Autonomous Materials Discovery using X-rays

Kevin G. Yager (kyager@bnl.gov)

Brookhaven National Laboratory | Center for Functional Nanomaterials | National Synchrotron Light Source II | Computational Science Initiative

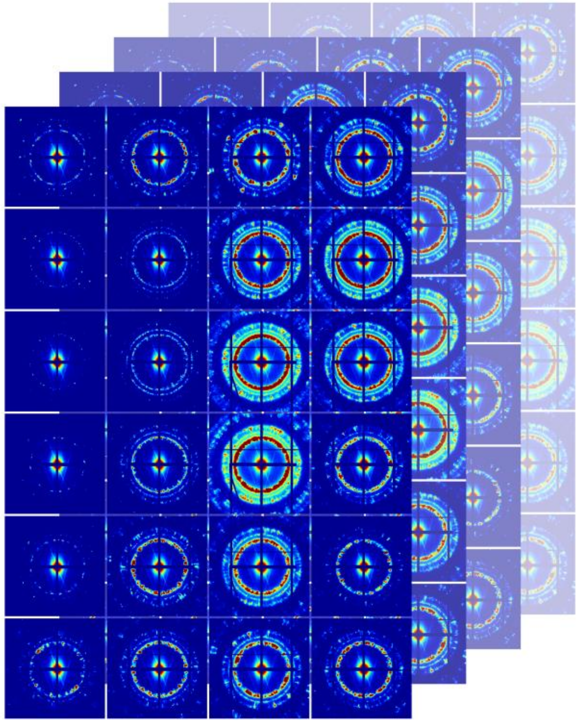
Synchrotron Data

- Synchrotron (electron accelerator) generates x-ray beams for material study

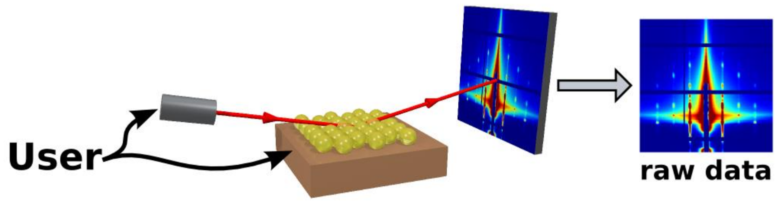
- Beamlines generate large amounts of data
  - Much of it is never analyzed

- Complex data analysis consumes scientists' precious time, distracting from deep scientific questions

- Leverage machine learning to automate experimental workflows?



Vision: Autonomous Experiments

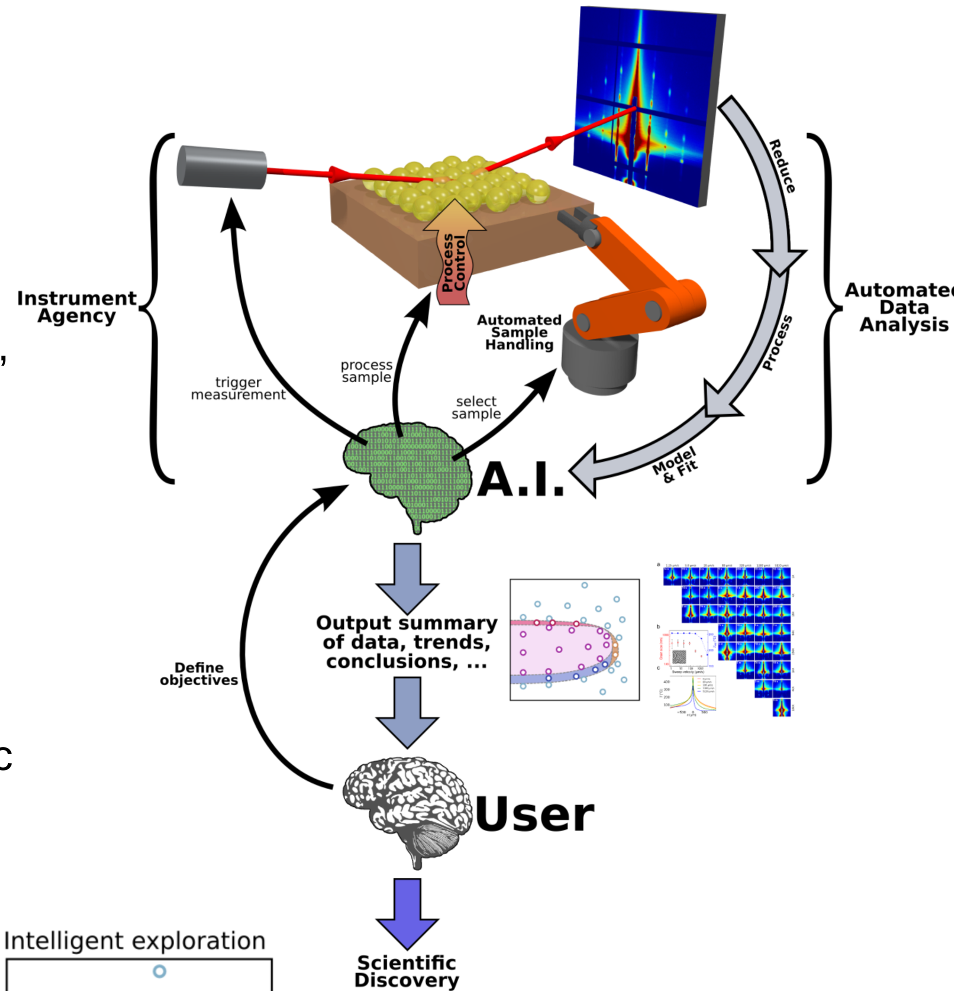


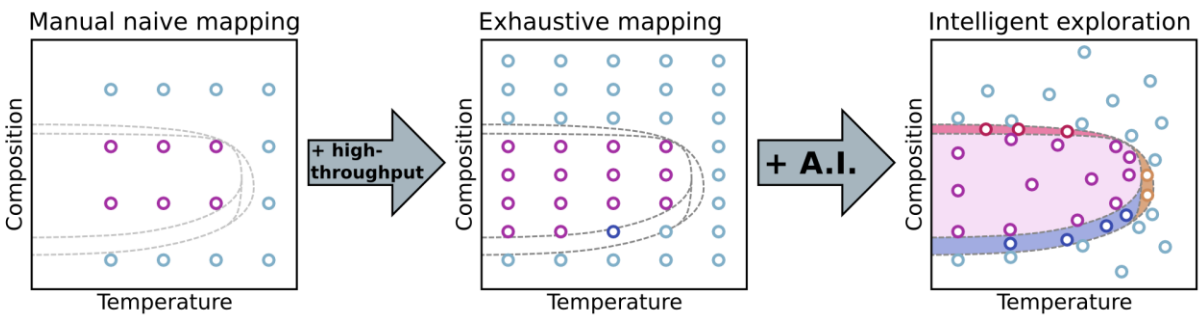
Past

- User manually loads samples, tweaks motors, collects data, rushes onto next sample…
- Human's time is wasted

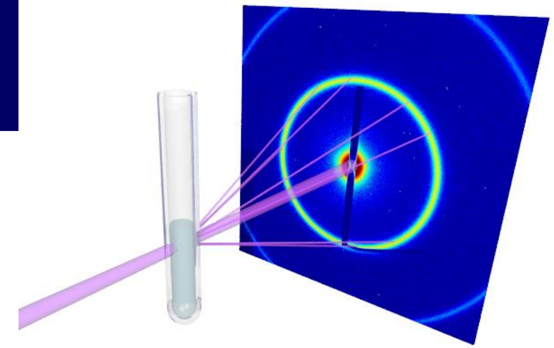
Future

- Automate the entire experiment, including **decision-making**
- **Liberate** human scientist to focus on scientific insight
- Accelerate **materials discovery**

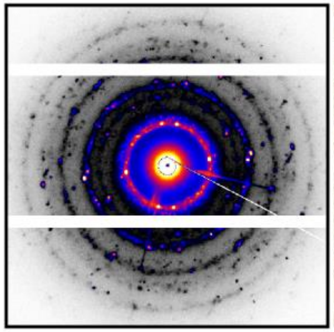


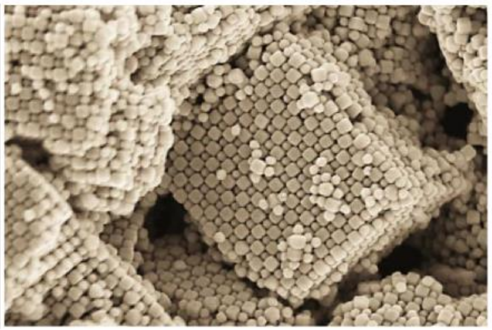


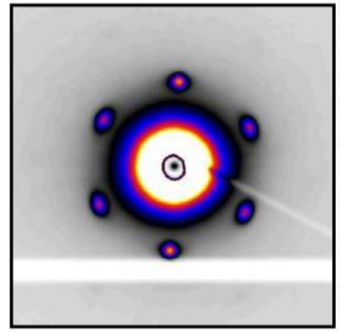
X-ray Scattering

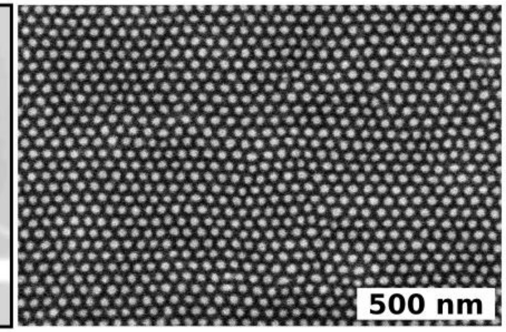


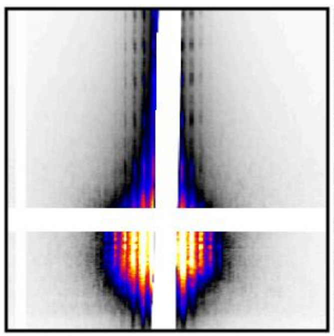
- The goal in x-ray scattering is to determine the structure of a material

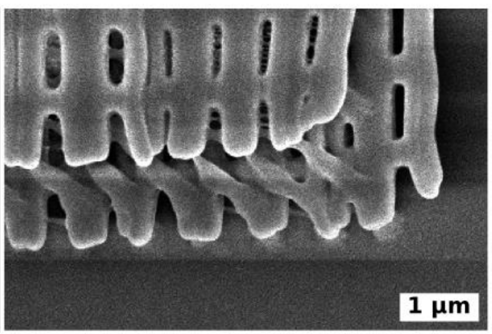


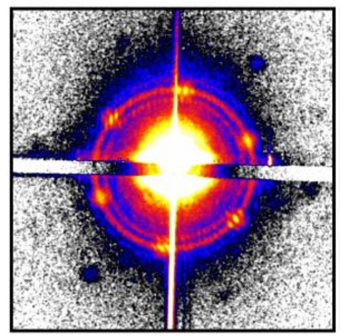


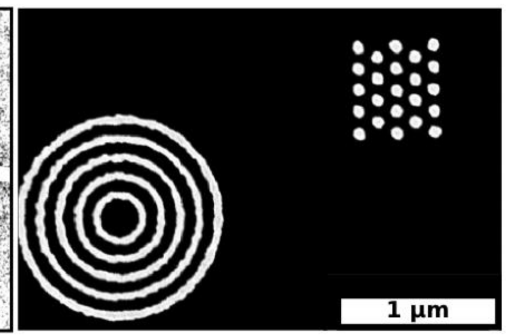


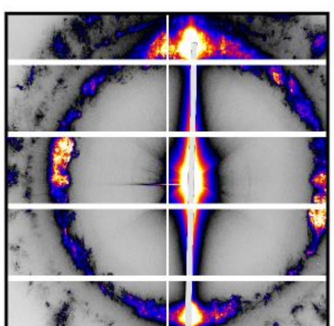




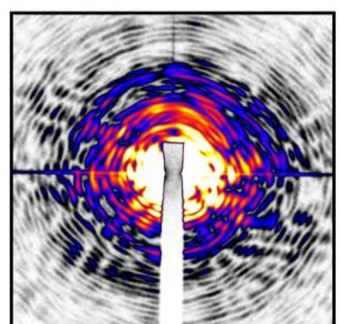








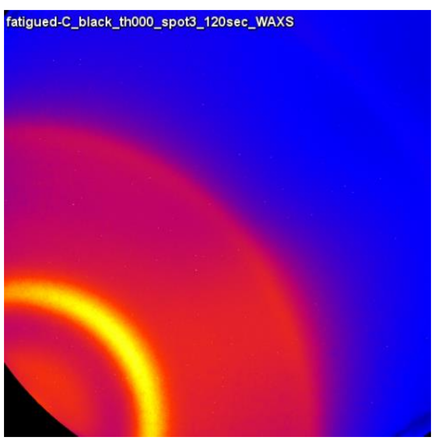
?



