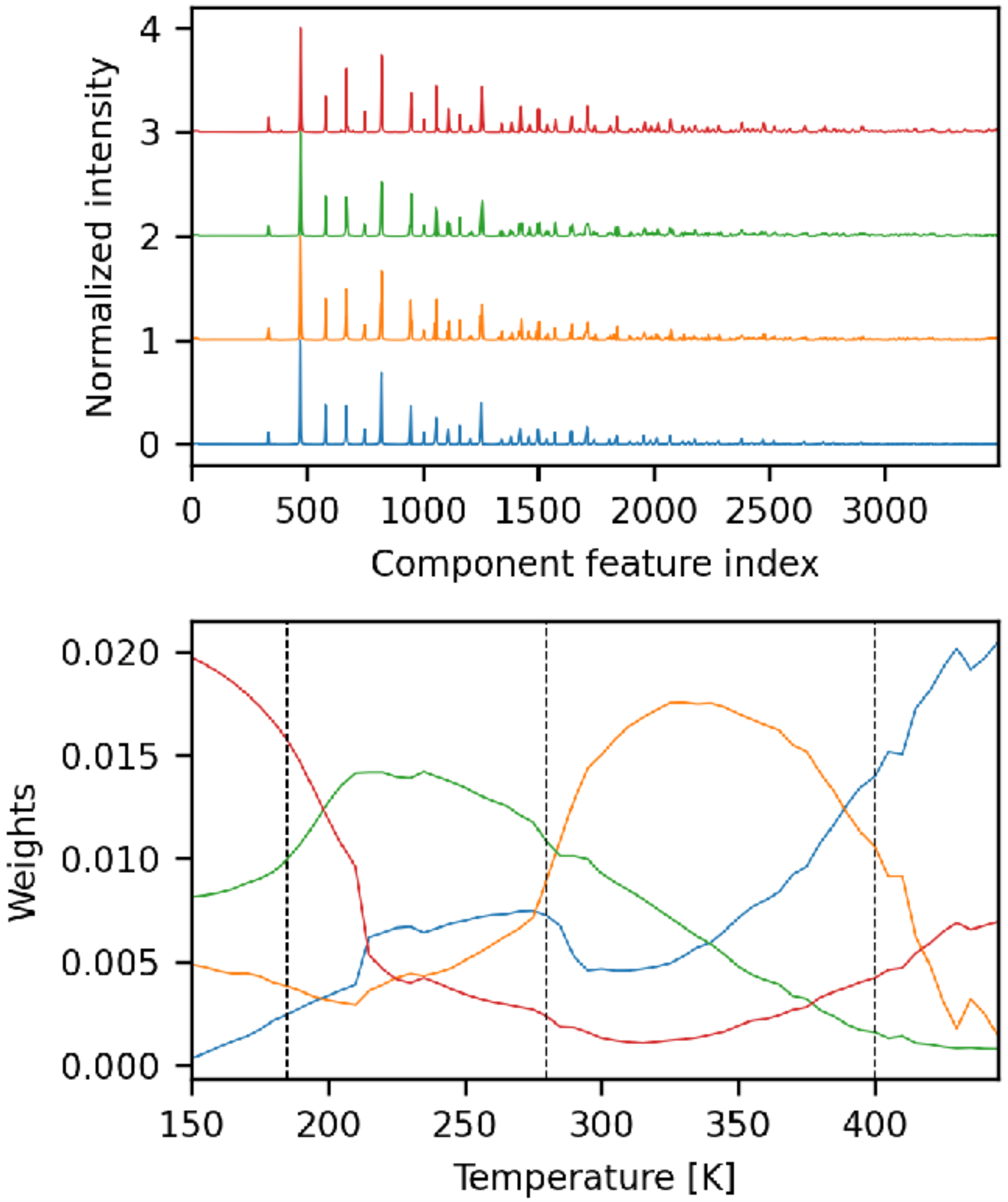


Appl. Phys. Rev. **Accepted**

U.S. DEPARTMENT OF ENERGY | Office of Science
BROOKHAVEN NATIONAL LABORATORY
FLATIRON INSTITUTE
National Synchrotron Light Source II

Constrained NMF produces physically realistic components and weights.

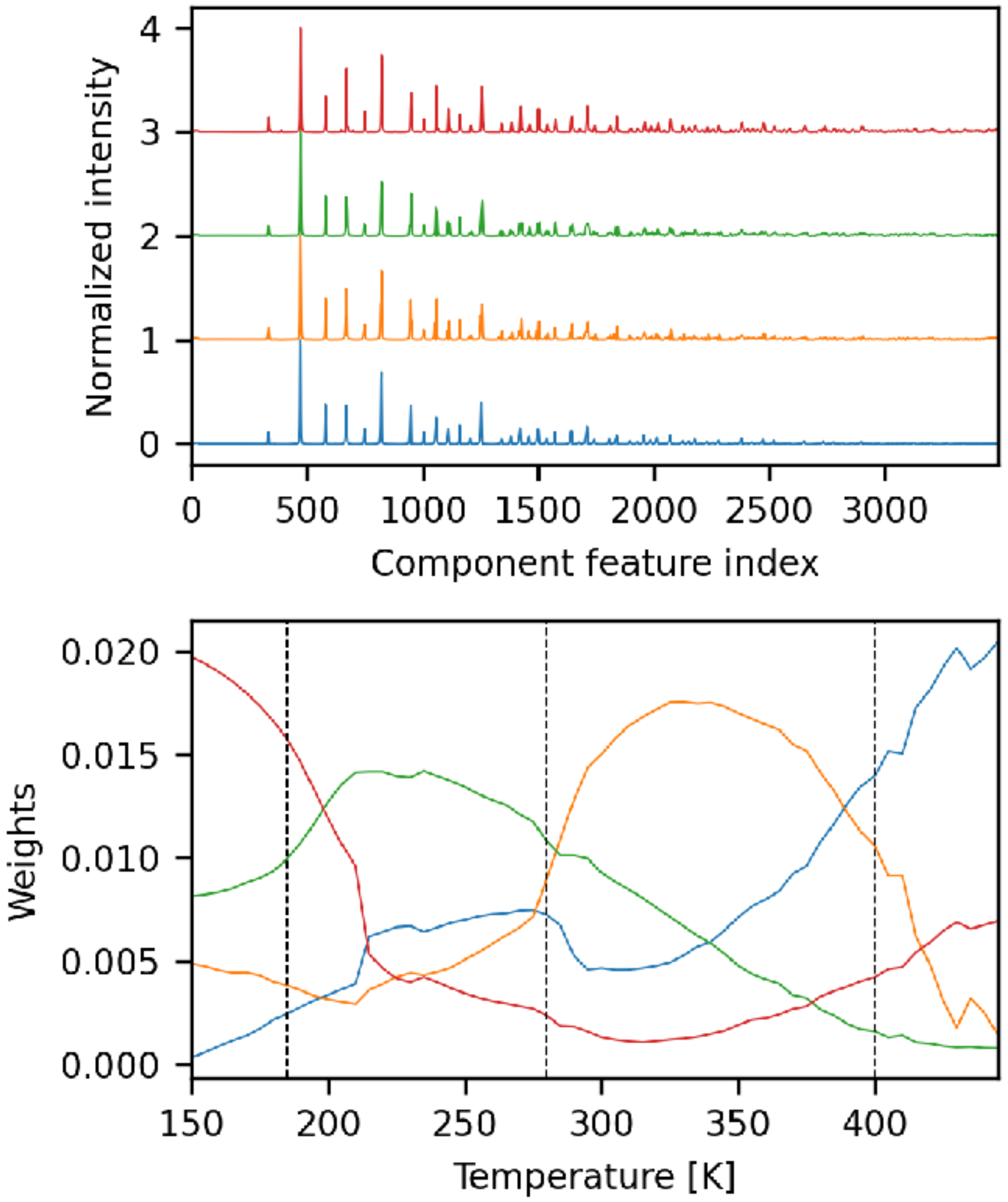
Canonical



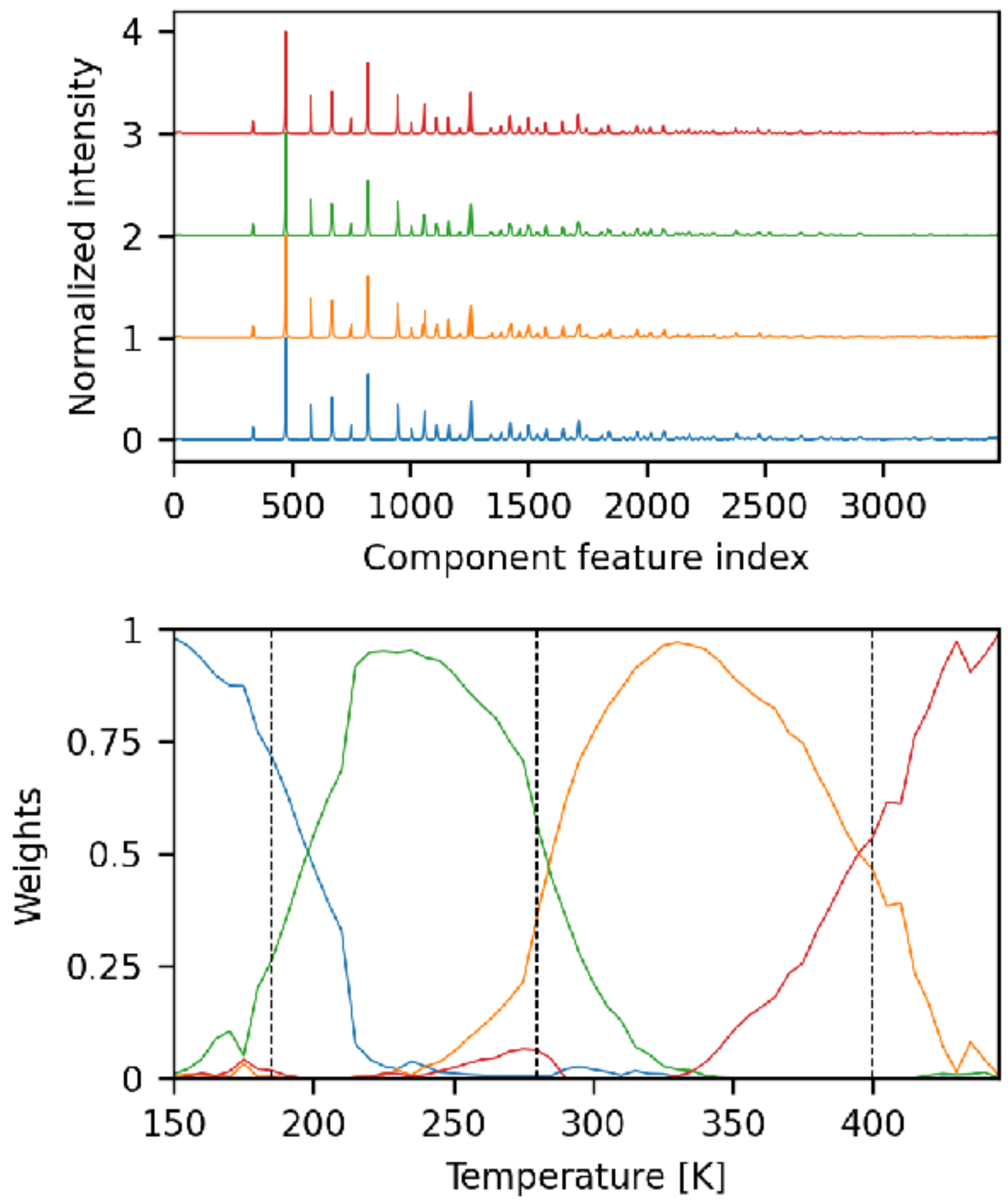
Appl. Phys. Rev. **Accepted**

Constrained NMF produces physically realistic components and weights.