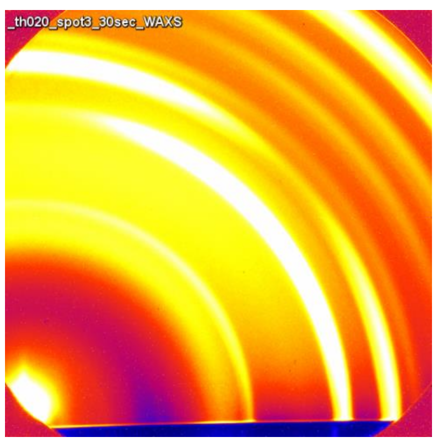
?

X-ray Detector Images

- Goal is to recognize the significance of an image (as a human does)
  - Is experiment working?
  - What features are in the image?
  - What's the sample's structure?
- We define an ambitious multi-classification problem (>100 tags)
  - Instrumental aspects (e.g. experiment type)
  - Image features (rings, peaks, symmetry, etc.)
  - Material detection (metal, polymer, etc.)
  - Holistic information ("sample is well-ordered", "material is aligned", etc.)

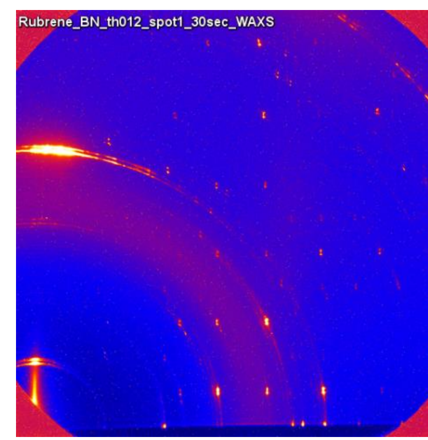


disordered



some ordering

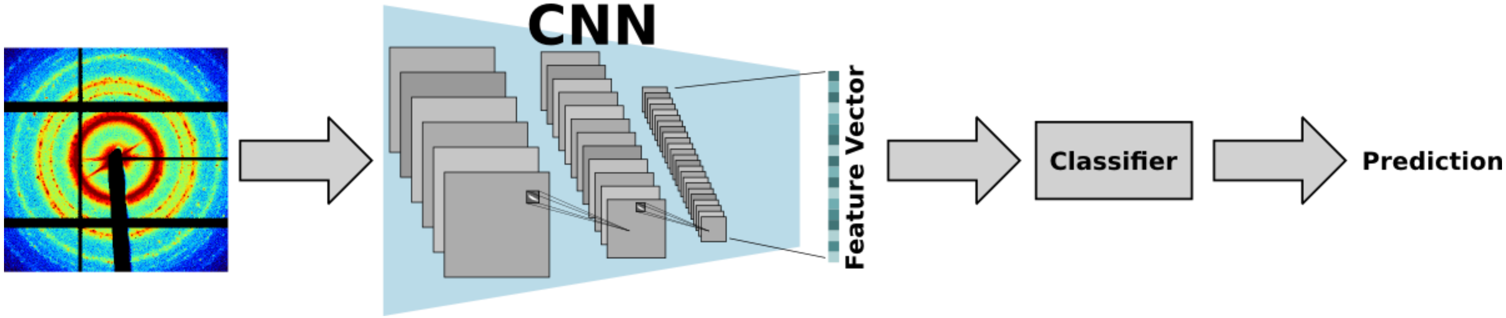
oriented, textured



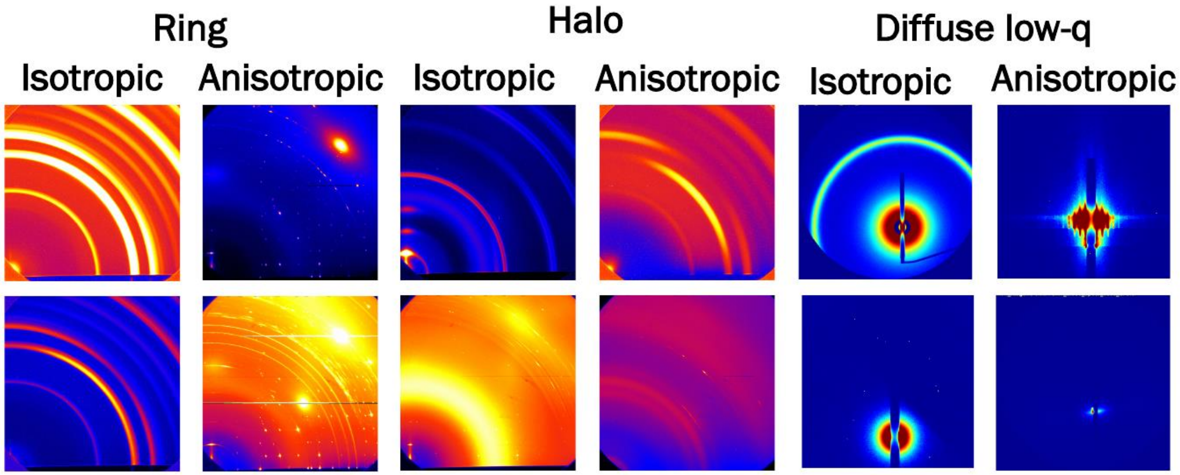
single crystal

X-ray Scattering

- Use **deep learning** (convolutional neural networks) for SAXS/WAXS detector images



- Synthetic data for training (mitigates sparsity and imbalance of tagged experimental data)
- Can automatically categorize images, identify scattering features or specific materials

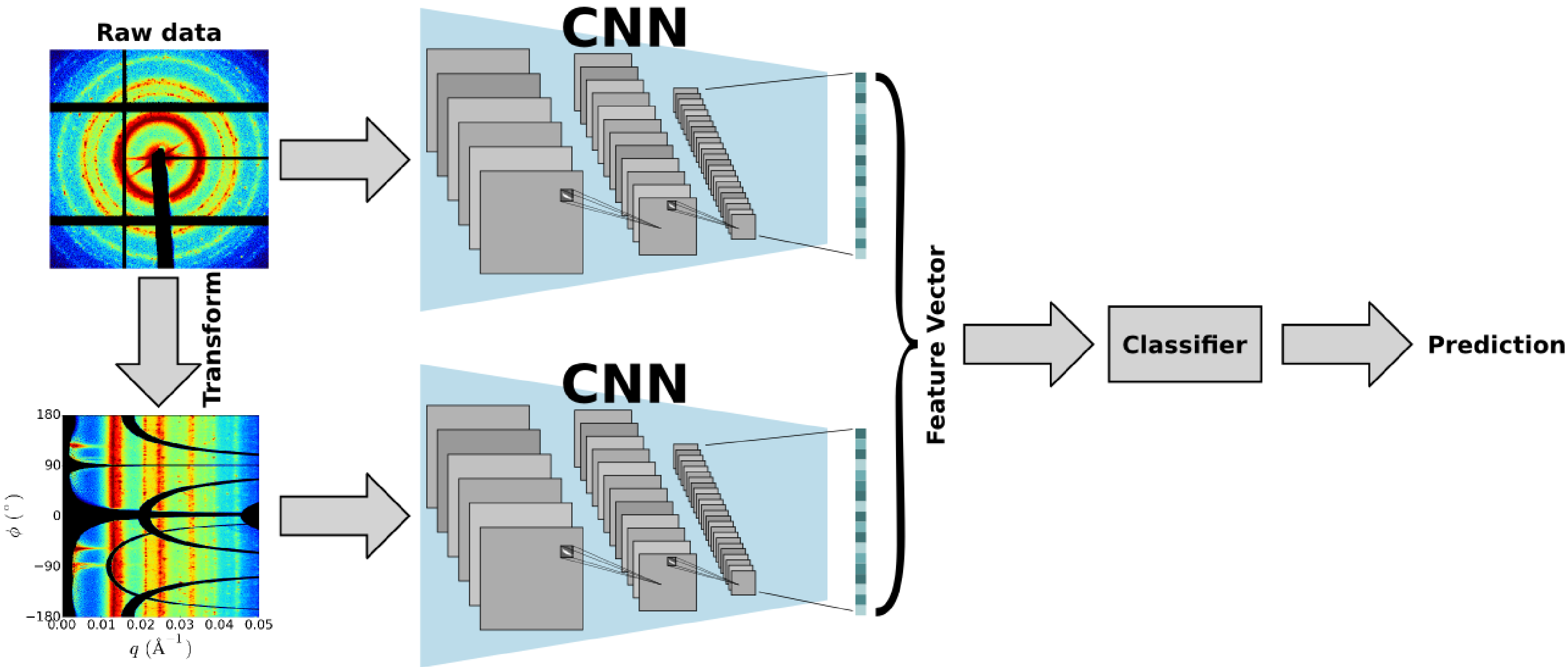


- Accuracy depends on class

Wang et al. *NYSDS* **2016**
Wang et al. *WACV* **2017**, *1*, 697

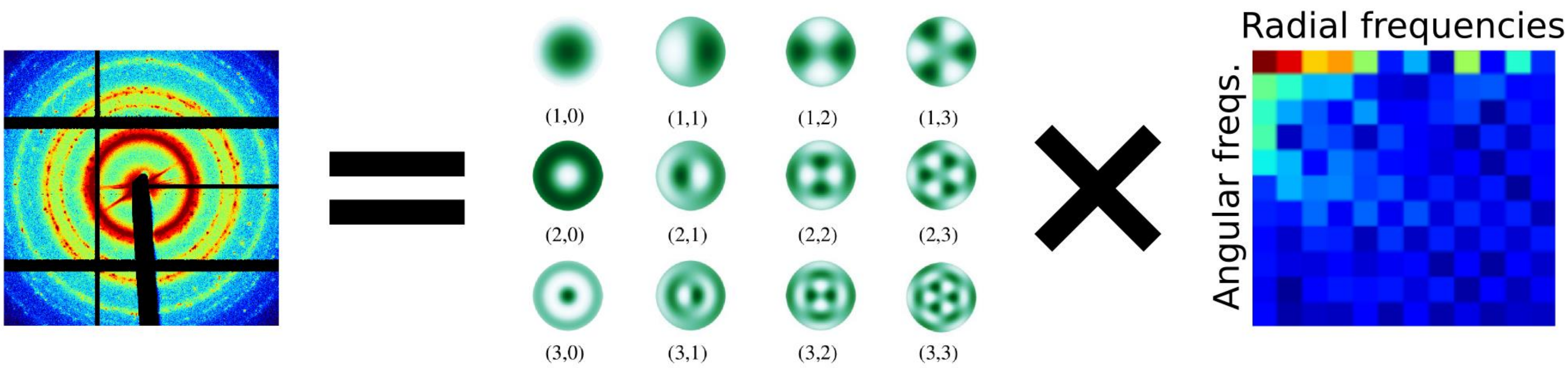
Multi-channel Learning

- Use multiple representations of the data as channels
- We can select representations that are particular well-suited to the scientific dataset
  - “Physics-aware” deep learning

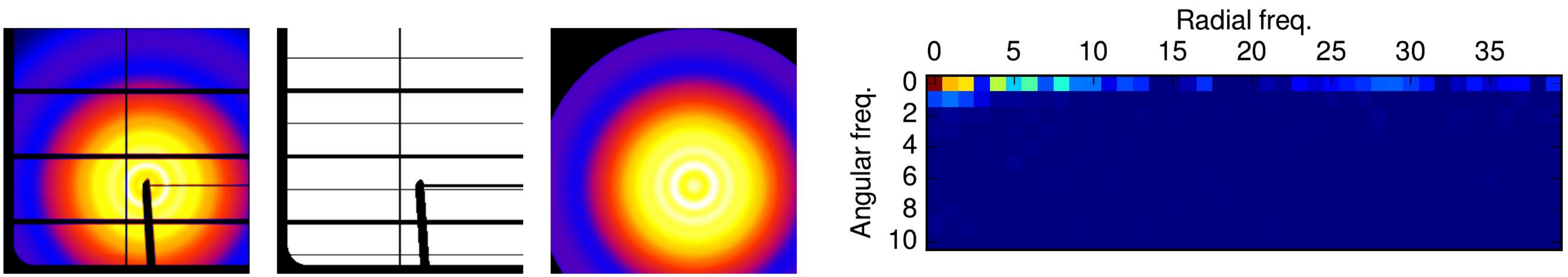


Fourier-Bessel Transform

- Fourier-Bessel decomposition is devised to “lock-in” on features that matter in scattering
  - Peaks
  - Symmetry
- Generates a matrix of terms (encoding important features)

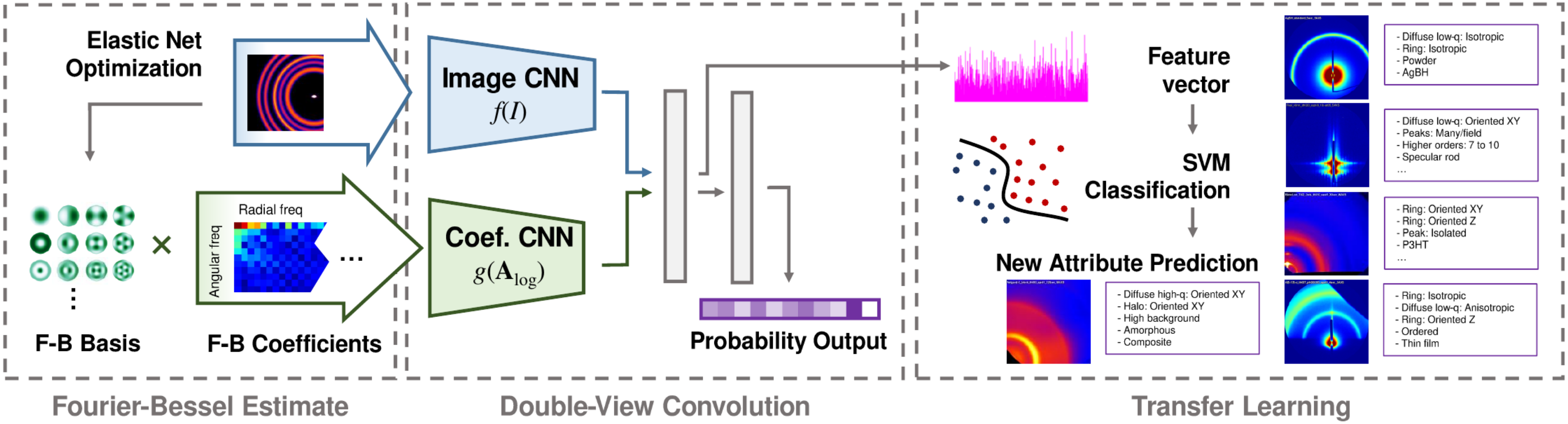


- Sparse matrix can be used as a “compressed representation,” which can be used for image healing and to improve classification



Fourier-Bessel Transform

- Compressed representation also useful for classification
- Fourier-Bessel decomposition designed to “lock-in” on features that matter in scattering
  - Peaks
  - Symmetry

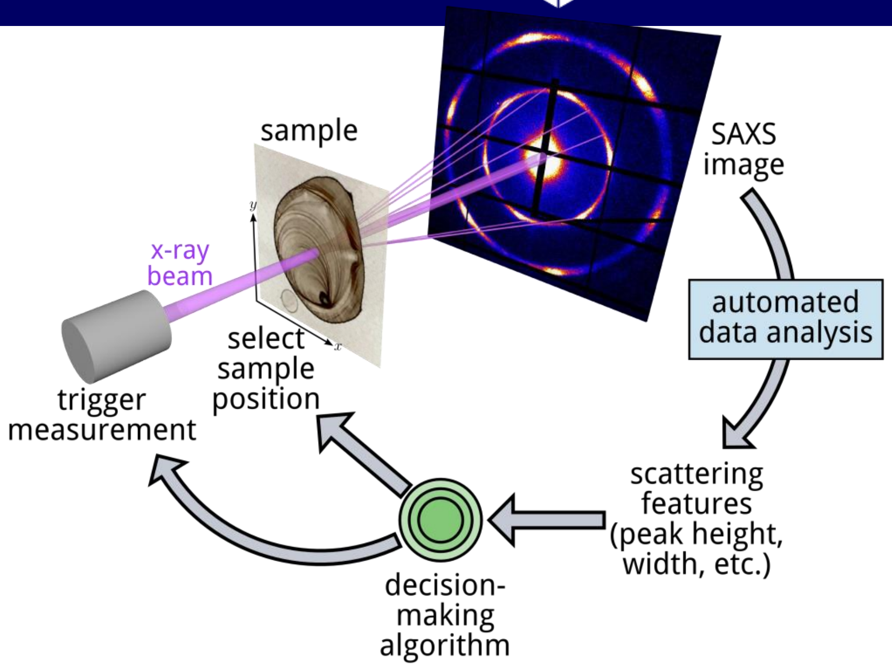


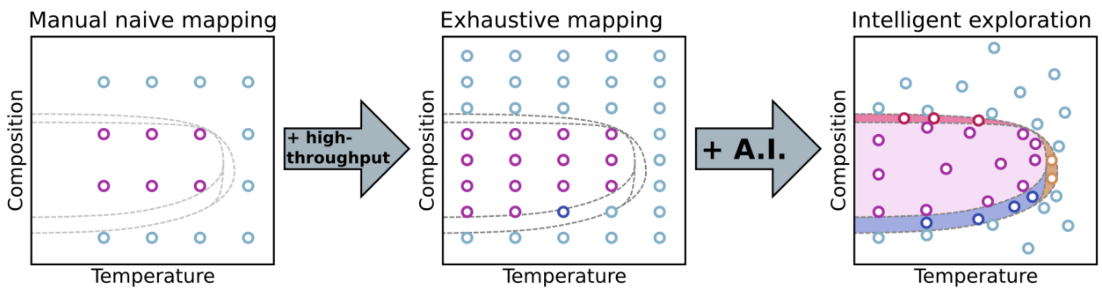
	mAP	Diff. lo-q	Diff. hi-q	Halo	Higher Ord.	Rings	Sym. halo	Sym. rings	2-fold sym.	4-fold sym.	6-fold sym.
Positive Ratio	–	0.1366	0.0840	0.2226	0.5776	0.5978	0.1440	0.1192	0.2422	0.1176	0.0838
Image CNN	0.6424	0.8945	0.8012	0.8839	0.9580	0.9568	0.5778	0.4873	0.4238	0.2199	0.2208
Coef. CNN	0.7450	0.8486	0.7044	0.8502	0.9451	0.9506	0.7165	0.5099	**0.6915**	0.5718	0.6612
Joint	**0.7779**	**0.9087**	**0.8178**	**0.9014**	**0.9604**	**0.9626**	**0.7494**	**0.5385**	0.6821	**0.5964**	**0.6621**

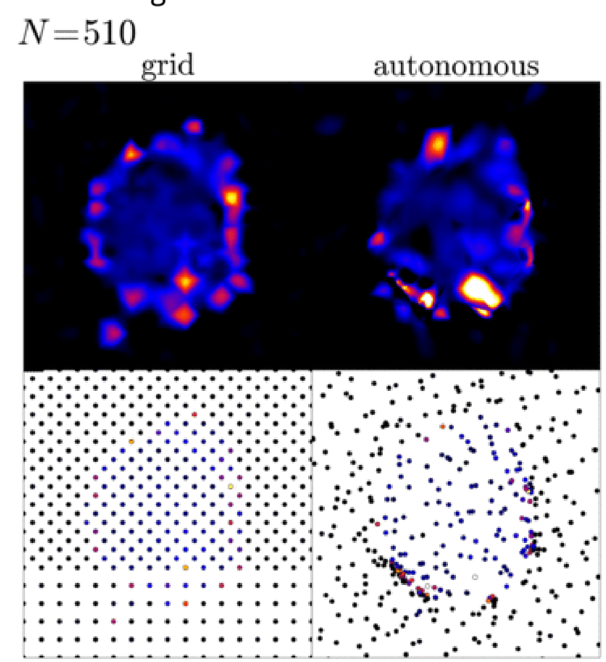
Guan et al. *BMVC* **2018**, *0828*, 1

Autonomous Experiments

- Automate entire experiment
  - Sample handling, align, data acquisition
  - Data analysis
  - **Decision-making**
- Improves efficiency
  - More optimal than naïve methods
  - Can define target (exploration, target, consider cost, find novelty)
  - AI/ML methods yield surrogate model, uncertainty
- **Liberates** humans to focus on science
  - Accelerated material discovery



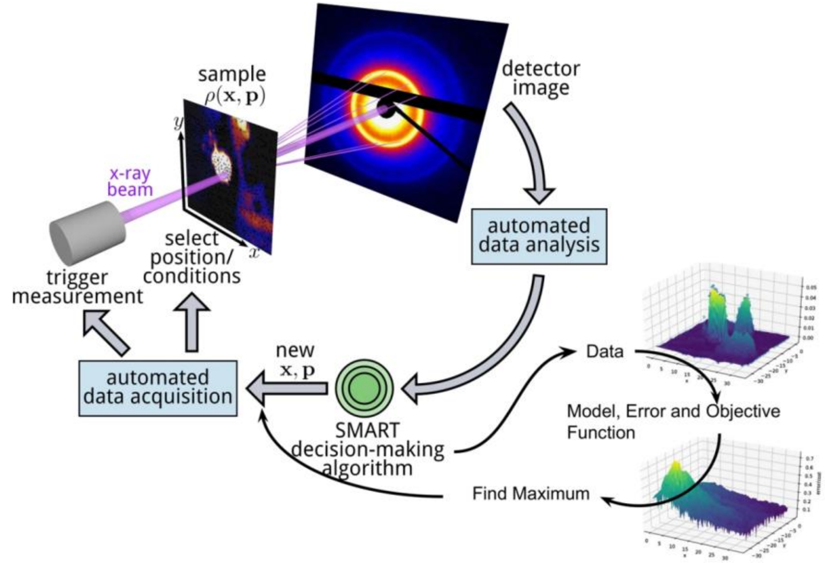


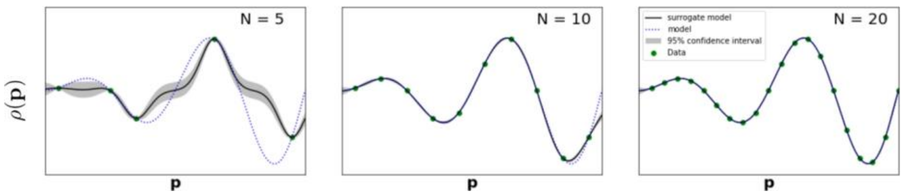


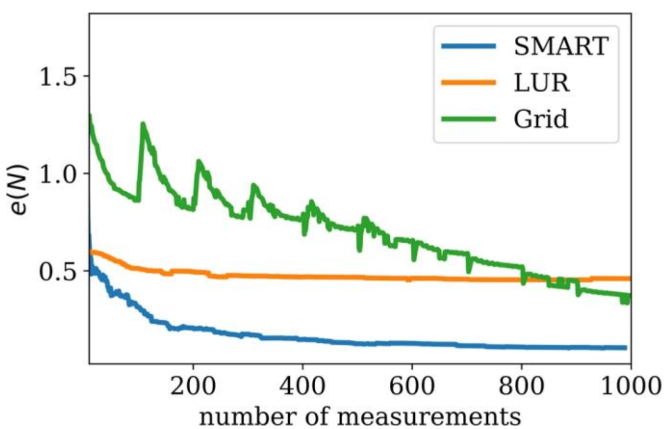
Noack et al. *Scientific Reports* **2019**, 9, 11809
Noack et al. *Scientific Reports* **2020**, 10, 1325
Noack et al. *Scientific Reports* **2020**, 10, 17663
Noack et al. *Nature Reviews Physics* **2021**, 3, 685