Canonical

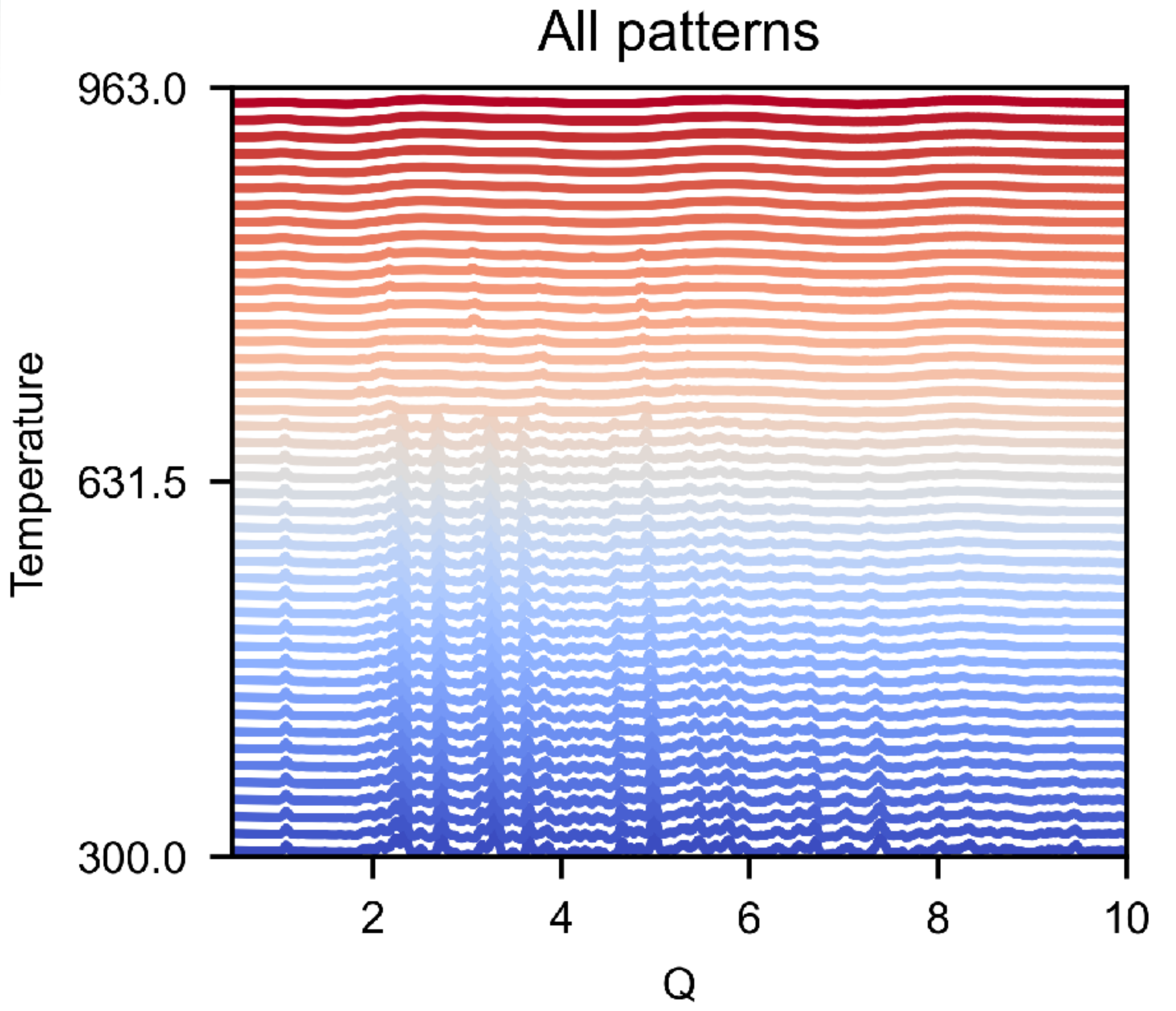


Constrained



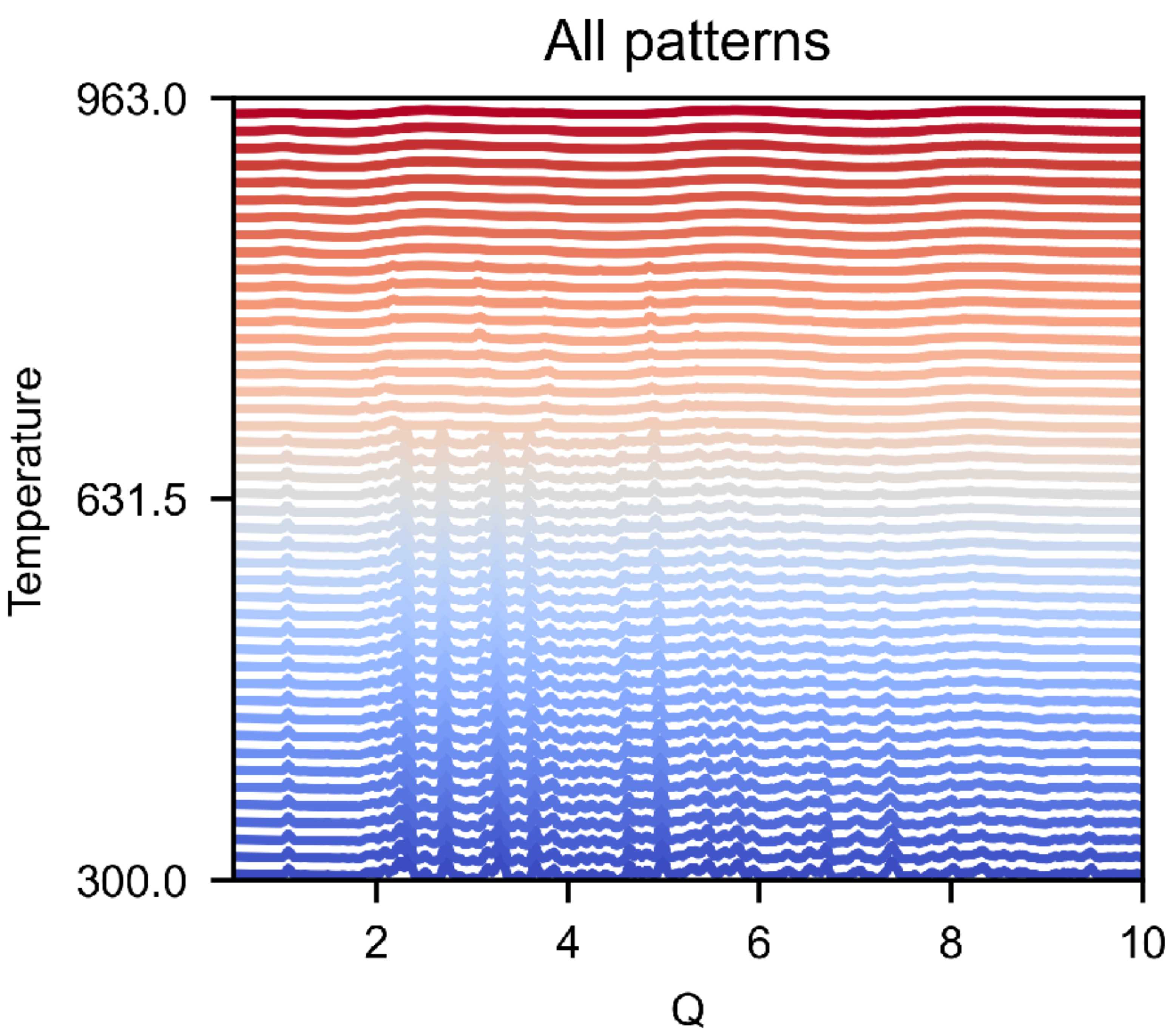
Appl. Phys. Rev. **Accepted**

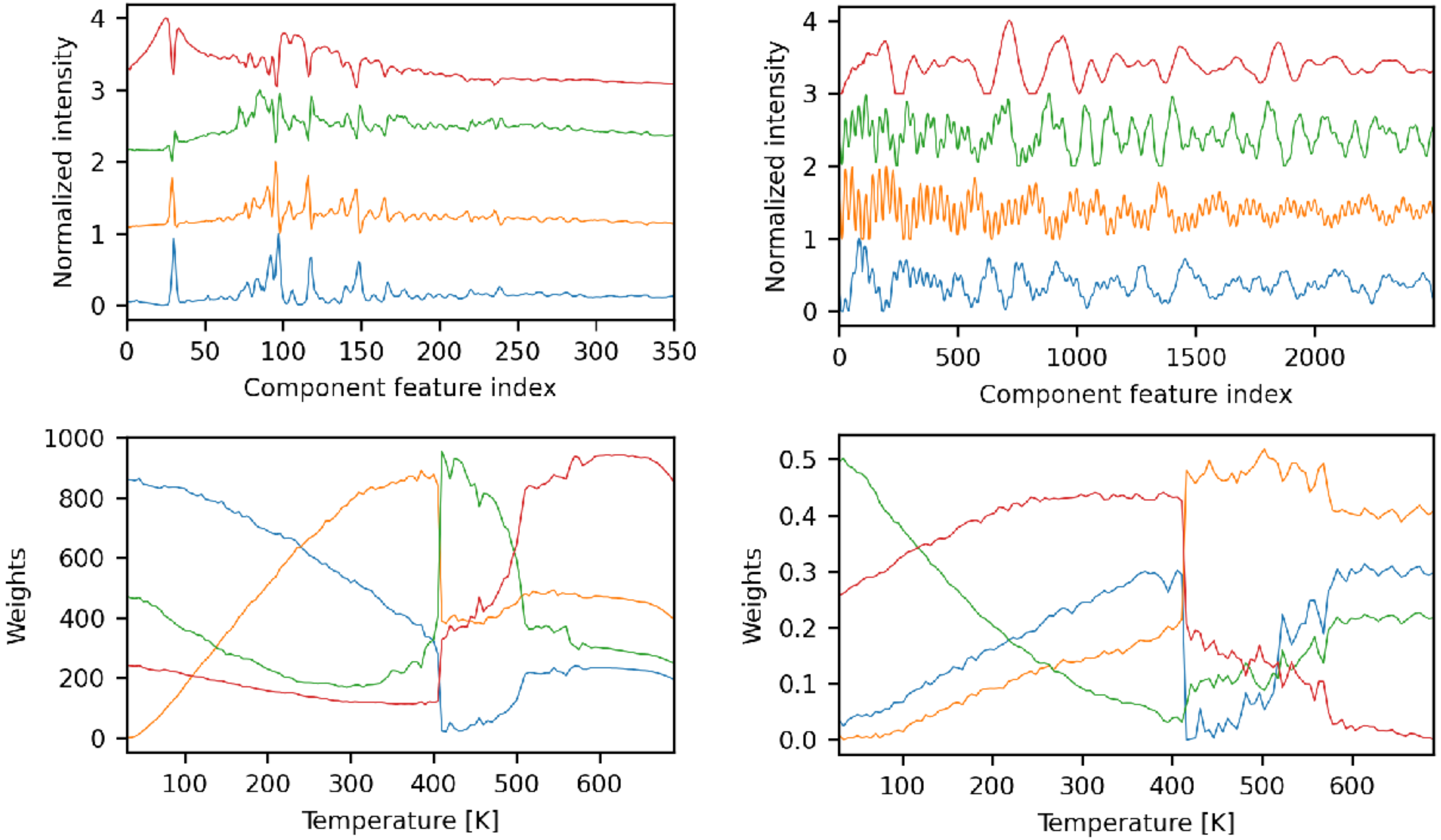
Canonical NMF leads to confusing results across a melting system.



Appl. Phys. Rev. **Accepted**

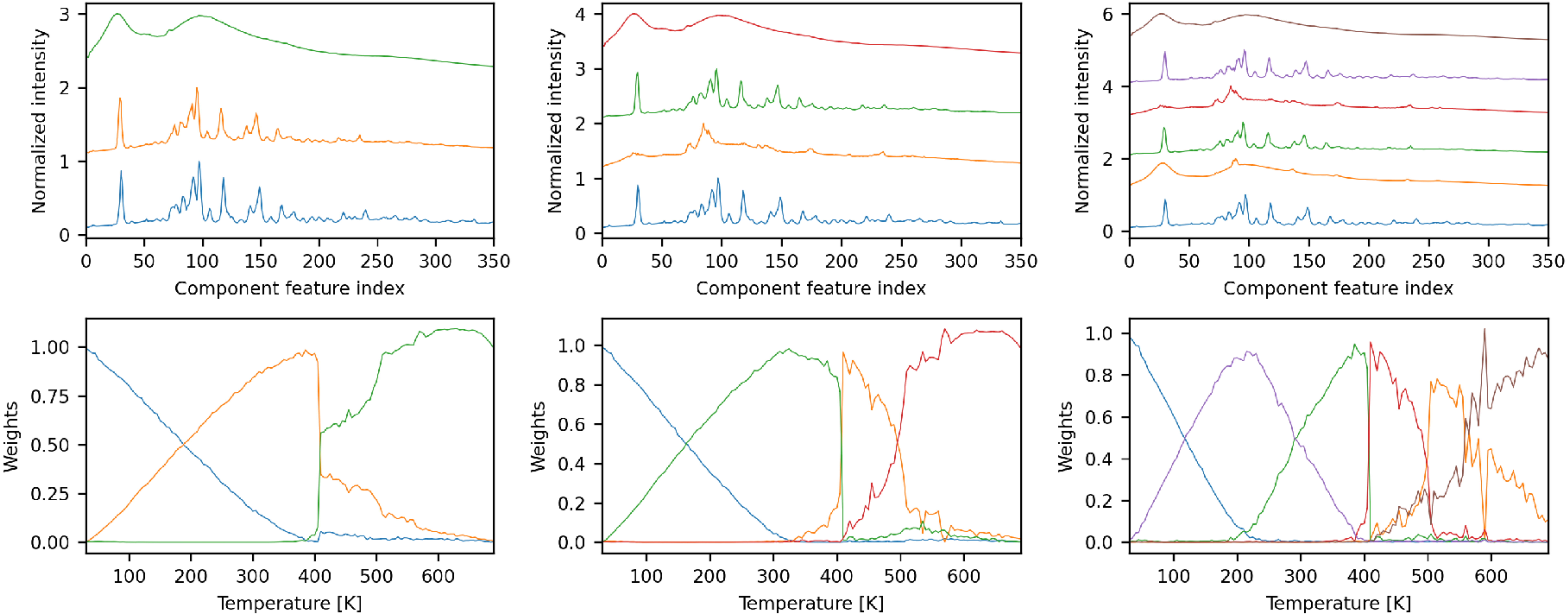
Canonical NMF leads to confusing results across a melting system.





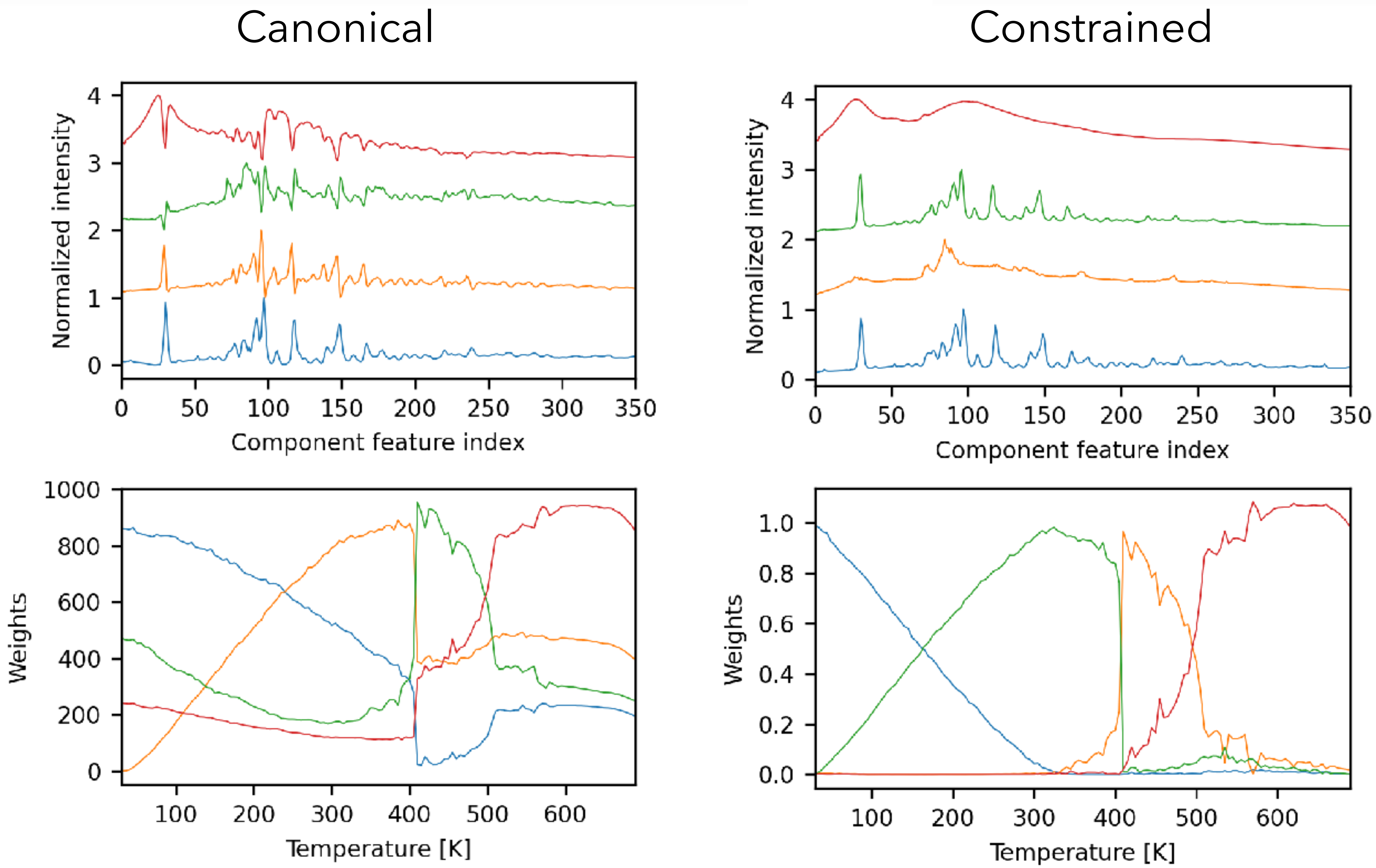
Appl. Phys. Rev. **Accepted**

Fast decomposition on-the-fly of a streaming dataset over a state variable.



Appl. Phys. Rev. **Accepted**

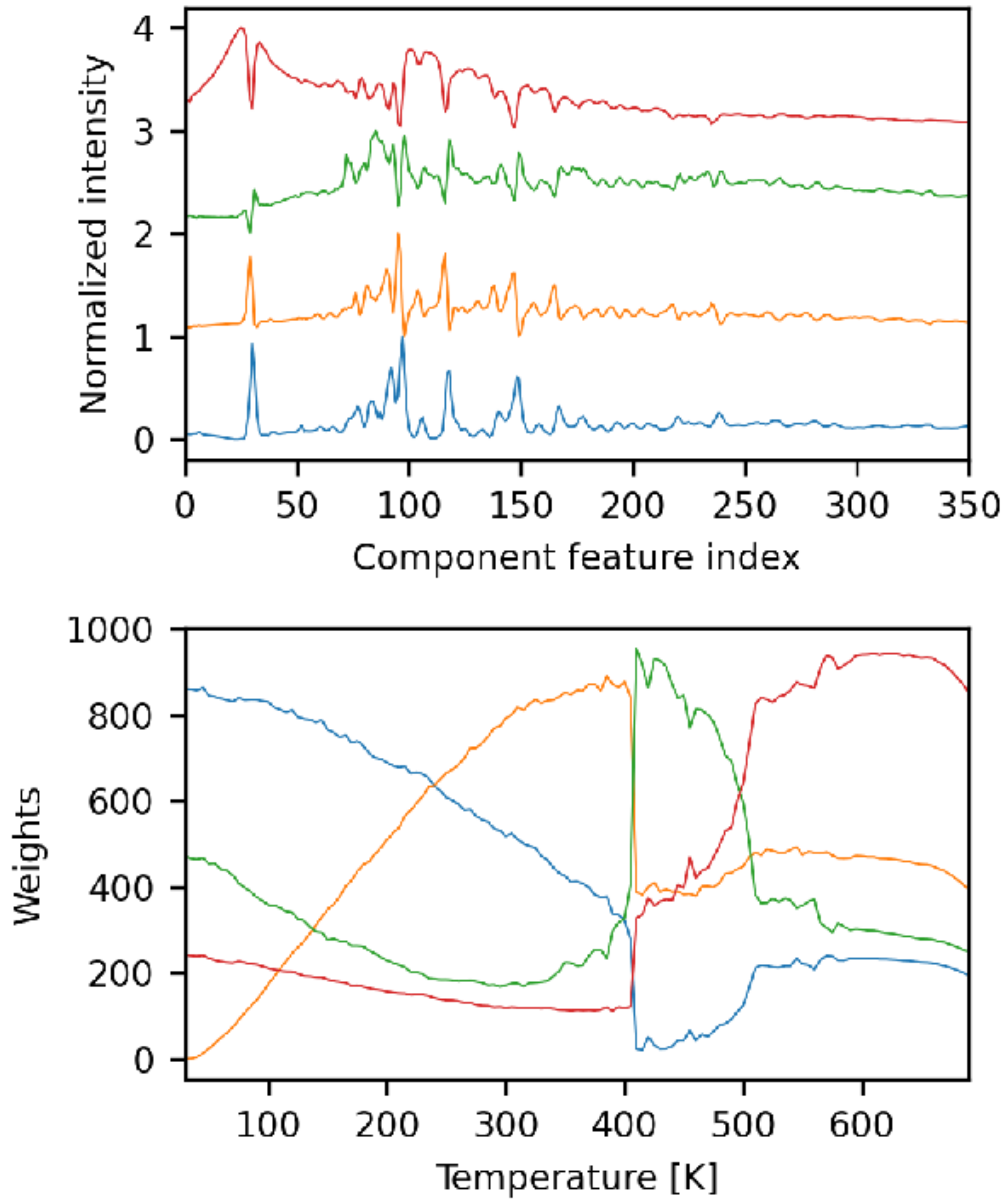
Dynamically adjusted constraints leads to directly interpretable decompositions