Algorithm: Gaussian Process

- Construct **surrogate model**
  - Fit/interpolate data
  - Select kernel to match physics
  - Estimate hyper-parameters that match data (lengthscales, periodicity, etc.)
- Calculate **uncertainty surface**
- Construct **objective function**
  - Search for maximum in objective
  - Optimization problem (genetic algorithm, differential evolution, deflation)
  - Control behavior: gradients, cost, etc.
- As we **iterate**, errors decrease (can terminate at desired error) and surrogate model improves



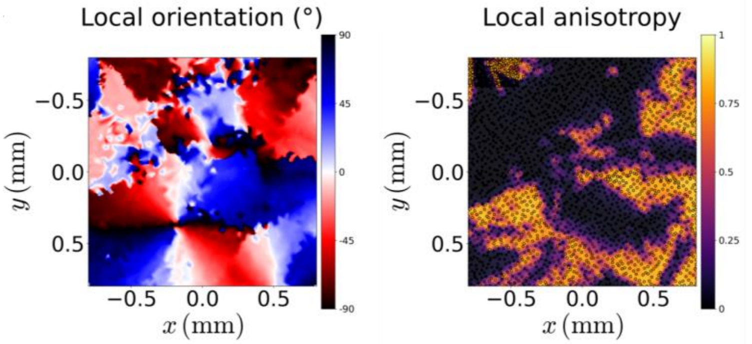




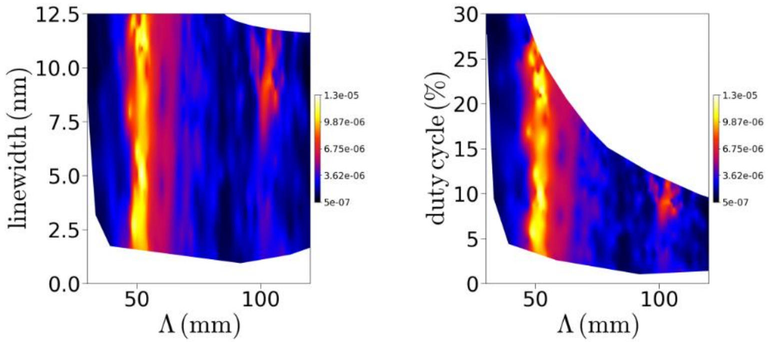
Noack et al. *Scientific Reports* **2019**, 9, 11809
Noack et al. *Scientific Reports* **2020**, 10, 1325
Noack et al. *Scientific Reports* **2020**, 10, 17663
Noack et al. *Nature Reviews Physics* **2021**, 3, 685

Autonomous X-ray Scattering

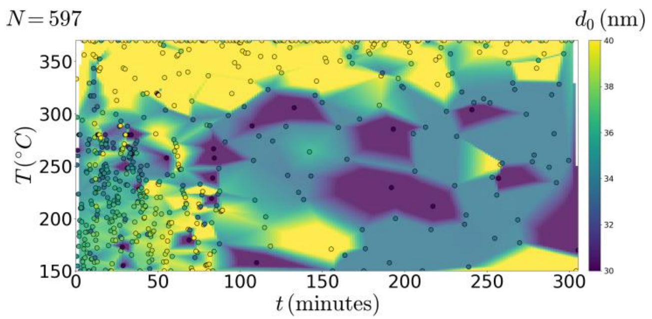
- **Mapping**
  - Coarse-to-fine imaging
  - Scale of heterogeneity not known *a priori*



- **Combinatorial sample arrays**
  - Explore physical parameter space
  - Spaces may be complex, high-dimensional



- **Real-time processing**
  - Control material ordering in real-time
  - Discover transient states and non-equilibrium process histories



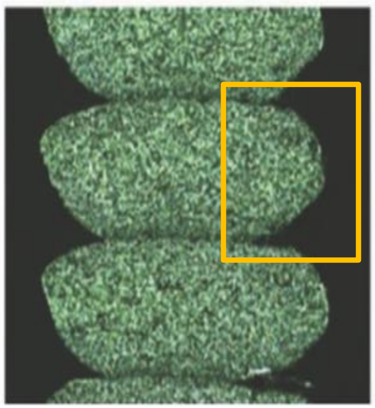
Kevin Yager (CFN), Masa Fukuto (NSLS-II)

Polymer 3D Printing

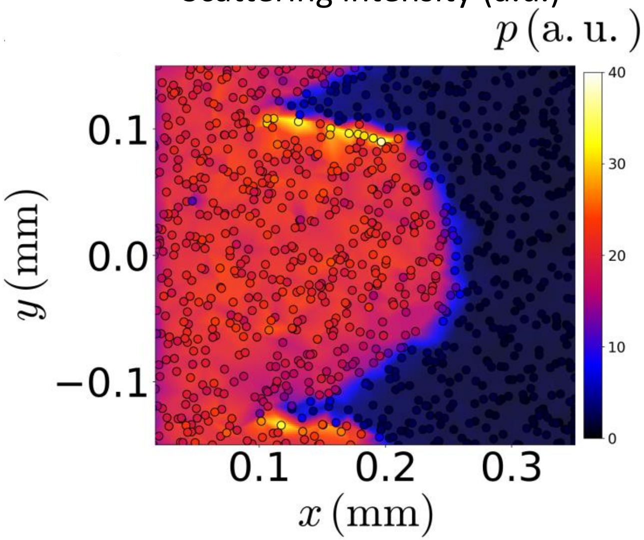
Materials science

- Polymer additive manufacturing depends critically on the welds between layers

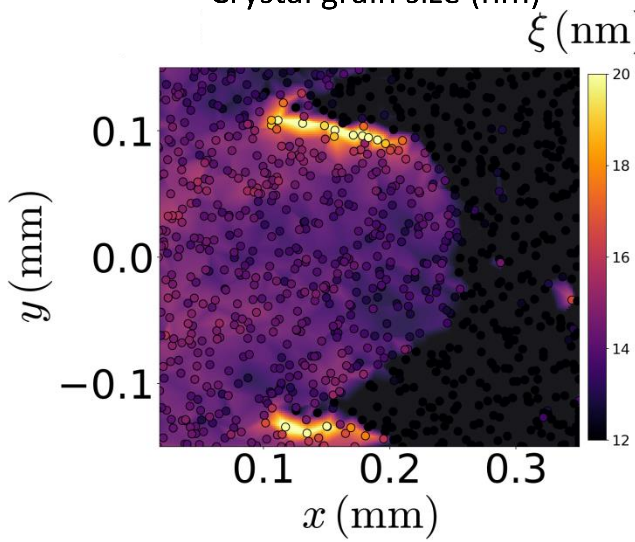
POM image



Scattering intensity (a.u.)

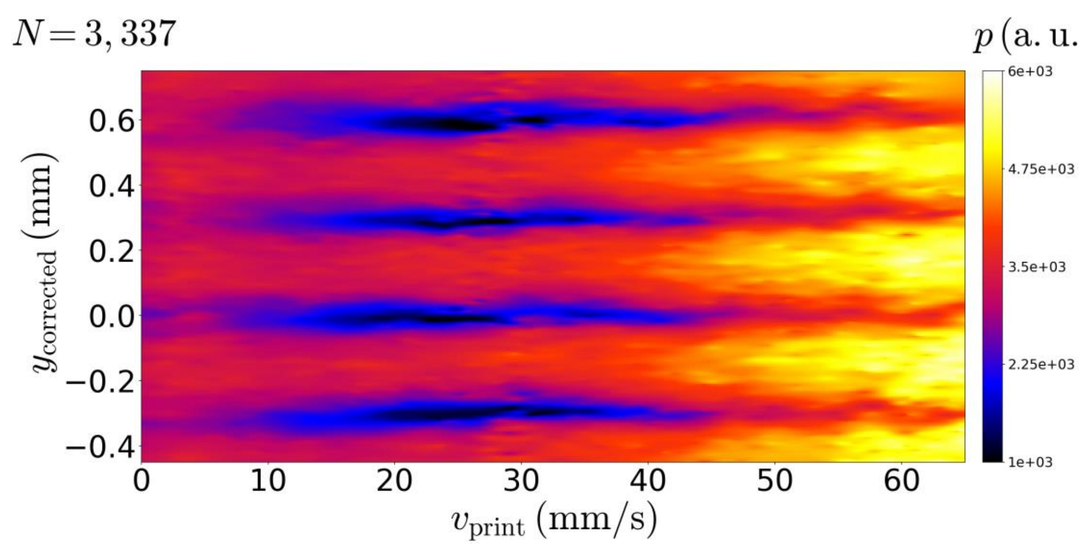


Crystal grain size (nm)



Experiment

- X-ray scattering mapping of polymer crystallization
- Identified surface ordering (after just ~40 measurements)
- Measure dependence on process parameters

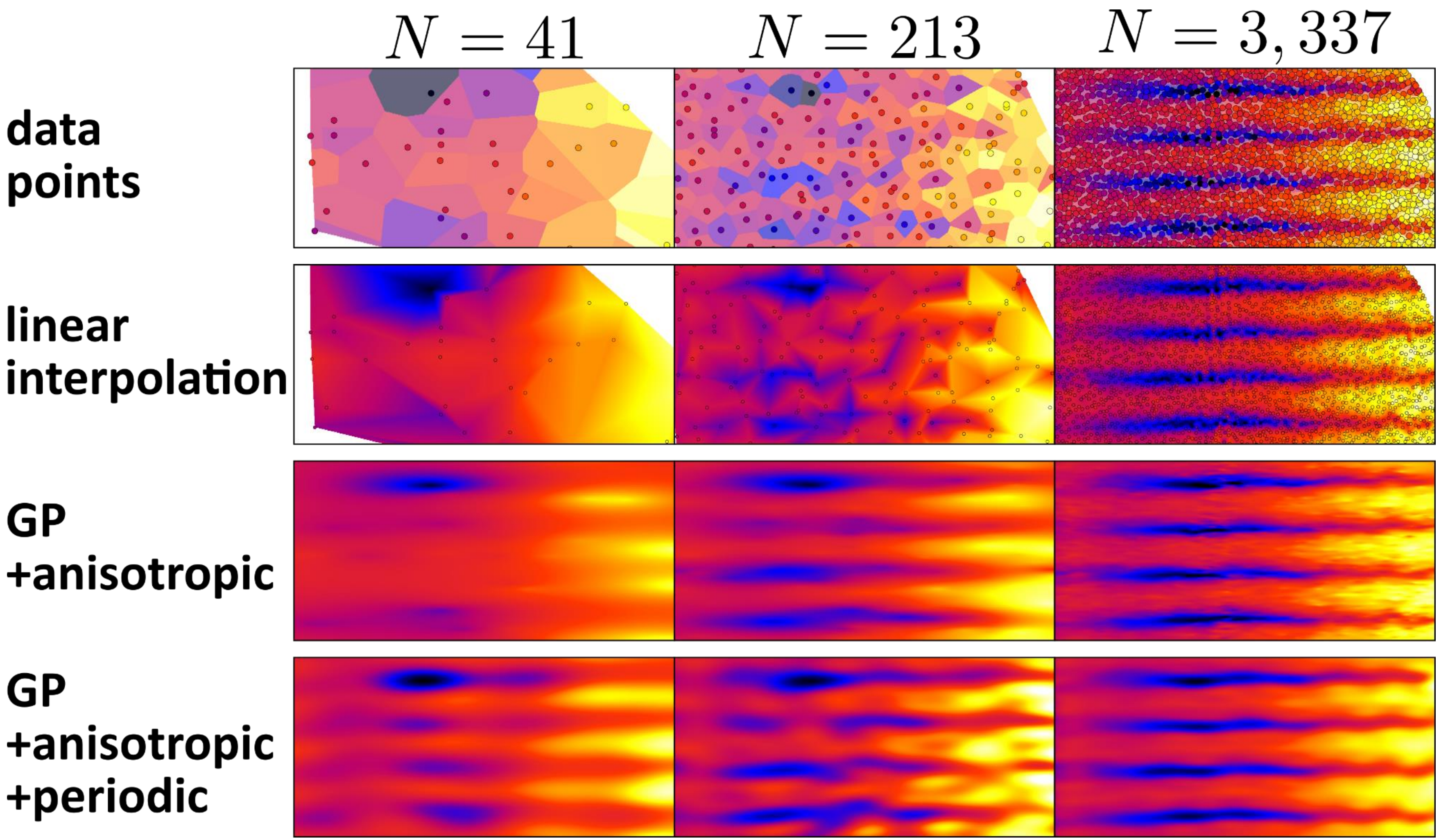


Collaboration: Jon Seppala, Tyler Martin (NIST)