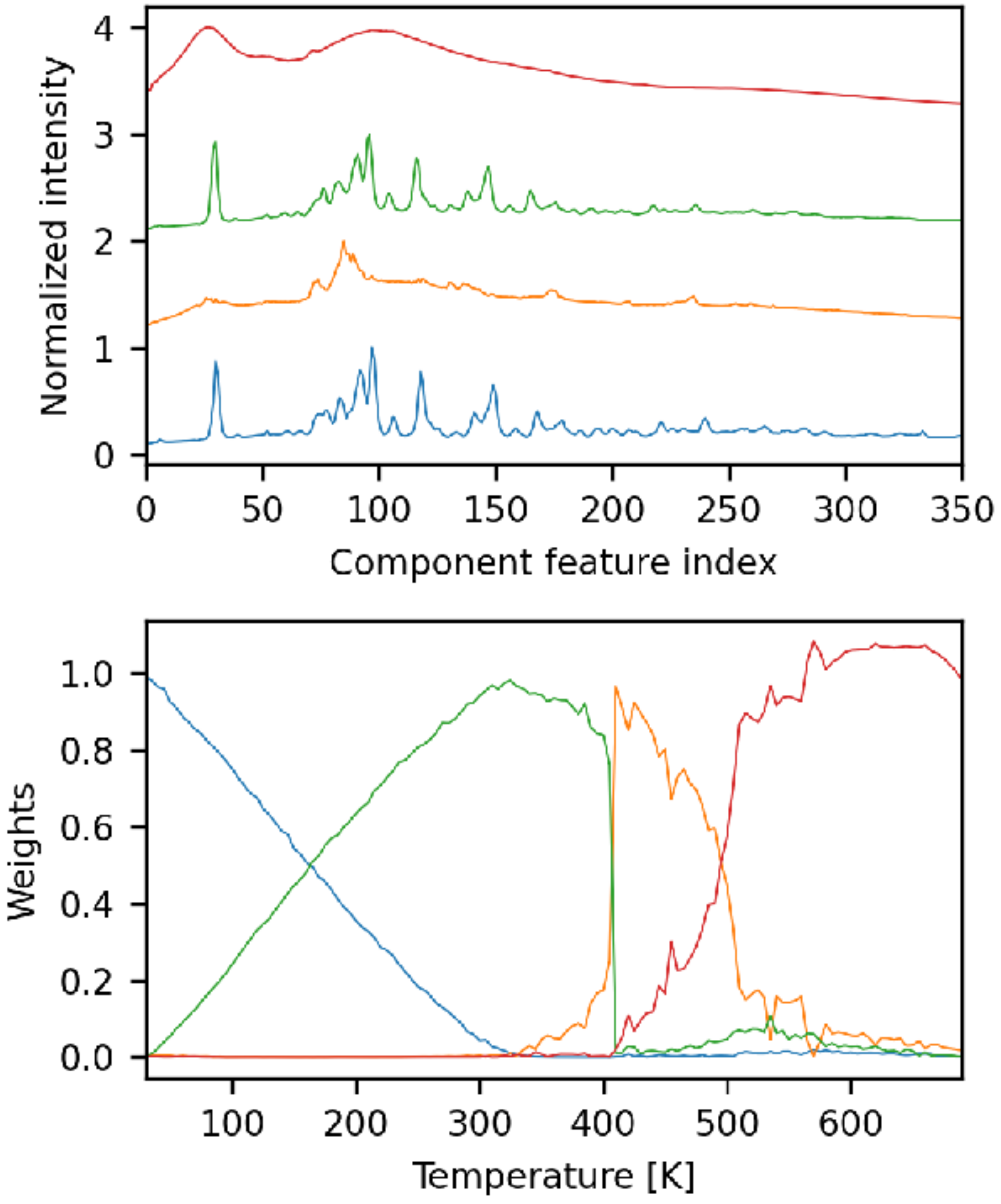
Appl. Phys. Rev. **Accepted**

Dynamically adjusted constraints leads to directly interpretable decompositions

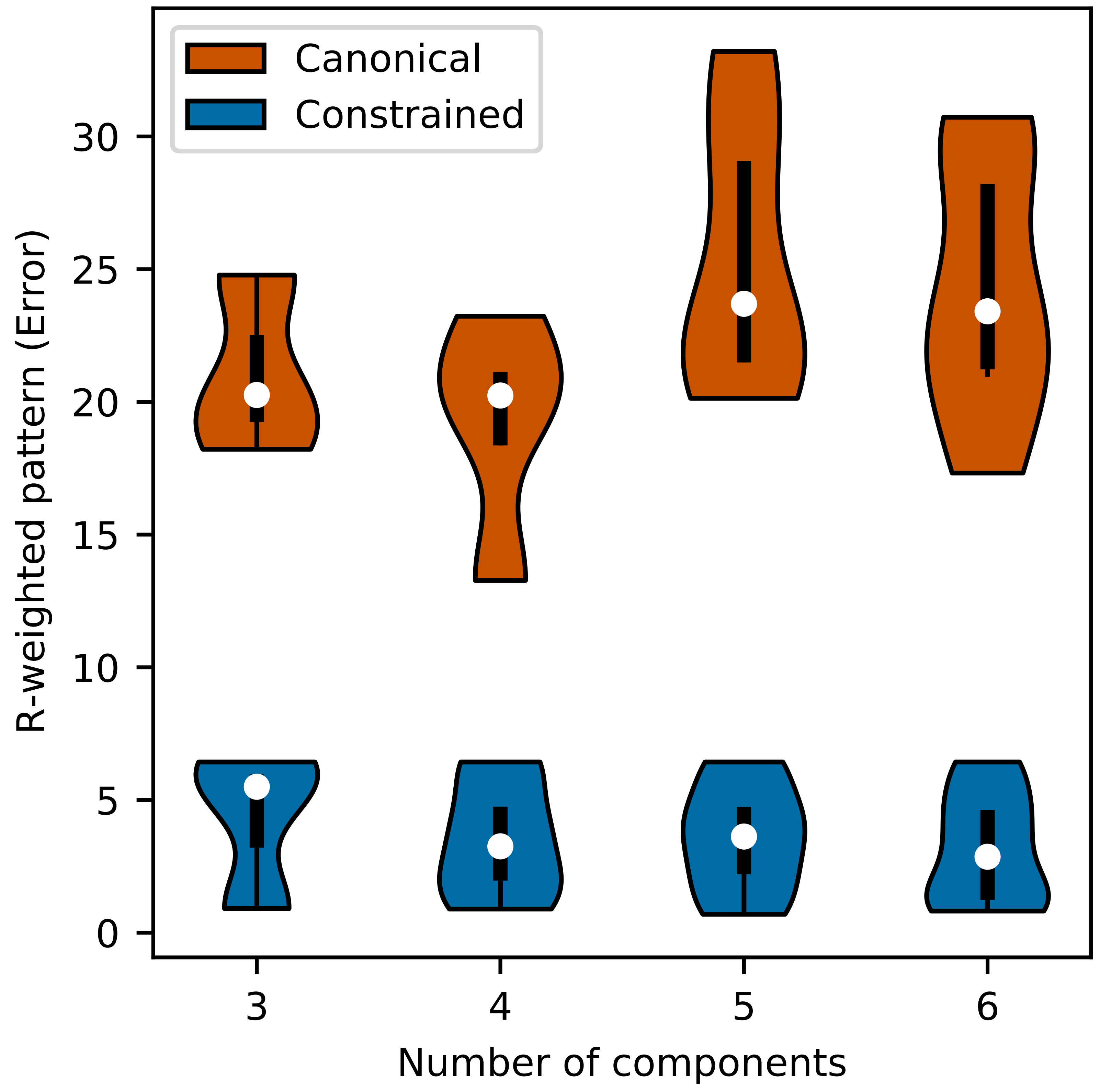
Canonical



Constrained



Refinement Results



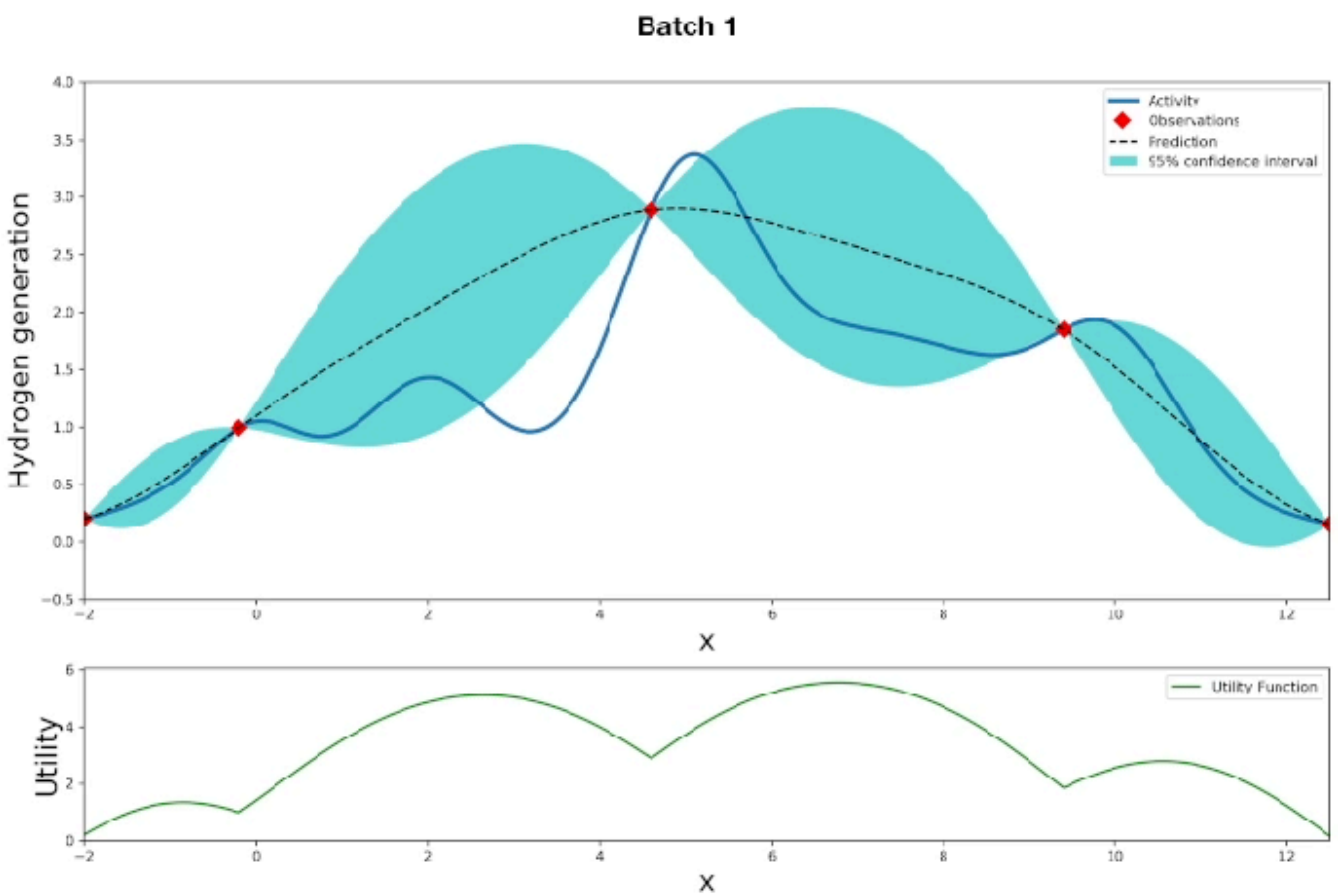
Appl. Phys. Rev. **Accepted**

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.

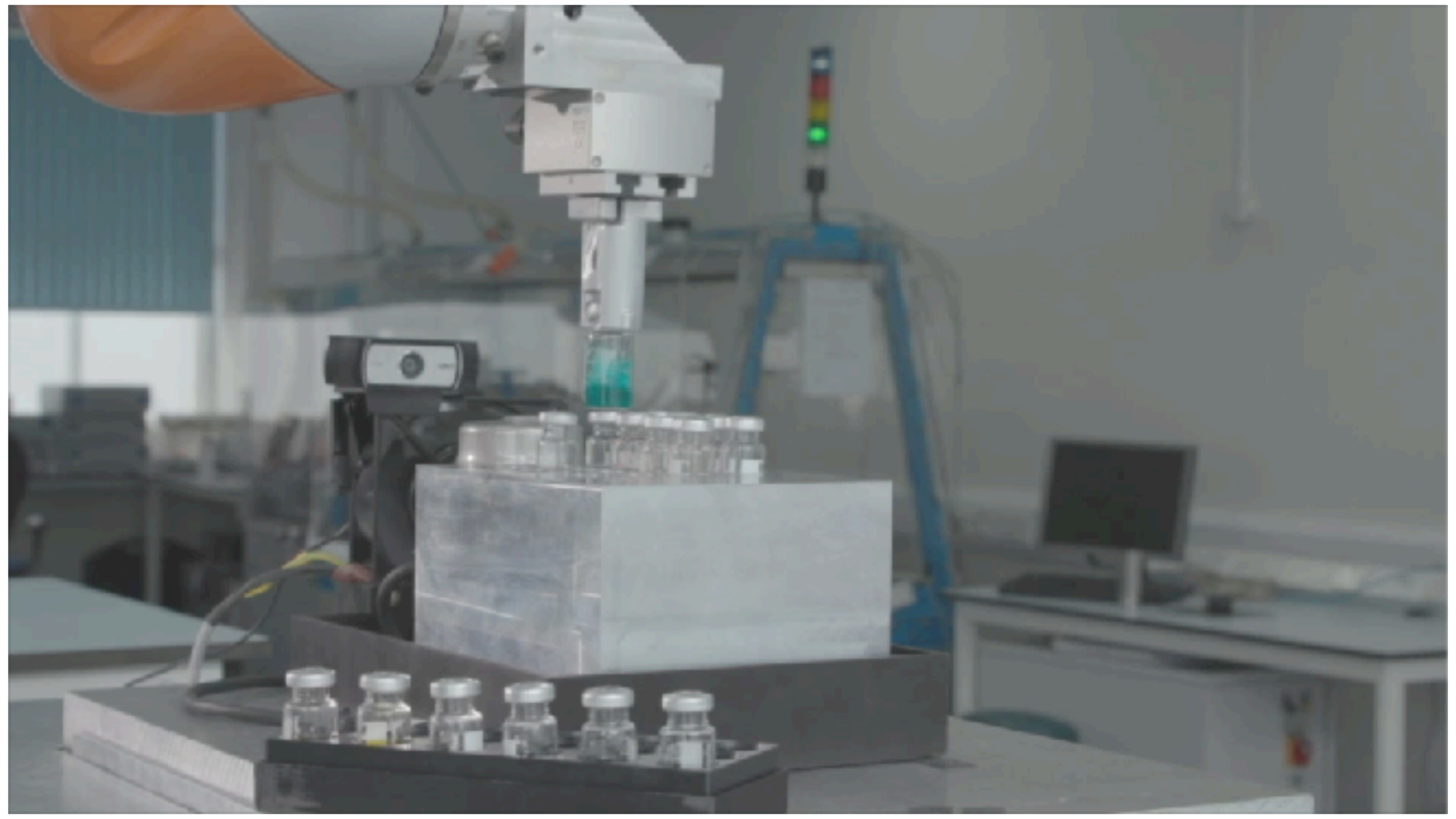
Nature **583**, 237-241 (2020)