Polymer 3D Printing

Model/kernel design matters

- Autonomous decisions (optimal exploration)
- Final reconstruction (insight)

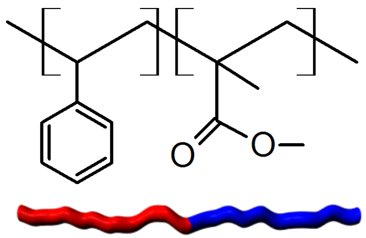


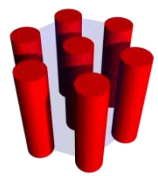
Collaboration: Jon Seppala, Tyler Martin (NIST)

Block Copolymers

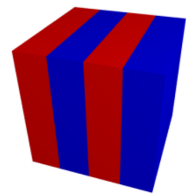
Structured Polymer Chains

- Spontaneously organize into nanoscale patterns

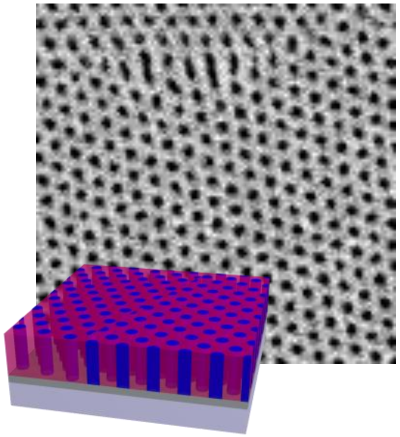


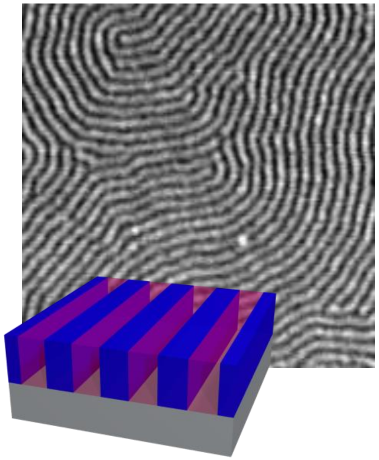


hexagonal cylinders



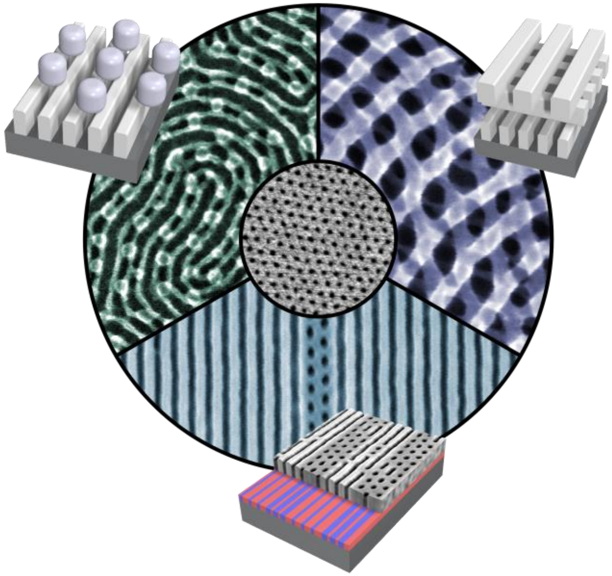
lamellae

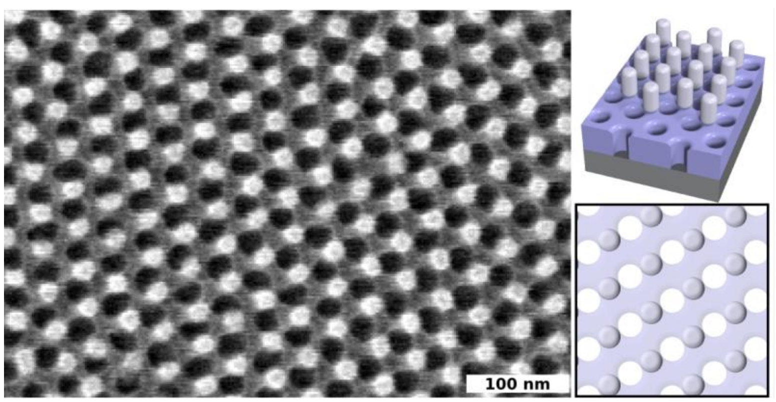


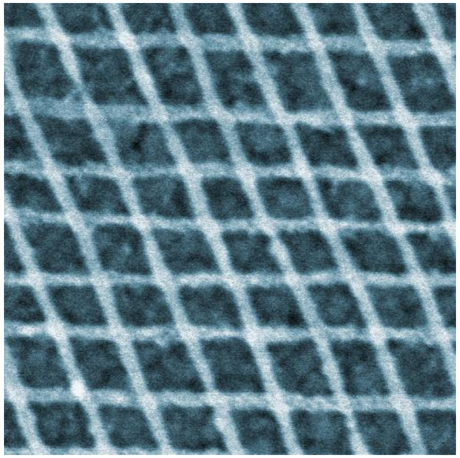


Responsive Materials

- Clever control of processing history can drive material into an exotic “non-native” state



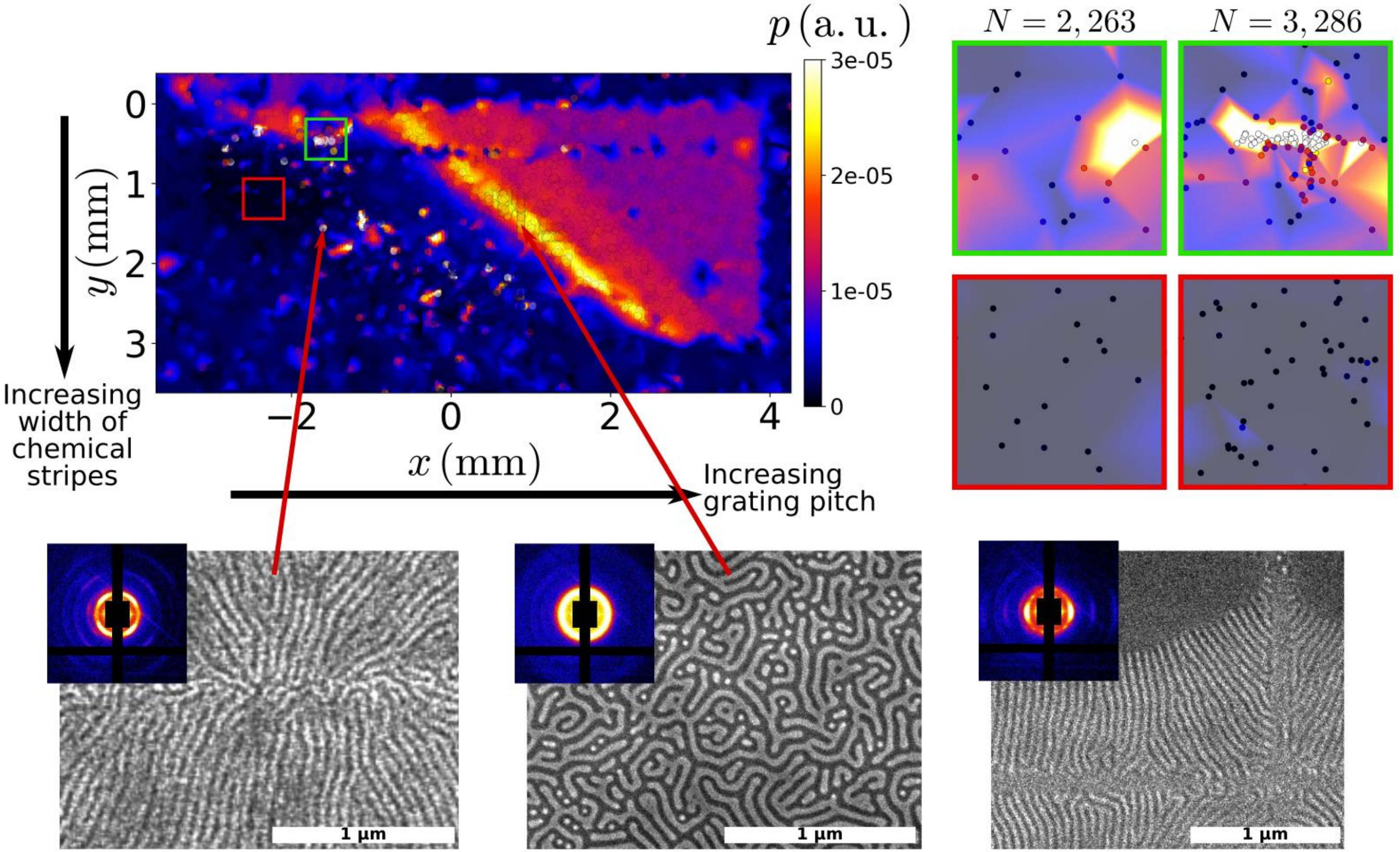




Directed Self-Assembly

Materials science

- Chemical grating controls ordering of block copolymer blends

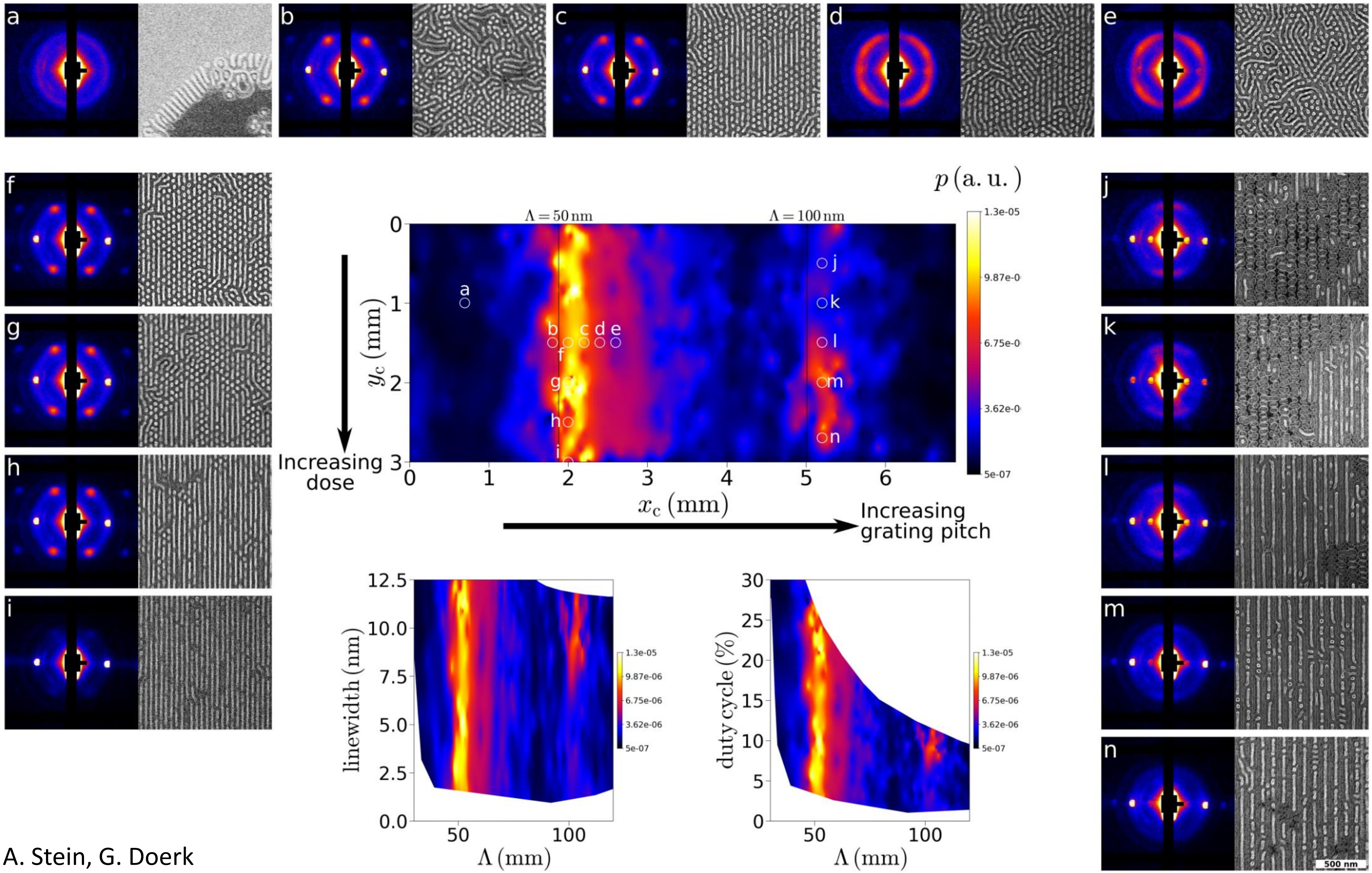


Demonstrates

- AE can find small patches of “unexpected” behavior; algorithm can seek “novelty”
- Enabled refinement of processing parameters

Directed Self-Assembly

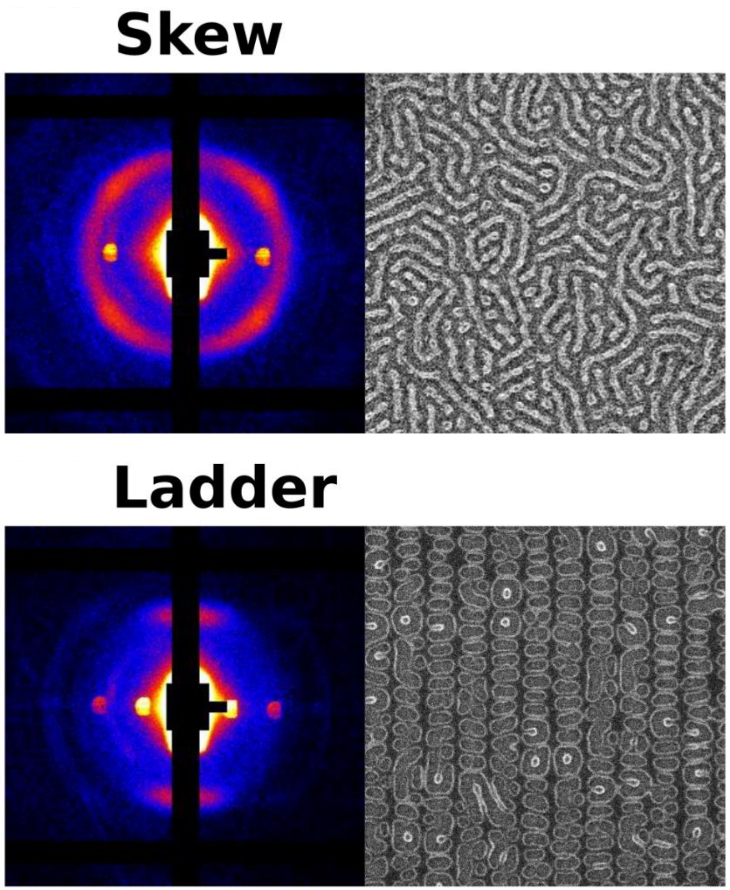
- Autonomy discovered numerous novel morphologies

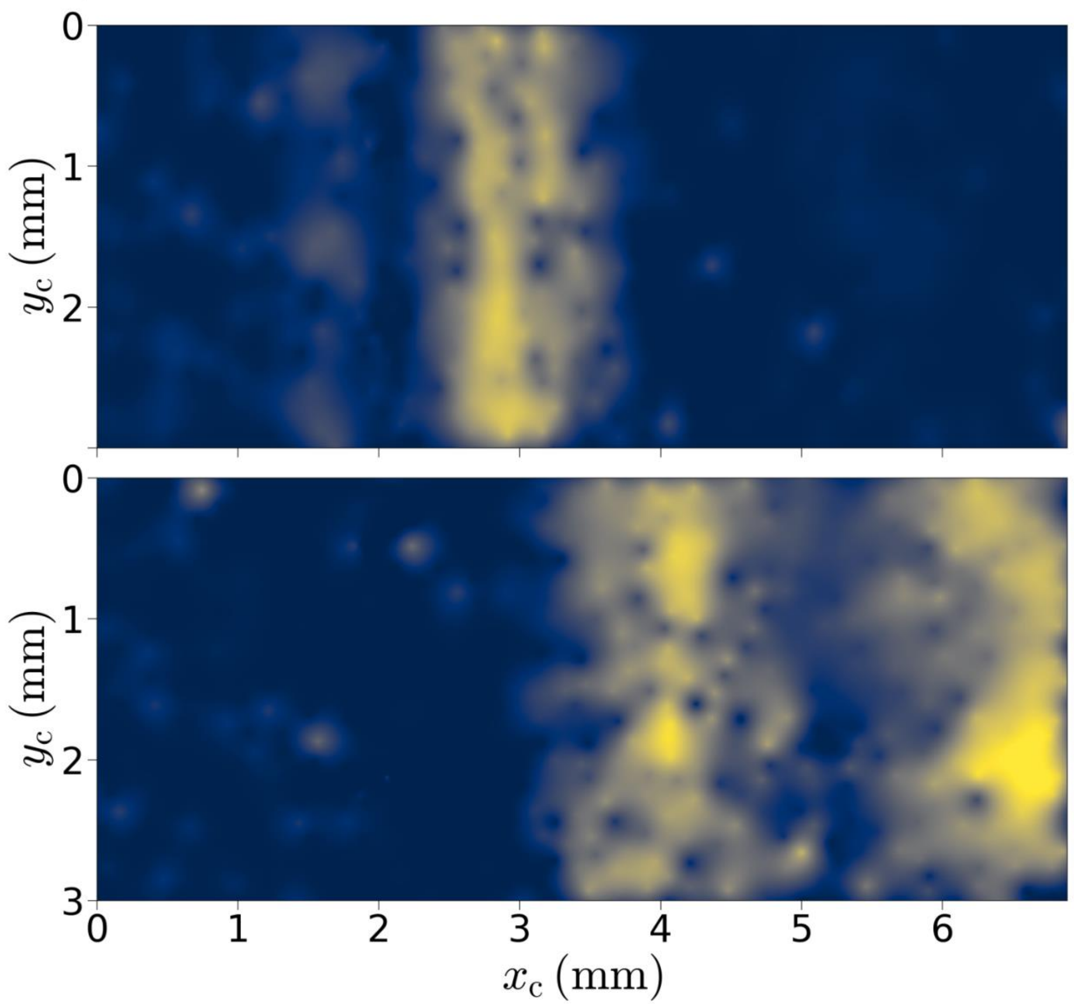


A. Stein, G. Doerk

Directed Self-Assembly

- Autonomy mapped new morphologies





A. Stein, G. Doerk

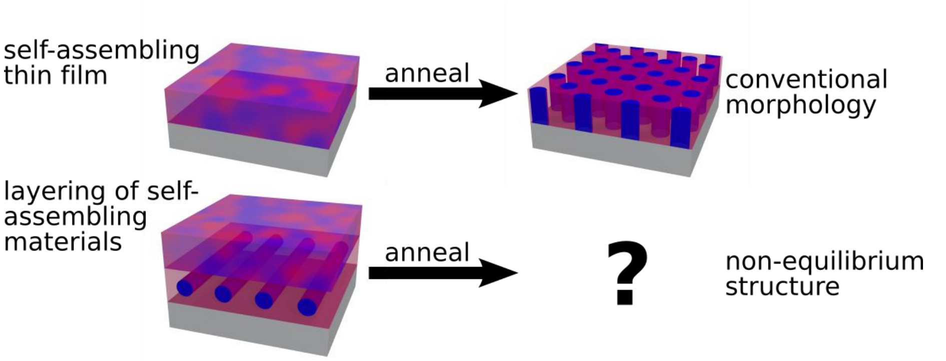
Block Copolymer Layering

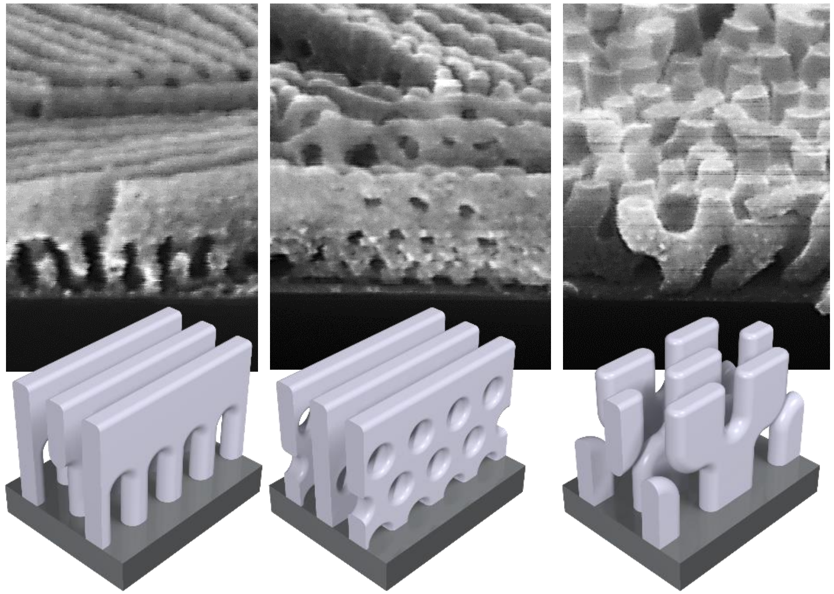
Materials science

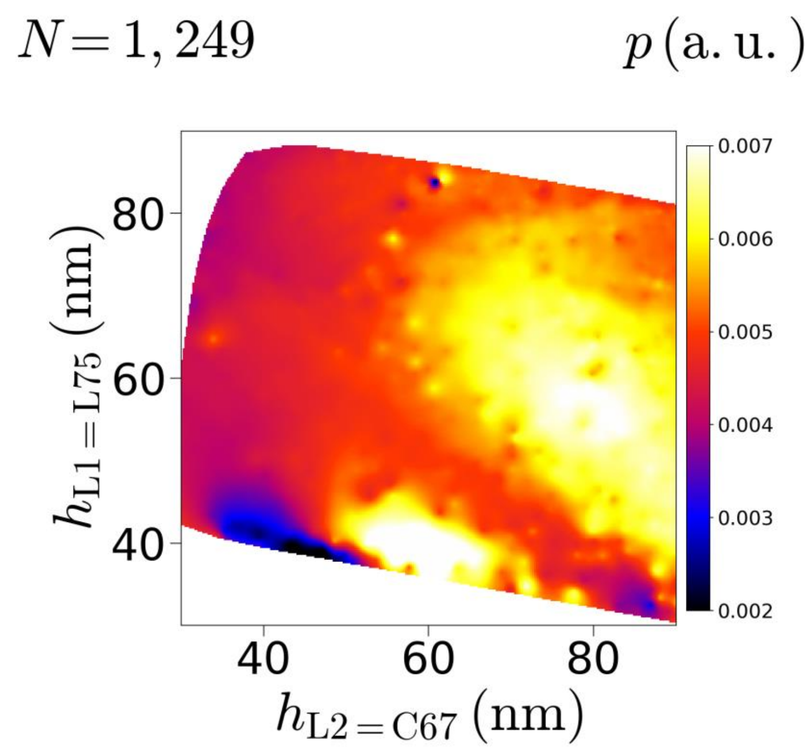
- Layering self-assembling materials can generate new (non-equilibrium) structures
- Enormous search space: material selection, layering sequence, annealing time, etc.