Autonomous discovery

Solid dispensing

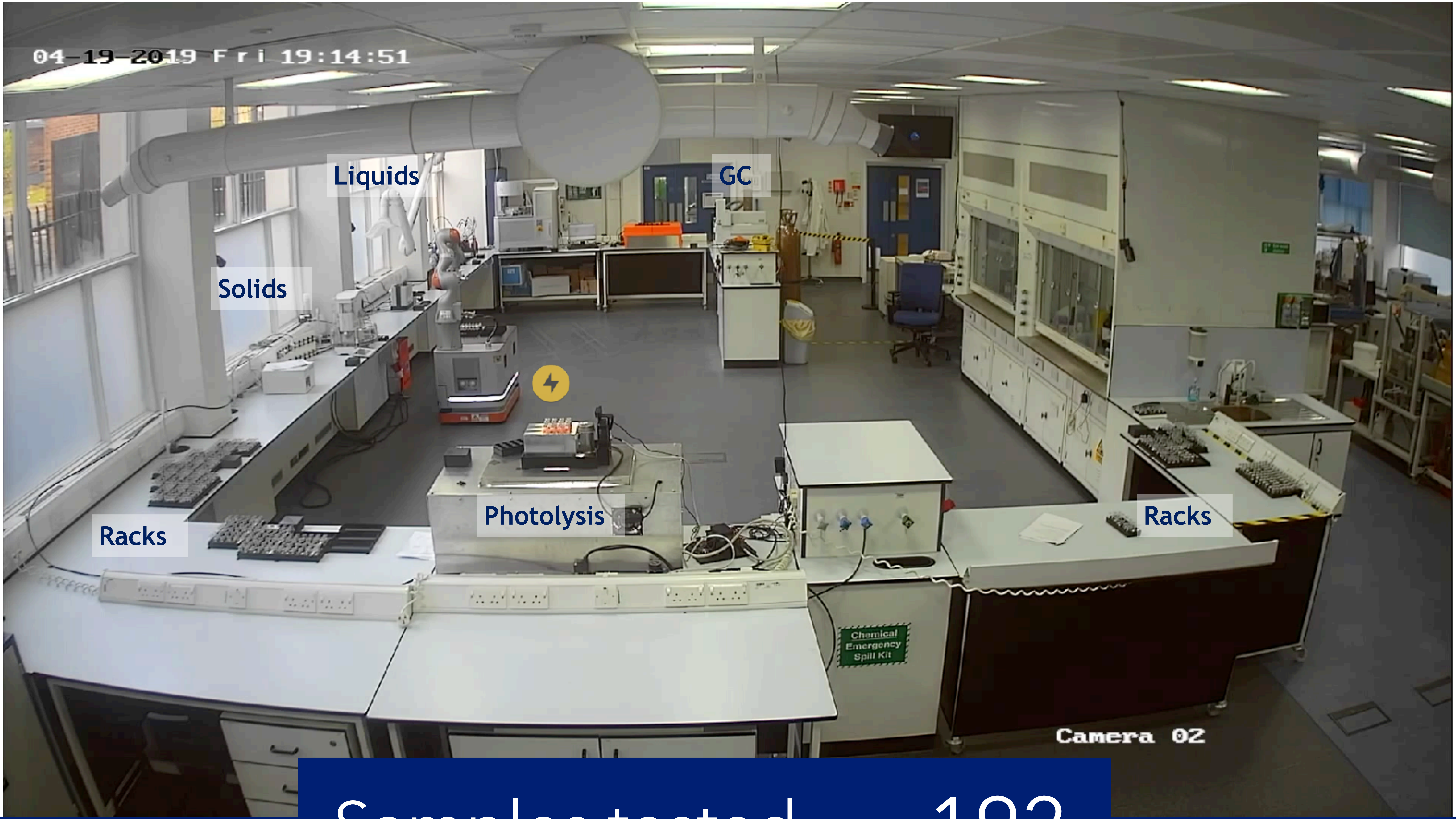
Liquid dispensing
Inertization

Photolysis



Measurement

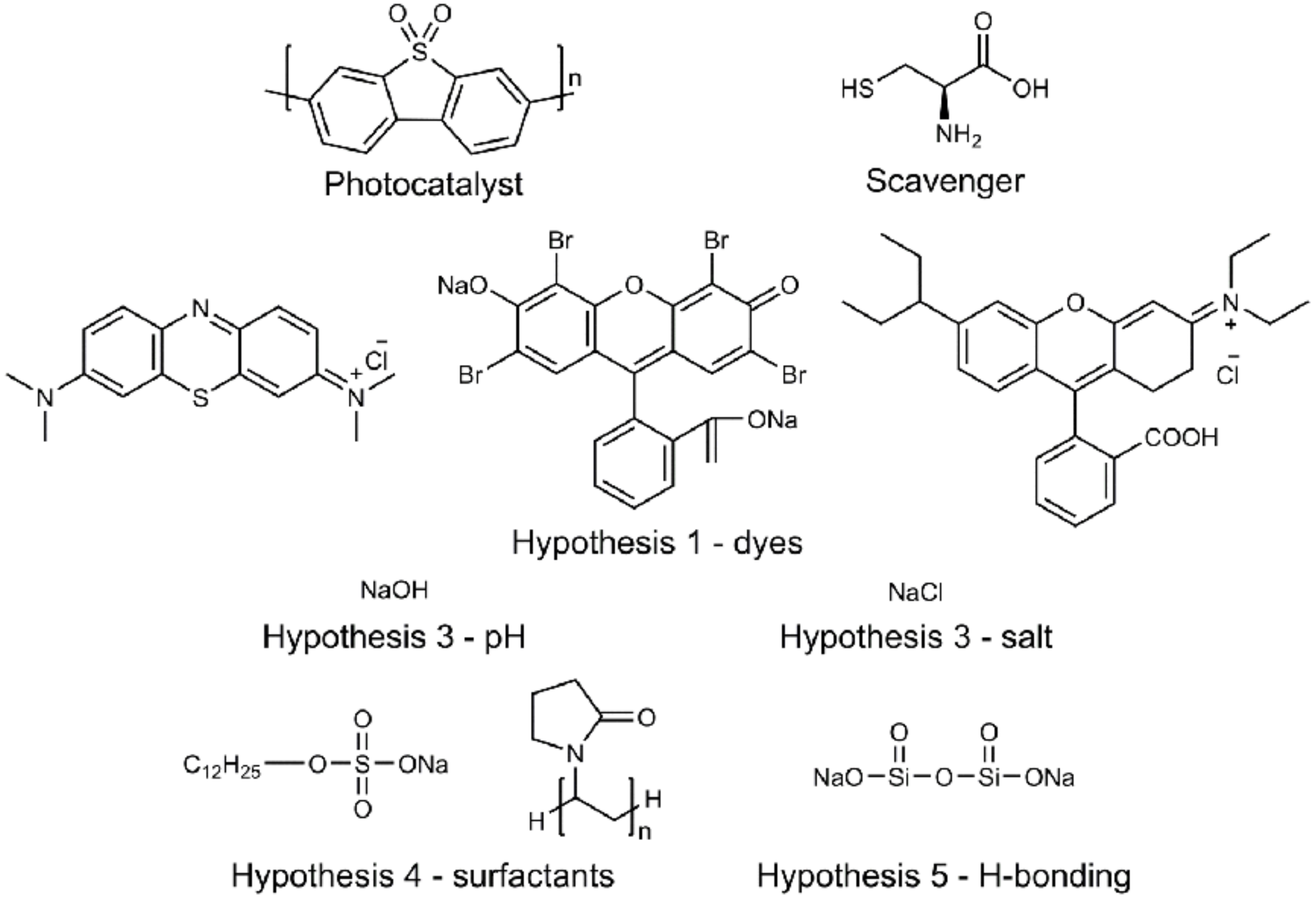
48 hours of research in 60 seconds

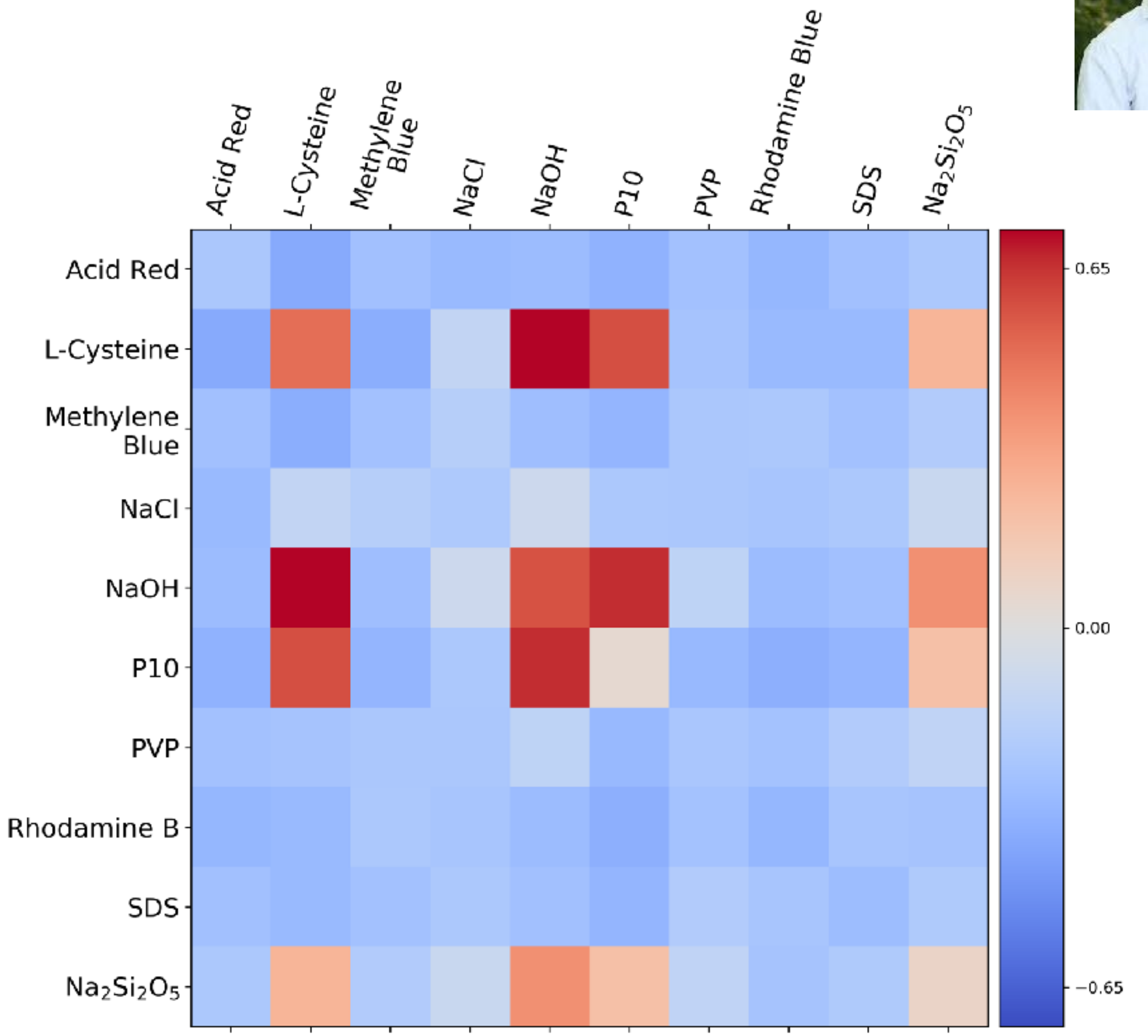


Samples tested = 192

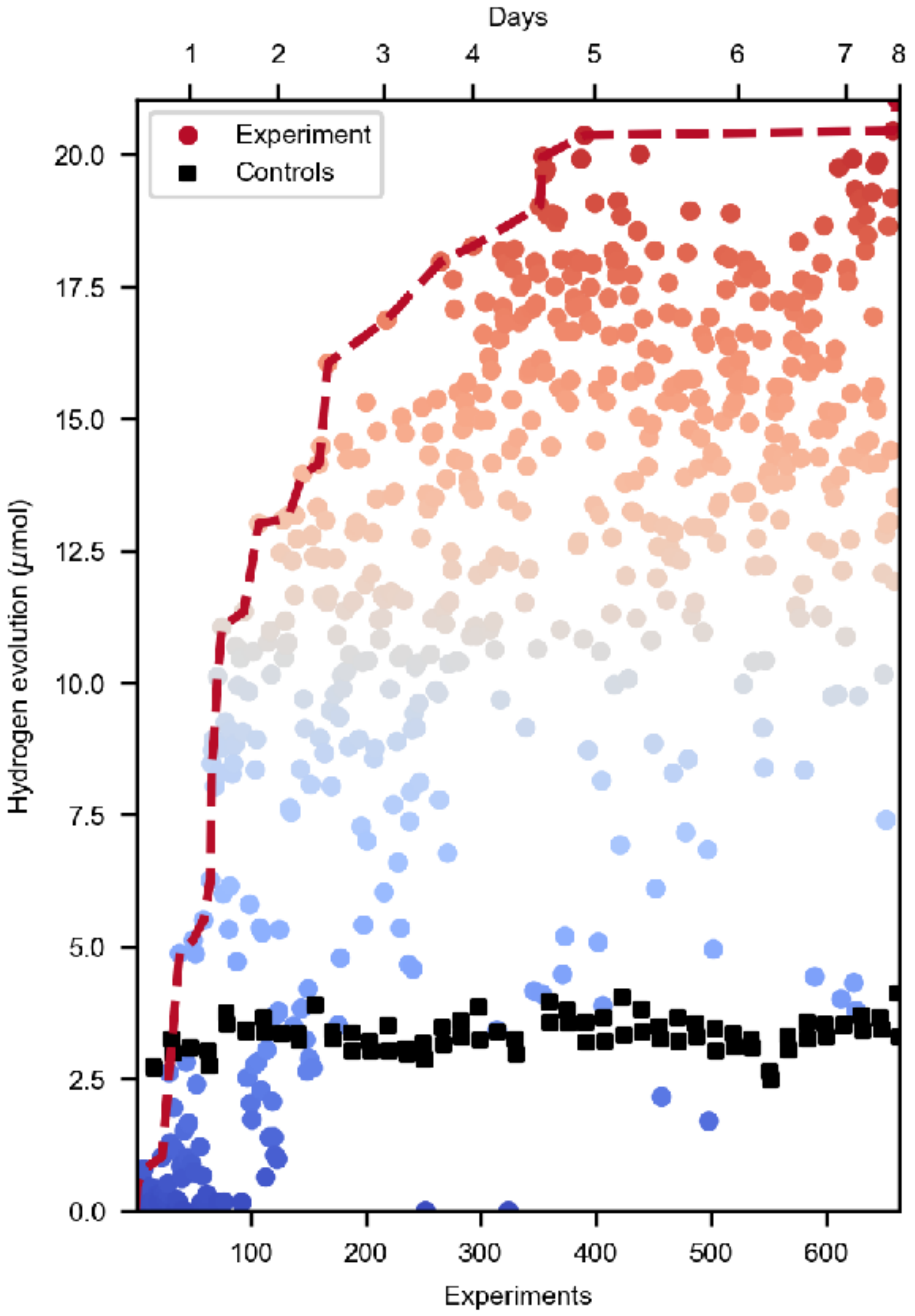
Nature **583**, 237-241 (2020).

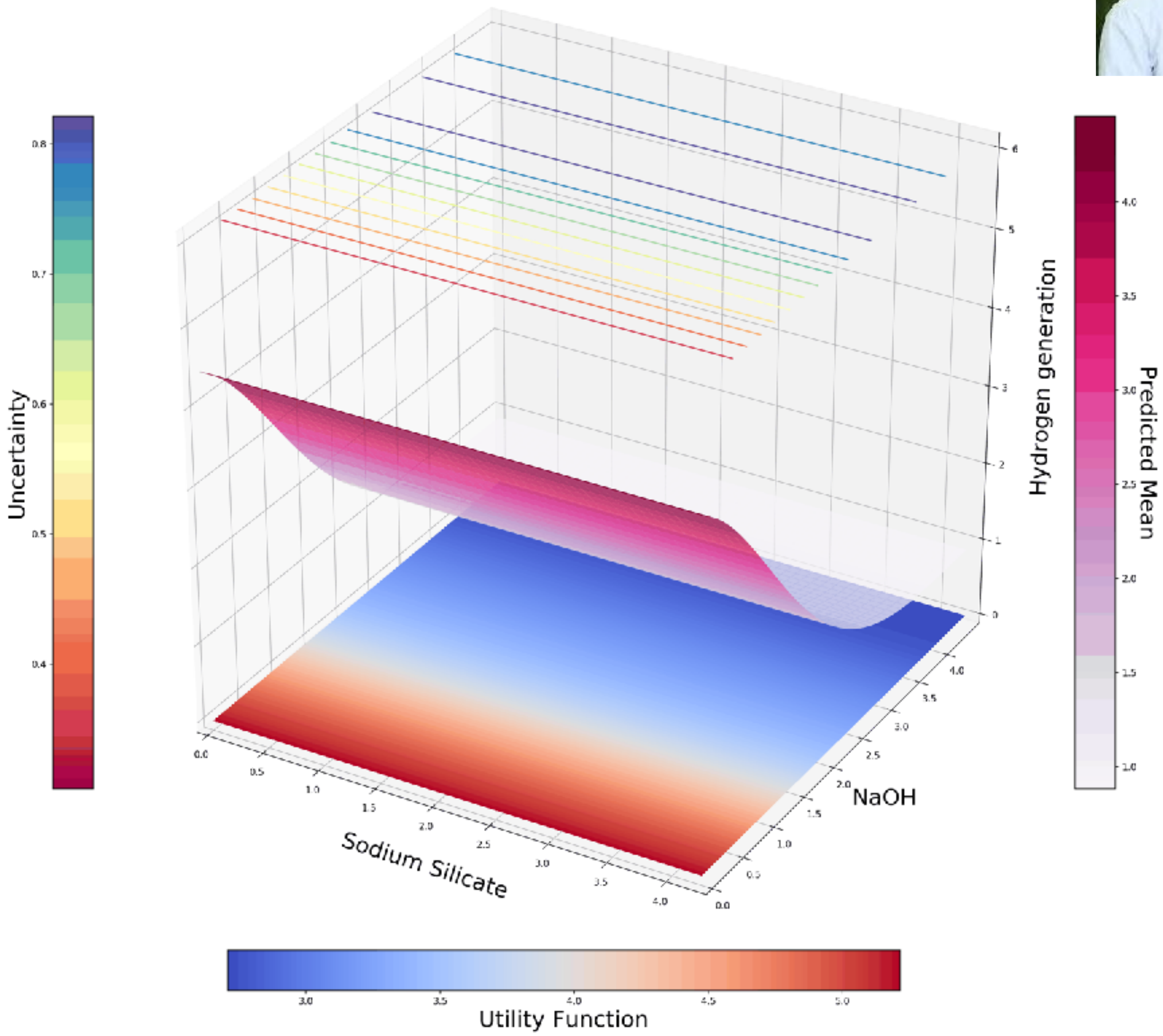
Human defined experiments ran by robot researchers



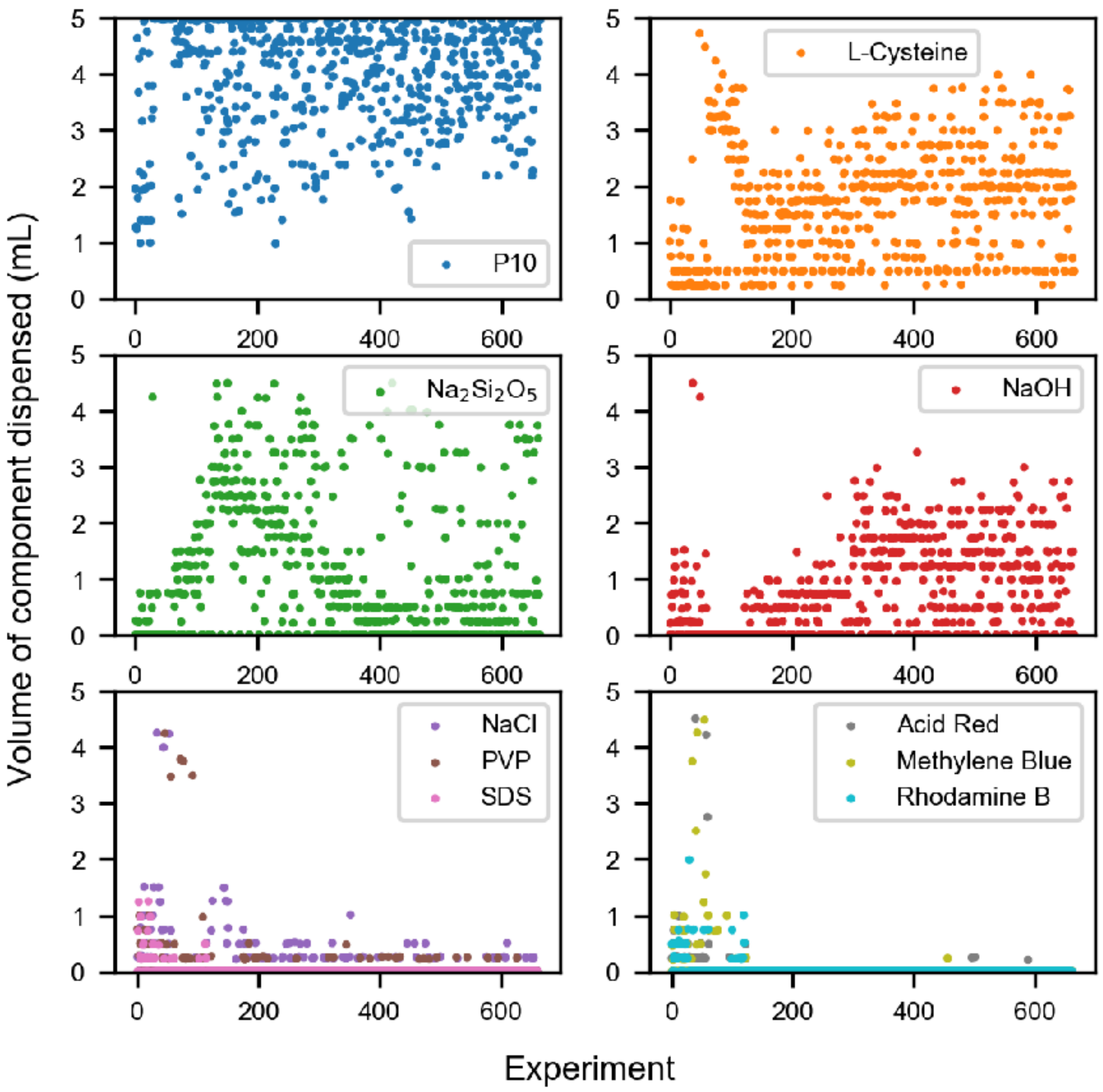


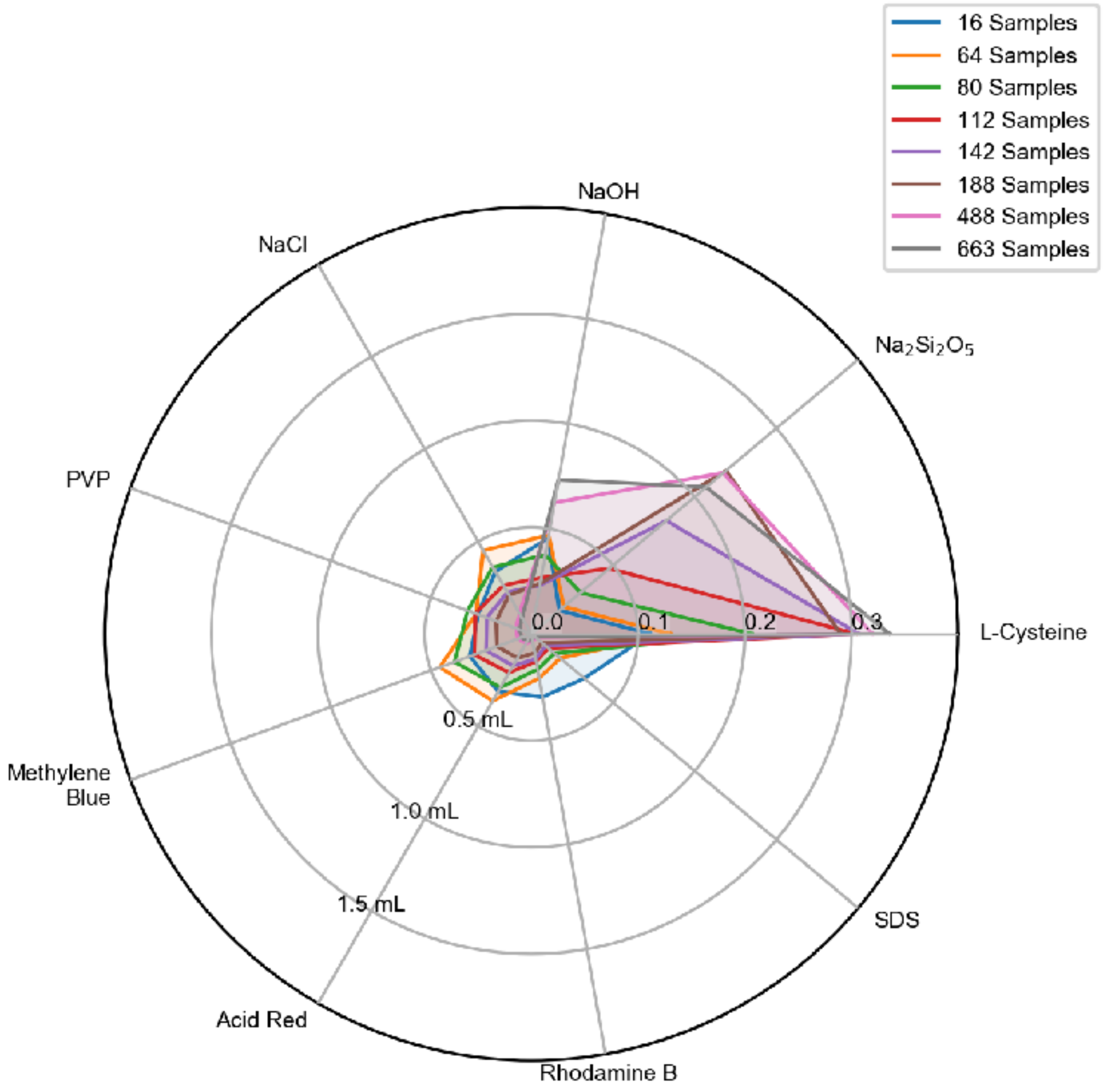
Models develop over time and balance exploration and exploitation





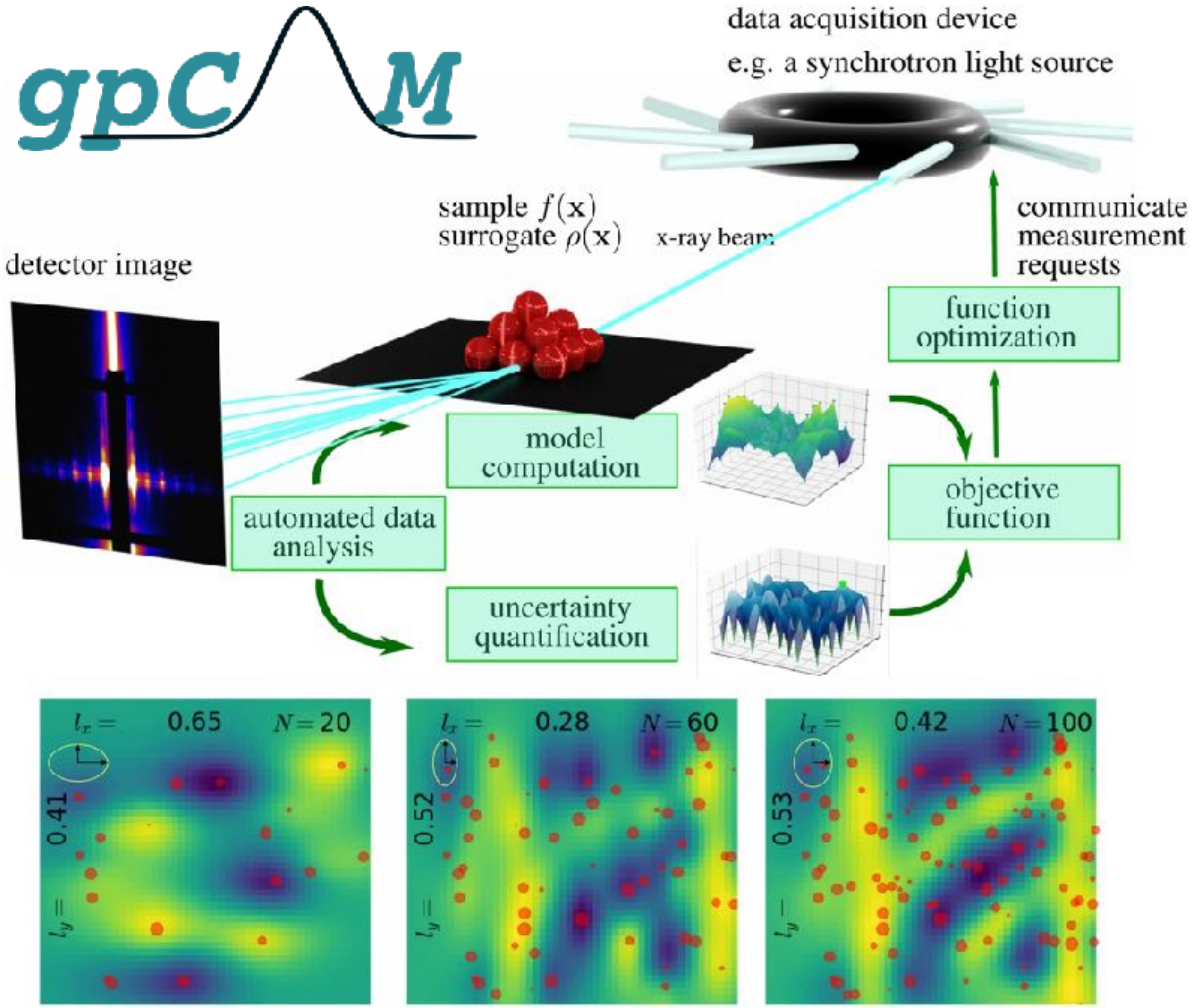
Important components are automatically selected