Experiment

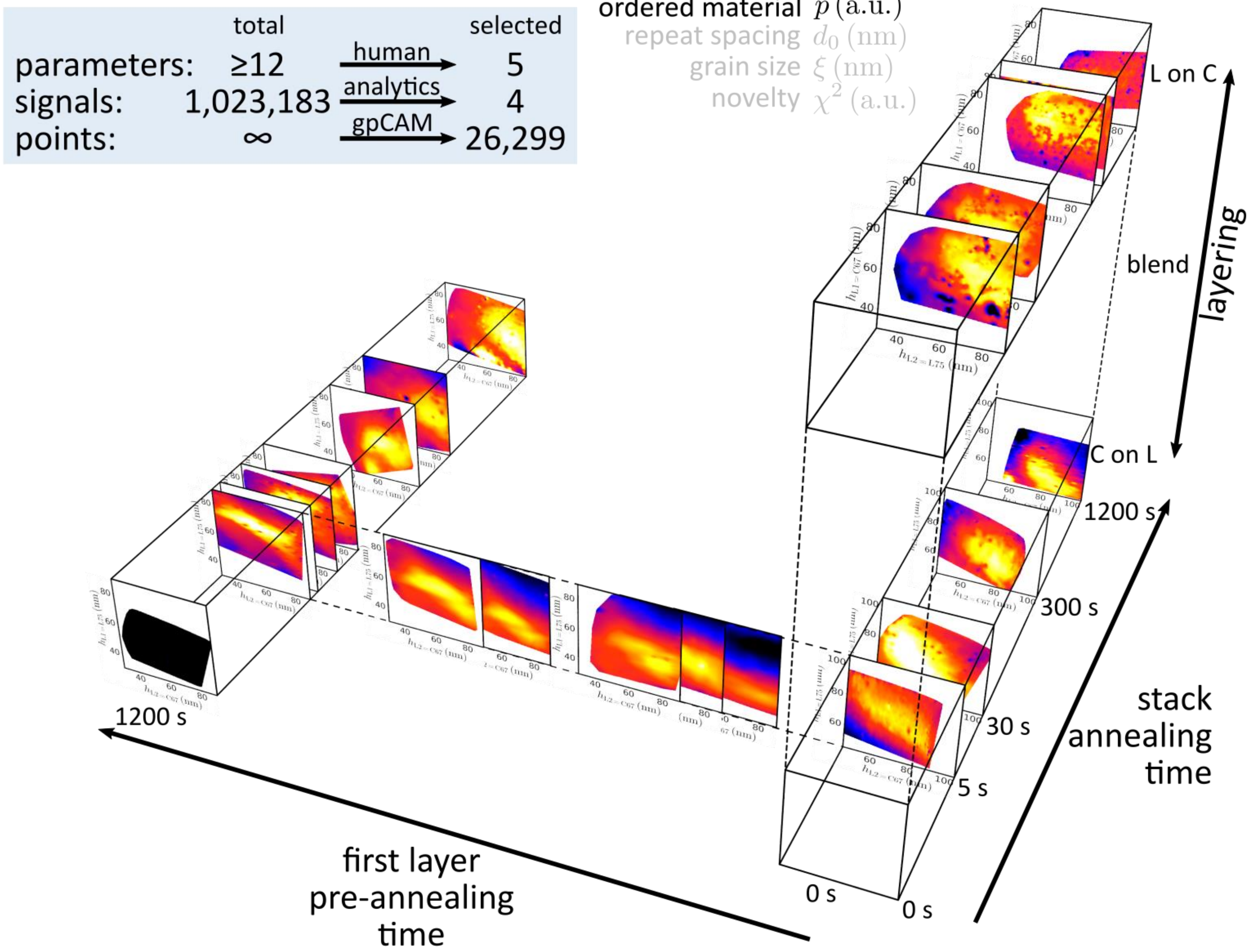
- Find rare, interesting transient states
- gpCAM explores 2D combinatorial gradients
- gpCAM selects next slice to make and measure







Sebastian Russell, Suwon Bae

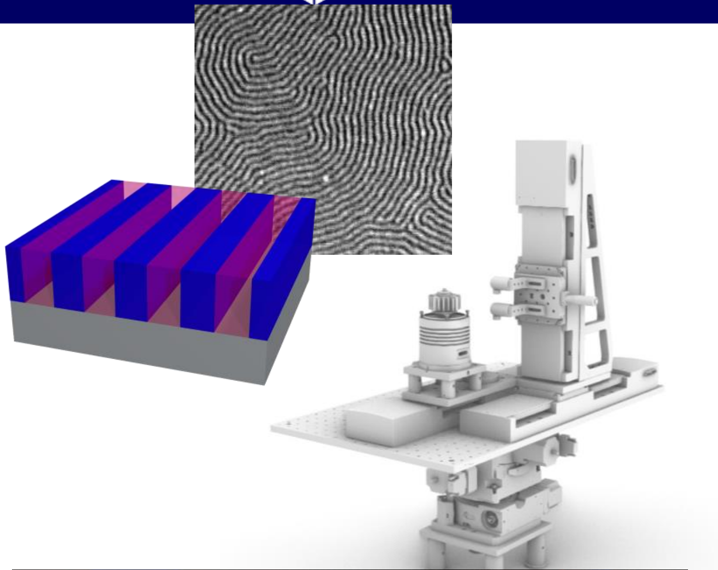


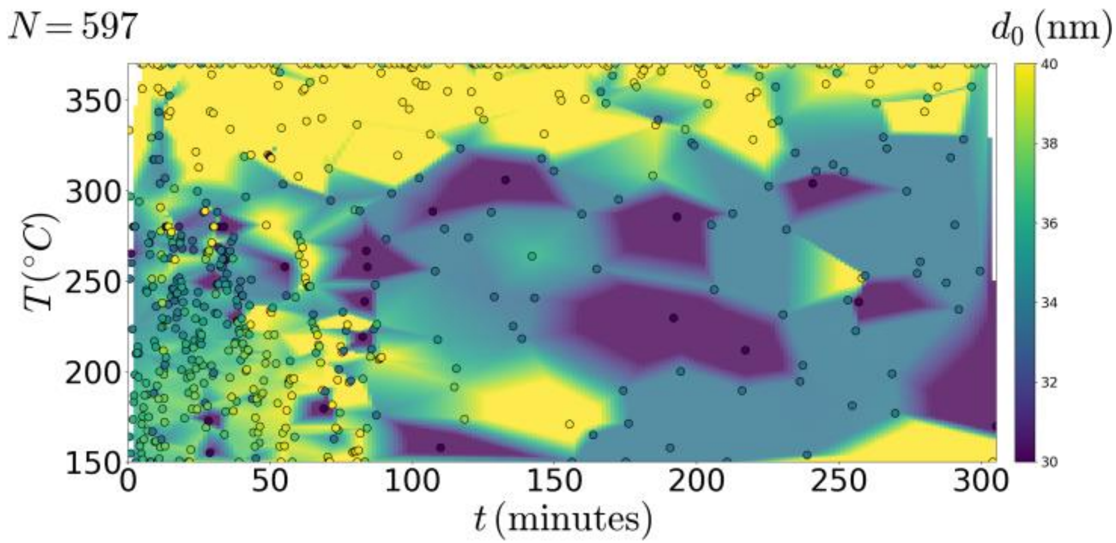
total
selected
parameters: ≥12 human 5
signals: 1,023,183 analytics 4
points: ∞ gpCAM 26,299
ordered material p (a.u.)
repeat spacing d_0 (nm)
grain size ξ (nm)
novelty χ^2 (a.u.)
L on C
blend
layering
C on L
1200 s
300 s
30 s
5 s
0 s
0 s
1200 s
stack annealing time
first layer pre-annealing time

Real-time Annealing of BCPs

Materials processing

- Block copolymer ordering depends on temperature and time
- Photo-thermal annealer allows thin film heating to be controlled
  - Fast and local heating, large gradients, realtime control
- Establish a thermal gradient, and measure changes over time





Demonstrates

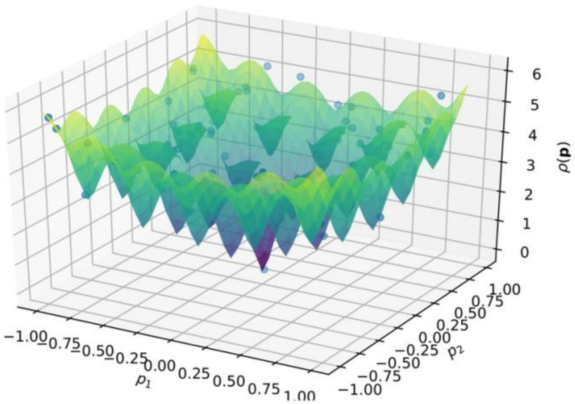
- Exploration of material *during* ordering

Collaboration: P. Majewski (U. Warsaw)

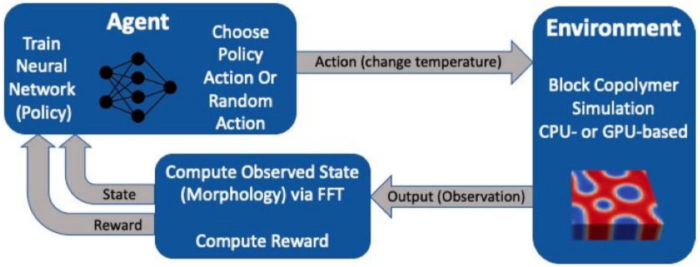
Future: Decision Methods

Brookhaven National Laboratory

- **Gaussian process**
  - Can be easily applied to any space; no pre-training required
  - Easy to tweak to suit experimental needs (target, cost, etc.)
  - Directly yields model uncertainty and termination criterion
  - Future: physics priors for guidance and hypothesis testing



- **Reinforcement learning**
  - Deep learning encodes “policy” for what action to take from a given state
  - Need training data (can pre-train on simulations)
  - *Collaboration with ExaLearn, CSI.*



Alexander et al. *Inter. J. H.P.C. App.* **2021**, 598

- **Bayesian model averaging**
  - Assume dynamics are part of a known class of models (plus a discrepancy term)
  - Real-time control guides towards target; data constricts distribution of likely models
  - As knowledge of dynamics improves, real-time control improves
  - *Collaboration with Kris Reyes (U. Buffalo, CSI)*