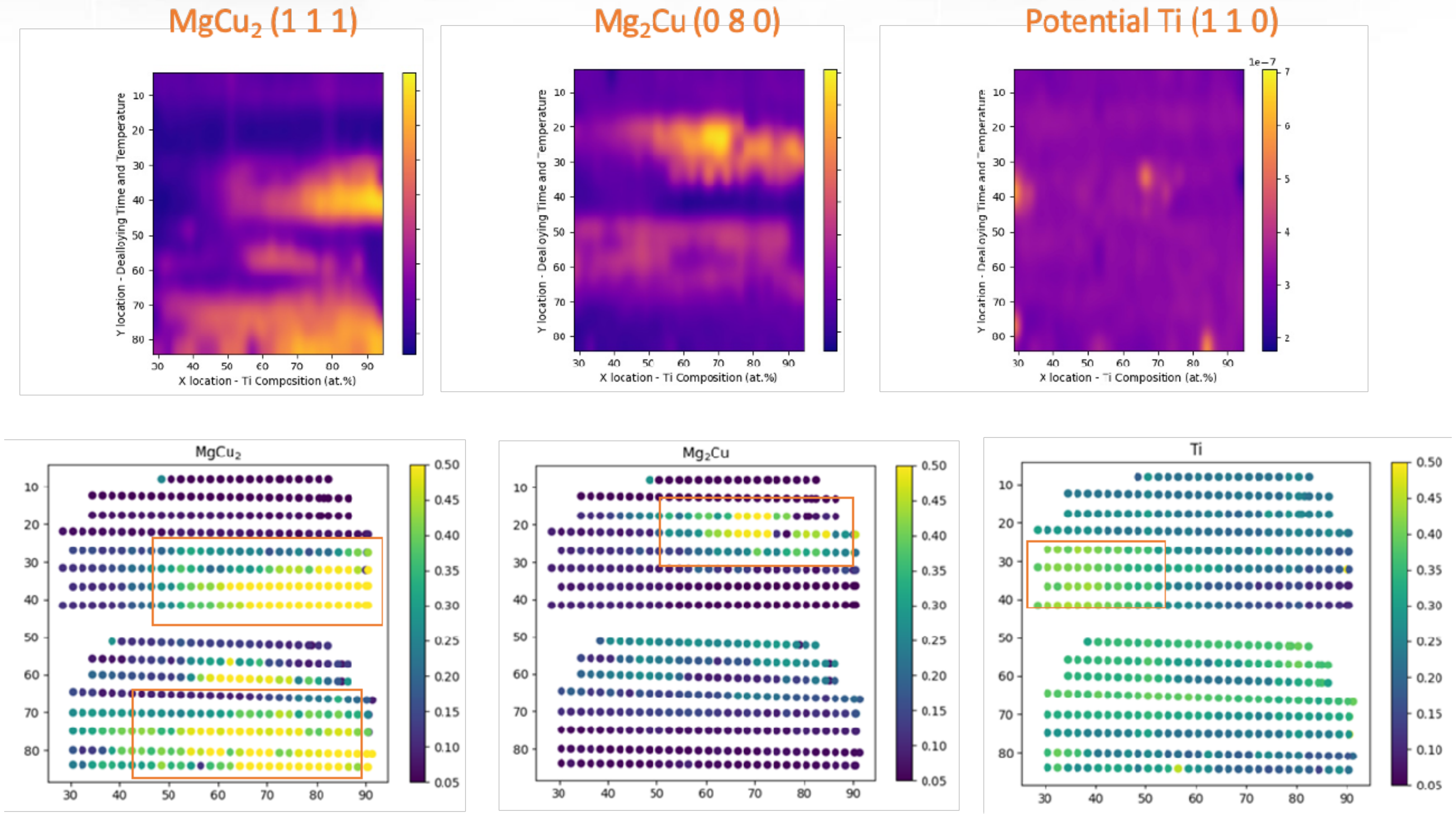
Bayesian optimization for autonomous characterization is enabled by bluesky-adaptive.

- Exploration for optimizing understanding of a sample phase space.
- Exploitation for optimizing a response function in a sample phase space.
- Seamless integration of heuristics and scientist-in-the-loop.



Noack, M. M. *et., al. Sci Rep* ***10,*** *17663 (2020)*

Probabilistic predictions from supervised models can also guide effective experimentation.



Office of Science

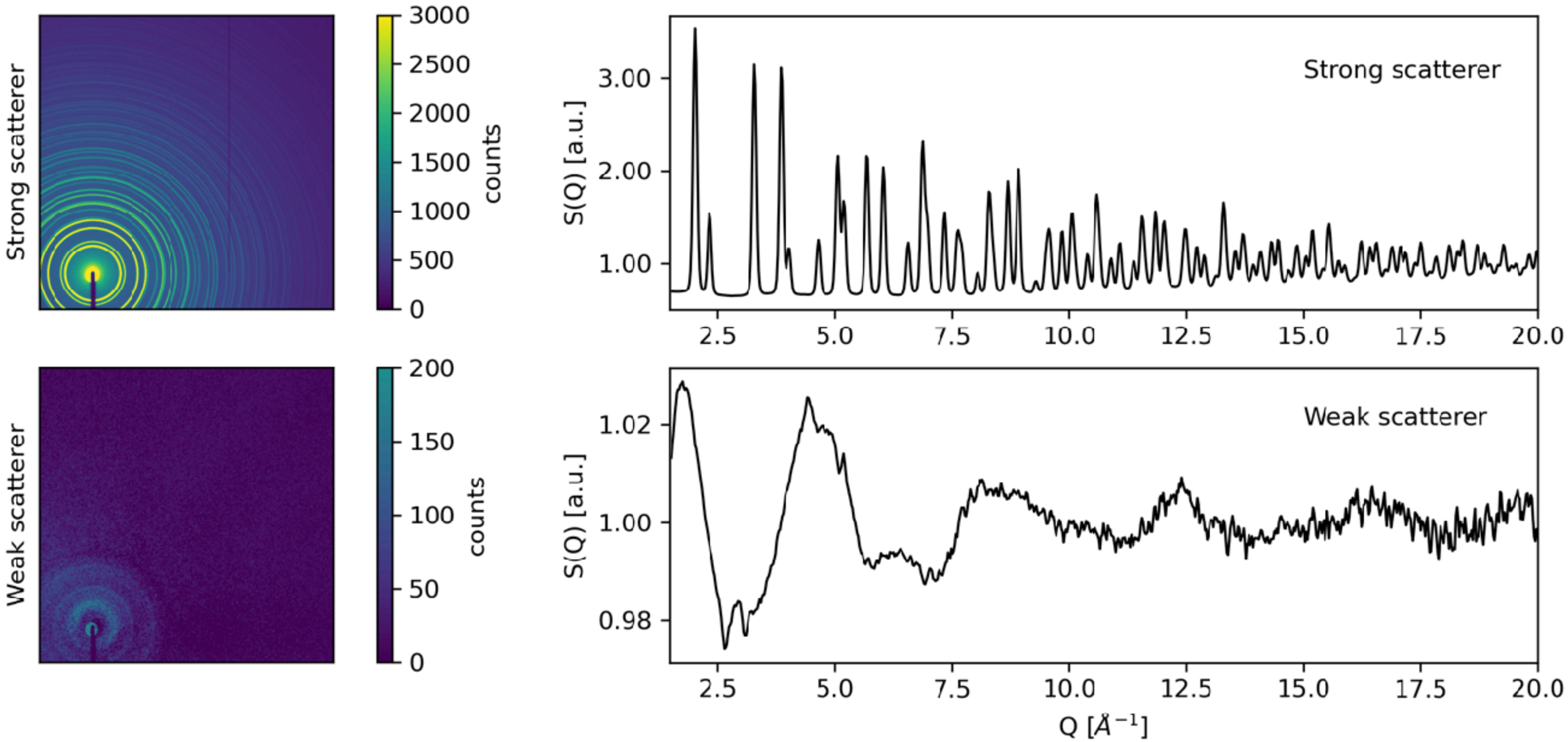
Stony Brook University

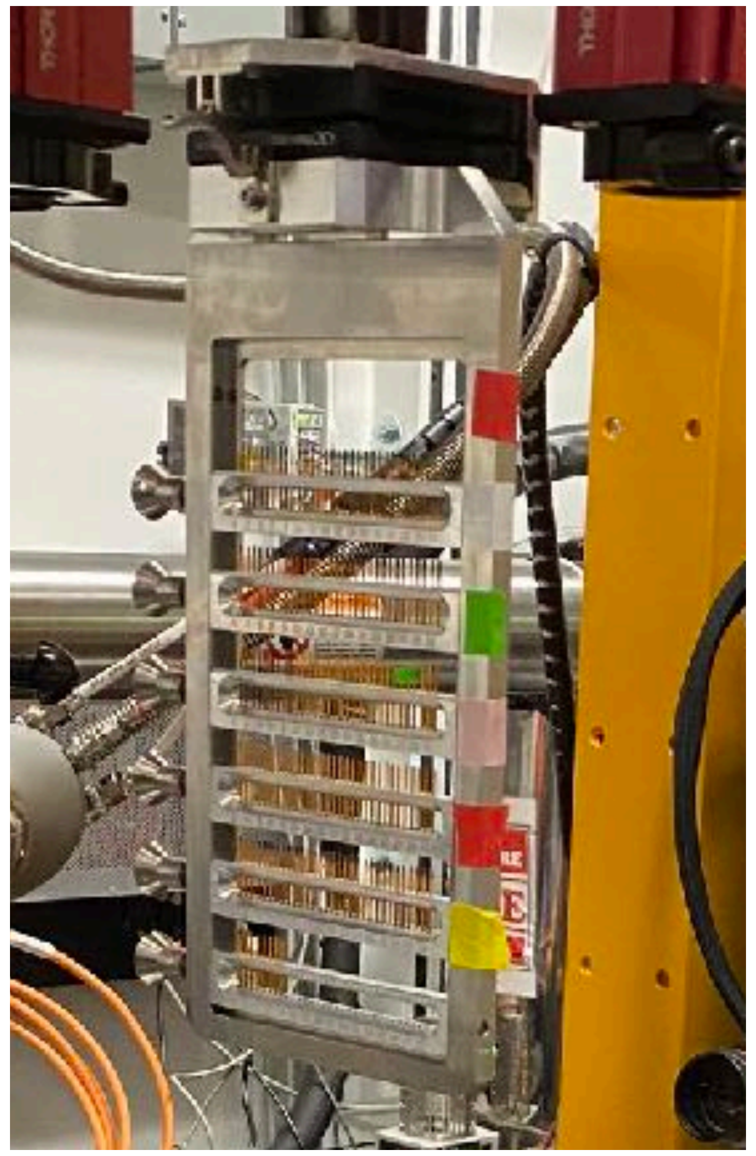
National Synchrotron Light Source II

Reinforcement learning:

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

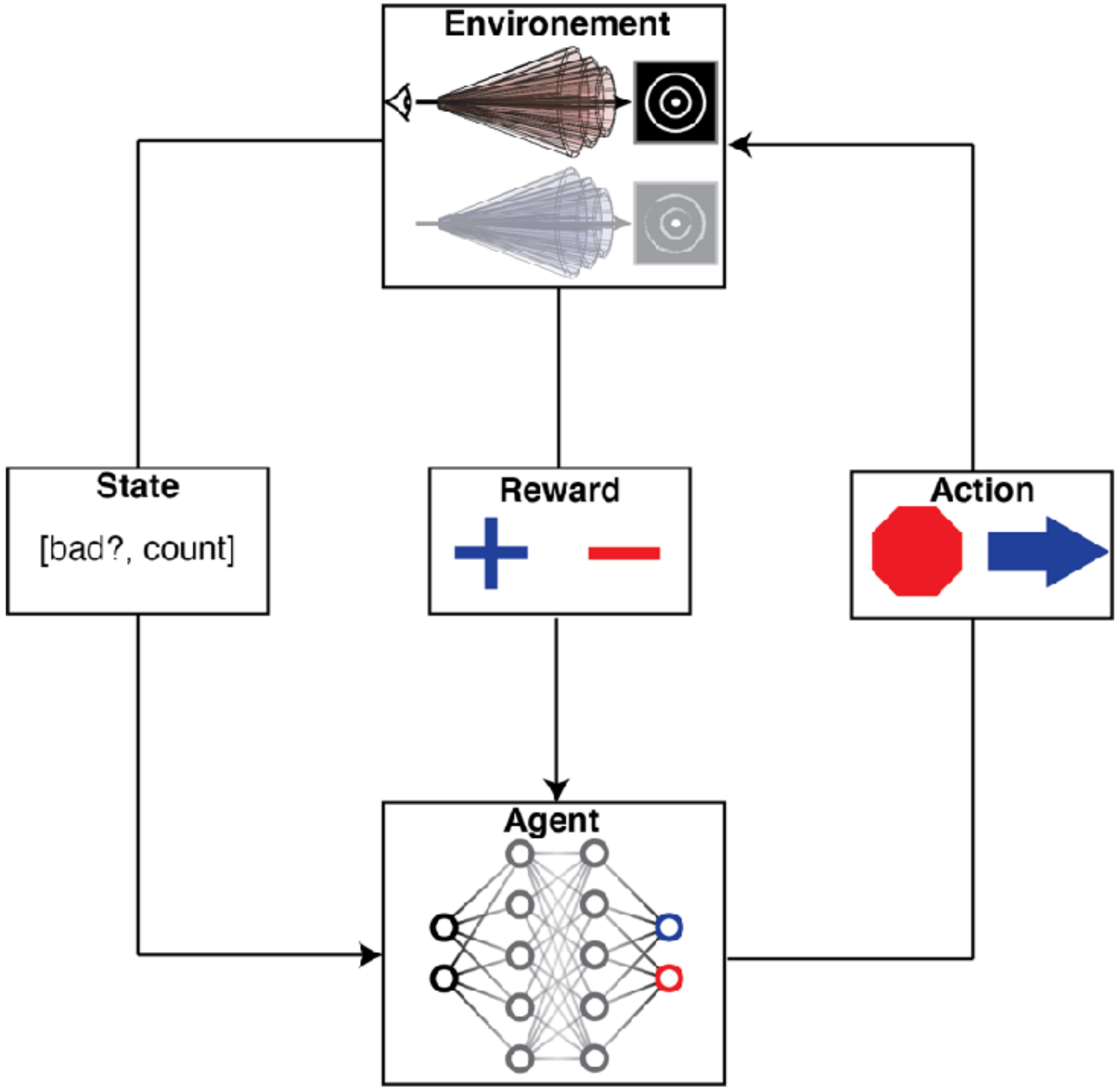
With hundreds of samples to run remotely, how do we best utilize our resources?





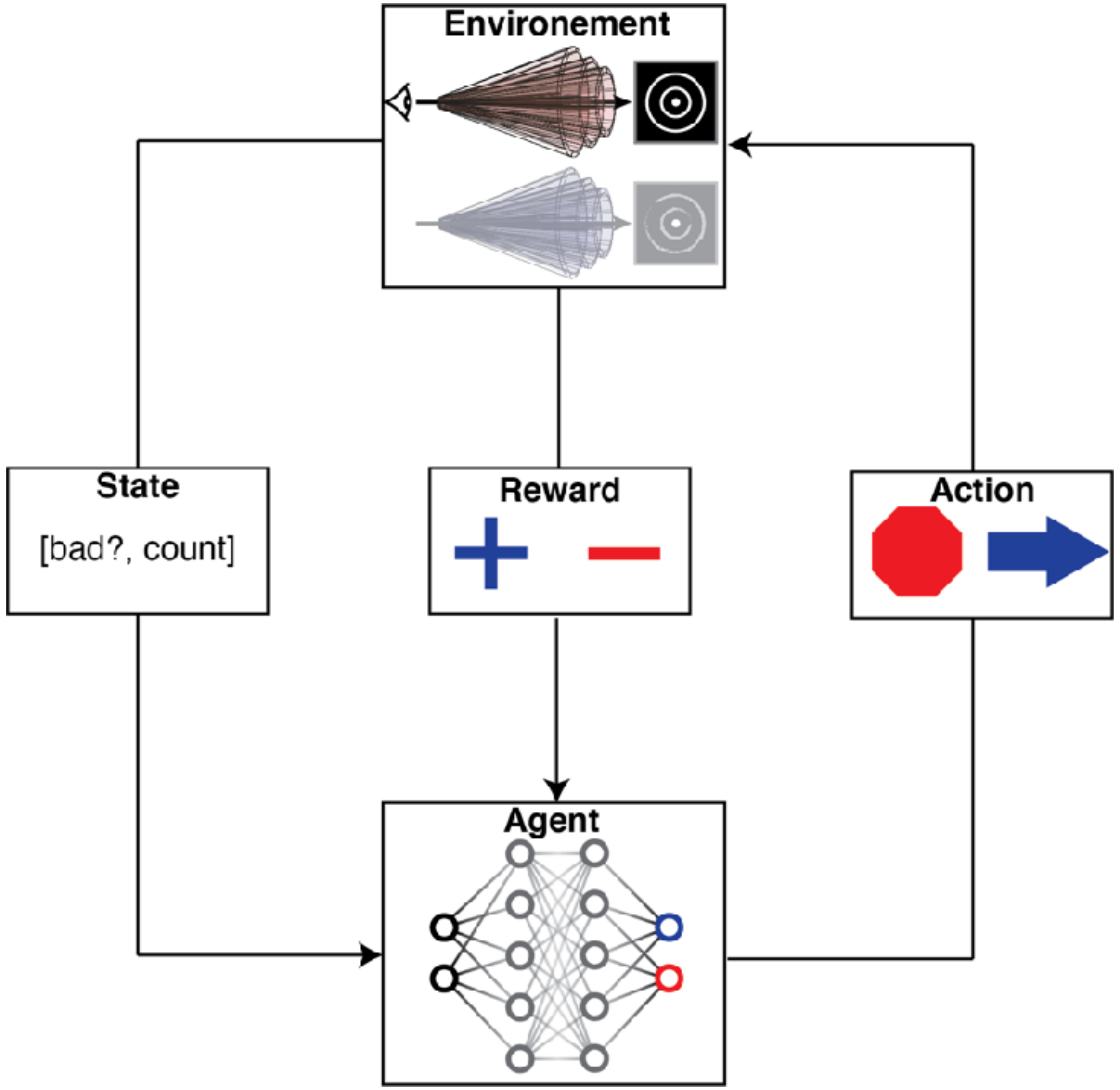
Mach. Learn.: Sci. Technol. **2**, 025025 (2021)

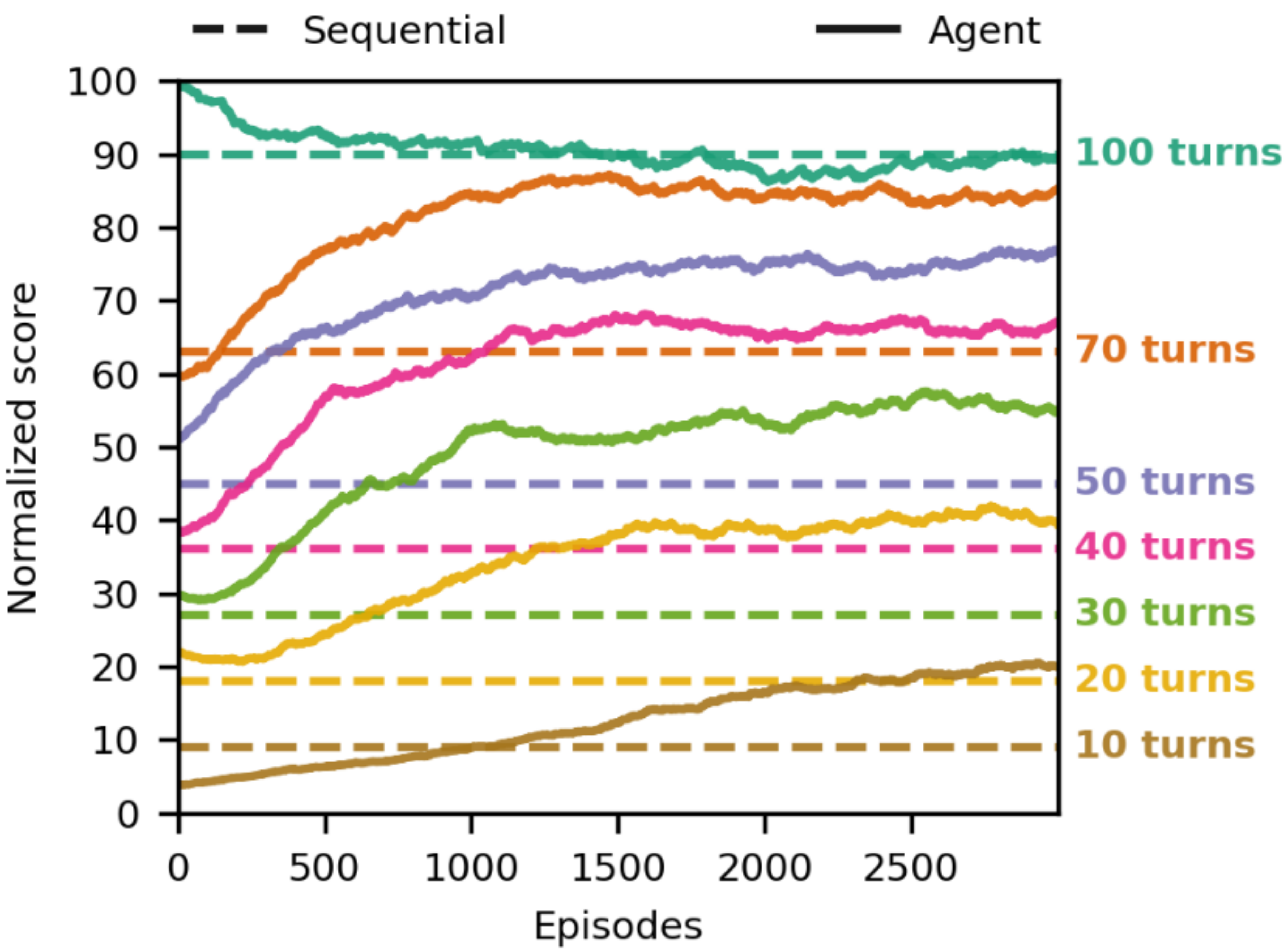
Reinforcement learning develops policies for optimal measurement strategies.



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

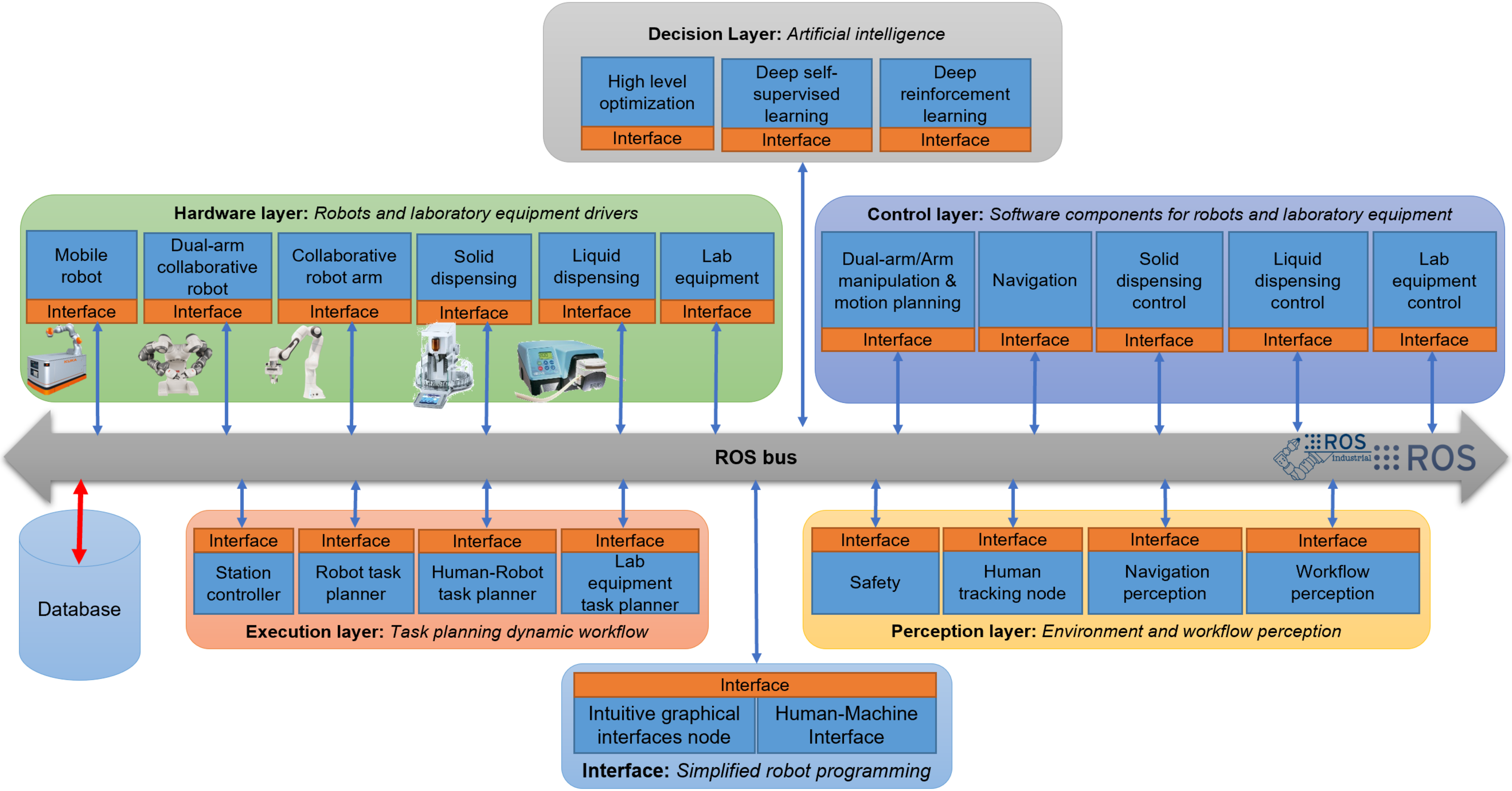
Reinforcement learning develops policies for optimal measurement strategies.





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

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

abb_driver

ROS industrial

Mettler-Toledo
QS30 Autosampler

Watson Marlow
323DU Pump

IKA
ULTRA-TURRAX

KUKA

ROS industrial