Future of AE

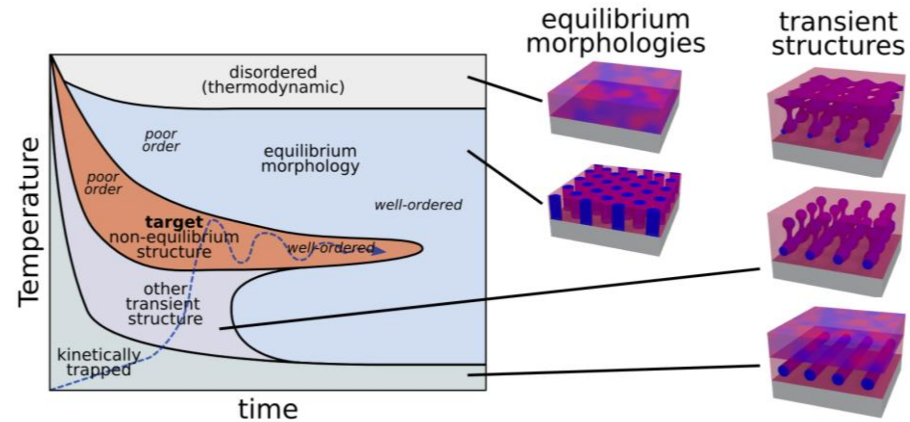
Brookhaven National Laboratory

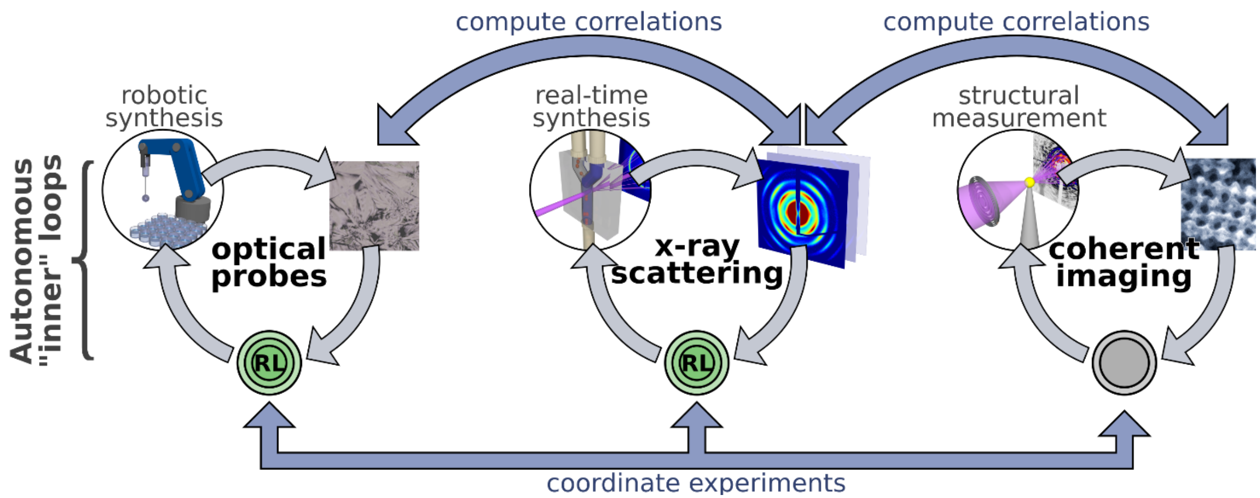
Physics-aware

- Control/constrain surrogate, kernel, cost, etc.
- “Prior” from simulations

Realtime processing

- Photo-thermal platform for complex annealing histories (block copolymers)
- Liquid handling for real-time synthesis
- Requires fast surrogates, real-time controllers?



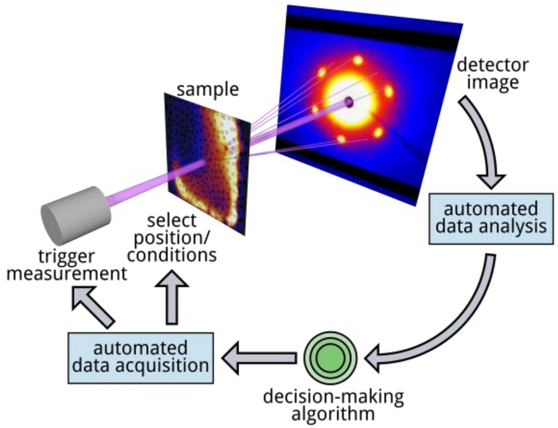


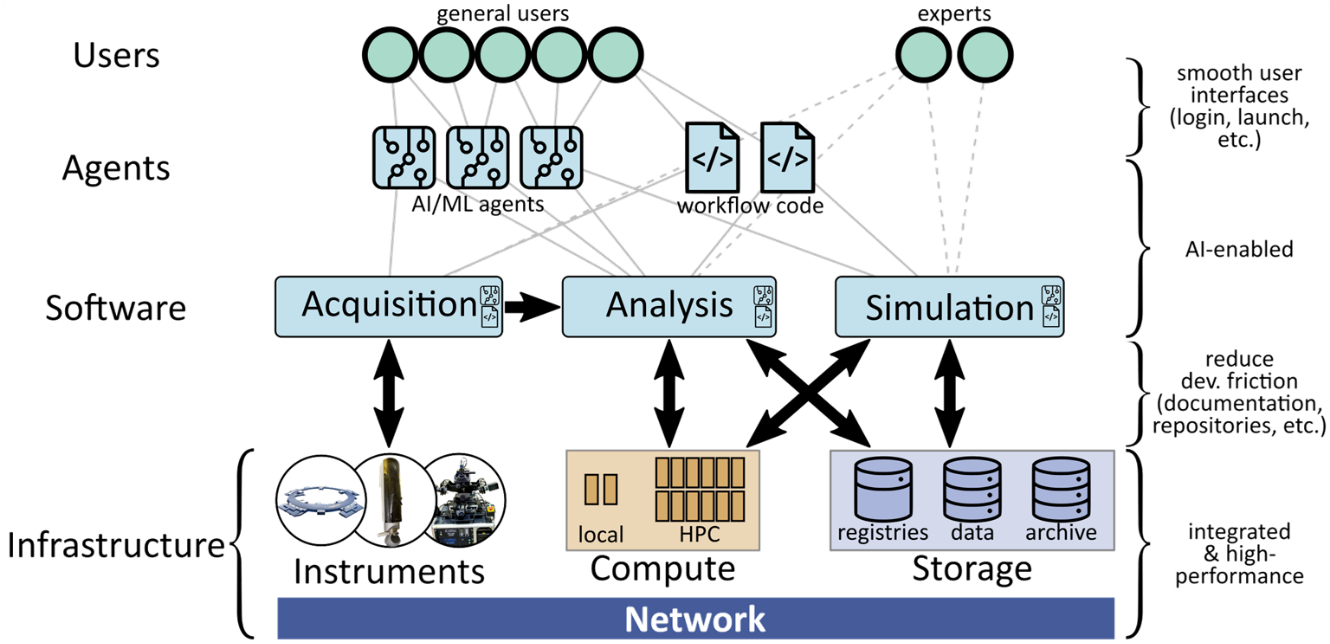
Multi-modal

- Balance tradeoffs in generality/specificity and rigor/speed
- Search for correlations across modules, experiments, etc.
- Transfer learning? Federated learning?

Future of Facility Science

- Advanced **computing** and **AI/ML**
- **Autonomous** experiments liberate scientists
- Aggregated and **open data**bases
- **Remote** analytics and instrument control
- Transition from a set of disconnected tools, into an AI/ML software-accessible **discovery ecosystem**
  - CFN strategic theme: Accelerated Nanomaterial Discovery
  - BNL initiative: Human-AI-facility integration
  - DOE: AI/ML for facilities

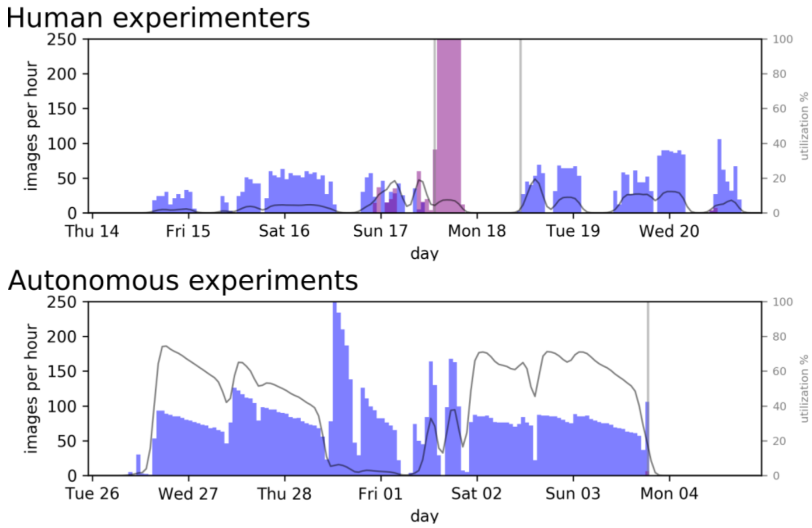




Summary

Brookhaven National Laboratory

- **Data Analytics**
  - ML powerful for scientific data (“understands” X-ray scattering)
  - Integrate domain expertise
- **Autonomous experiments** can explore parameter spaces
  - Increases beamtime utilization
  - More efficient exploration
  - GP directly yields model and uncertainty



- **Outlook**
  - Apply to even more challenging materials problems
    - Physics-informed
    - Realtime processing/synthesis
    - Multi-modal

Acknowledgments

- Machine vision
  - Tamara Berg (UNC Chapel Hill)
  - Alex Berg (UNC Chapel Hill)
  - Hadi Kiapour (UNC Chapel Hill)
- Deep learning
  - Dantong Yu (BNL, CSI)
  - Minh Hoai Nguyen (Stony Brook)
  - **Boyu Wang** (Stony Brook)
  - Hong Qin (Stony Brook)
  - **Ziqiao Guan** (Stony Brook)
- Data analysis
  - Jiliang Liu (BNL, CFN)
  - Julien Lhermitte (BNL, CFN)
  - Youngwoo Choo (Yale)
- Visualization
  - Wei Xu (BNL, CSI)
  - Klaus Mueller (Stony Brook)
- Autonomous implementation
  - **Masa Fukuto** (BNL, NSLS-II)
  - **Marcus Noack** (**CAMERA**)
  - Jamie Sethian (CAMERA)
  - Ruipeng Li (BNL, NSLS-II)
  - Esther Tsai (BNL, CFN)
- Beamlines (BNL, NSLS-II)
  - Mikhail Zhernenkov
  - Guillaume Freychet
  - Lutz Wiegart
  - Sanjit Ghose
  - Dan Olds
  - Phillip Maffettone
  - Joshua Lynch
  - Tom Caswell
- Directed assembly (BNL, CFN)
  - Greg Doerk
  - Aaron Stein
  - Sebastian Russell
  - Suwon Bae
- Electrospray (Yale, U.Penn)
  - Chinedum Osuji
  - Kristof Toth
- Photo-thermal annealing (U. Warsaw)
  - Pawel Majewski
  - Andrzej Sitkiewicz
  - Arkadiusz Leniart
- Polymer composites (Columbia)
  - Sanat Kumar
  - Andrew Jimenez
  - Alejandro Krauskopf
- Nanoparticle superlattices (U. Penn)
  - Chris Murray
  - Katherine Elbert
- Nanorod assembly (AFRL)
  - Richard Vaia
  - Jason Streit
- 3D printing (NIST)
  - Jon Seppala
  - Tyler Martin
- Metal dealloying (SBU)
  - Karen Chen-Wiegart
  - Chonghang Zhao

CAMERA Applied Math

U.S. DEPARTMENT OF ENERGY | Office of Science

BROOKHAVEN NATIONAL LABORATORY | Center for Functional Nanomaterials | National Synchrotron Light Source II | Computational Science Initiative

Penn
Yale University
AFRL
NIST
UNIVERSITY OF WARSAW
Stony Brook University
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK