Autonomous analysis of synchrotron X-ray experiments with applications to metal nanoparticle synthesis

Anthony Fong, Lenson Pellouchoud, **Sathya Chitturi**, Christopher Tassone
Stanford University & Stanford Synchrotron Radiation Lightsource

Outline

1. Pipeline for nanoparticle optimization based on Small-Angle X-ray Scattering (SAXS) data (Anthony/Lenson)
   - Automatic analysis of SAXS data
   - Bayesian optimization of reaction conditions

2. Automatic analysis of powder diffraction / Wide-angle X-ray scattering (WAXS) data
   - Lattice parameter estimation using CNNs
   - Effects of experimental non-idealities on predictions
   - Approaches for tackling multiphase data

Introduction

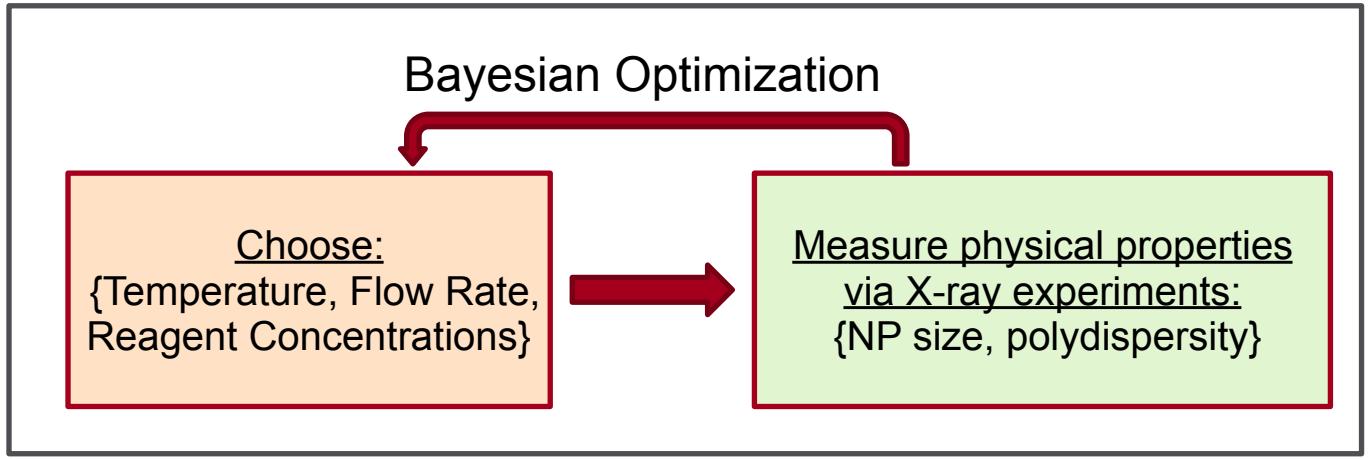
SLAC

Project Goals

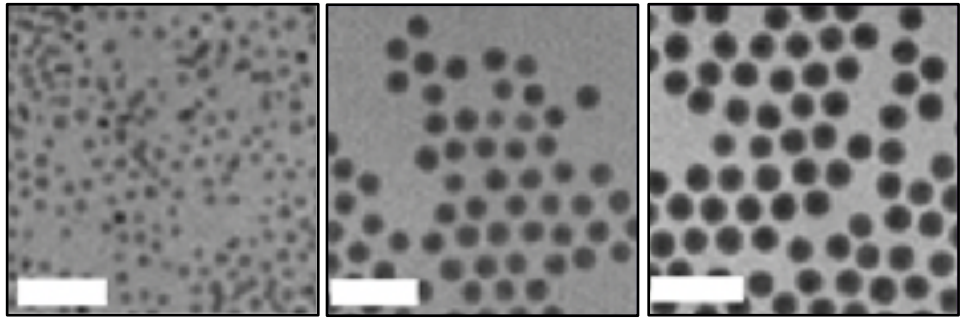
1) Synthesize libraries of monodisperse Pd nanoparticles with user specified sizes

2) Determine synthetic conditions for desired nanoparticle sizes

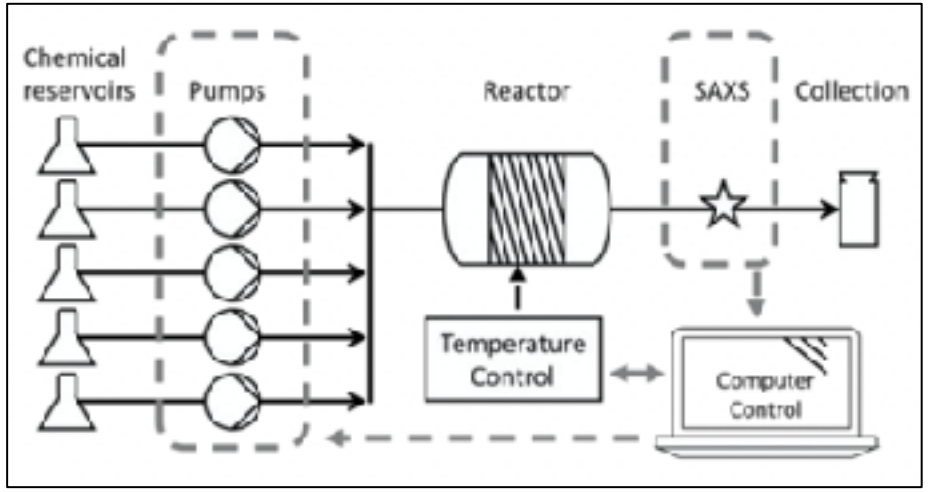
3) Perform these experiments in an autonomous fashion using a flow-reactor setup



TEM of Pd NPs with different sizes



Automatic Synthesis Schematic

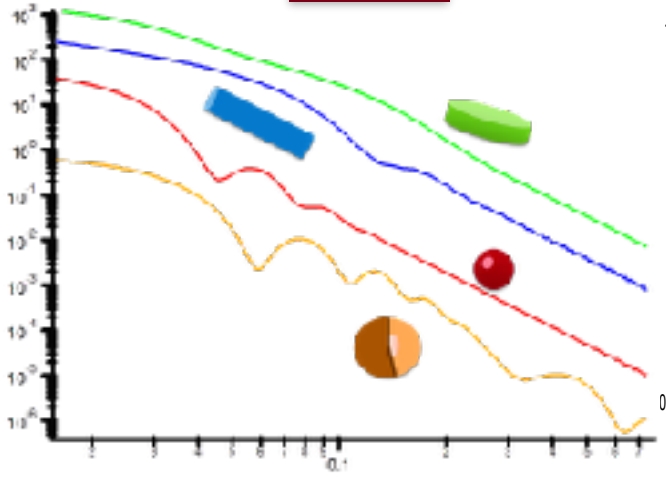


Fong, A et al. Utilization of machine learning to accelerate colloidal synthesis and discovery (in review).

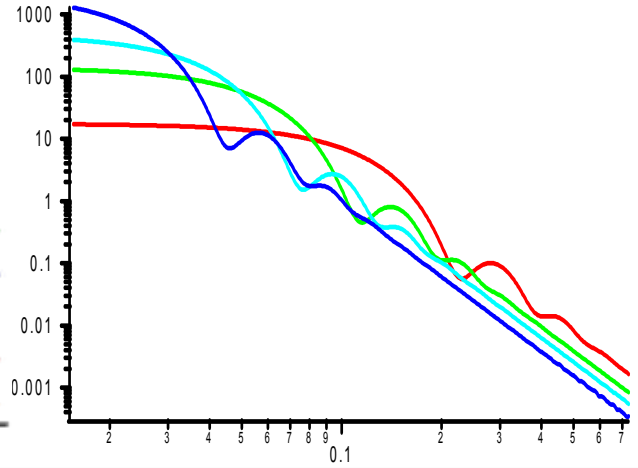
Matching the Probe to the Metric

Small Angle X-ray Scattering (SAXS) characterizes

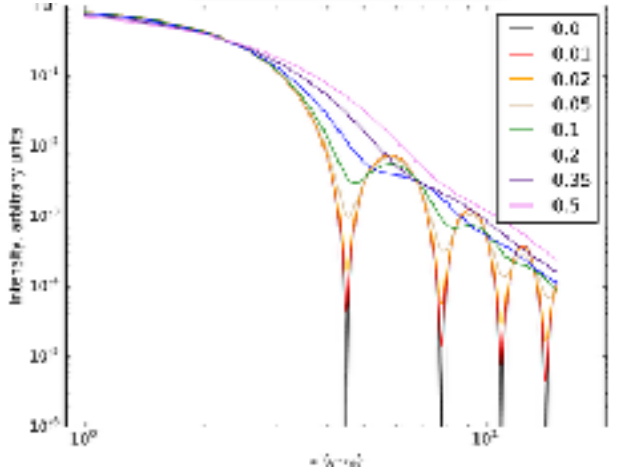
Shape



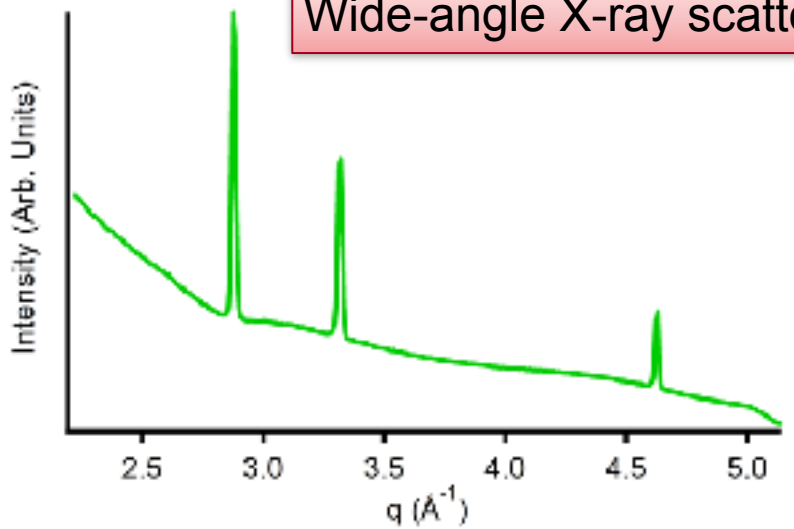
Size



Polydispersity

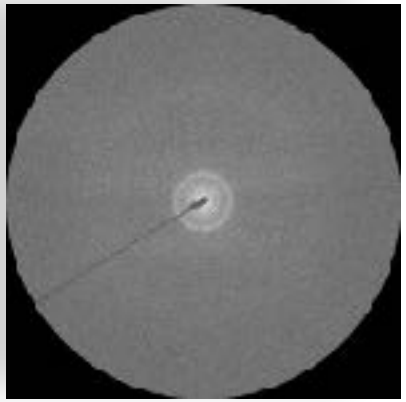


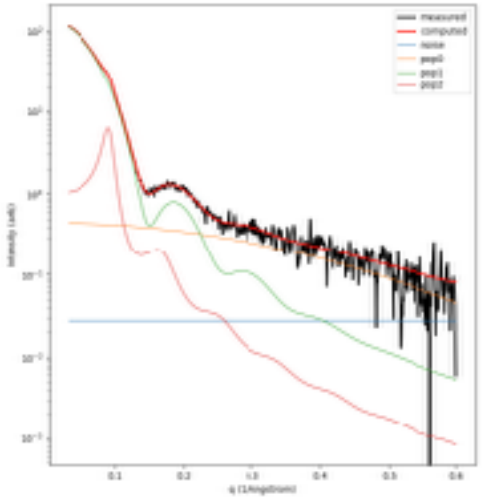
Wide-angle X-ray scattering (WAXS) characterizes



- Phase
- Composition
- Degree of Crystallinity
- Strain
- Crystallite Size

Automated Fitting of SAXS data





$$\mathbf{s} = \begin{bmatrix} I_{\text{noise}} \\ I_{0,\text{pre}} \\ I_{0,\text{dil}} \\ I_{0,\text{cond}} \\ r_{\text{g,pre}} \\ r_{\text{dil}} \\ \sigma_{\text{dil}} \\ r_{\text{cond}} \\ \sigma_{\text{cond}} \\ r_{\text{h,cond}} \\ \phi_{\text{cond}} \end{bmatrix}$$

SAXS data populations

-- Dilute spherical NPs (desired)
-- Condensed / disordered
-- Superlattice
-- Precursor (metal ligand)

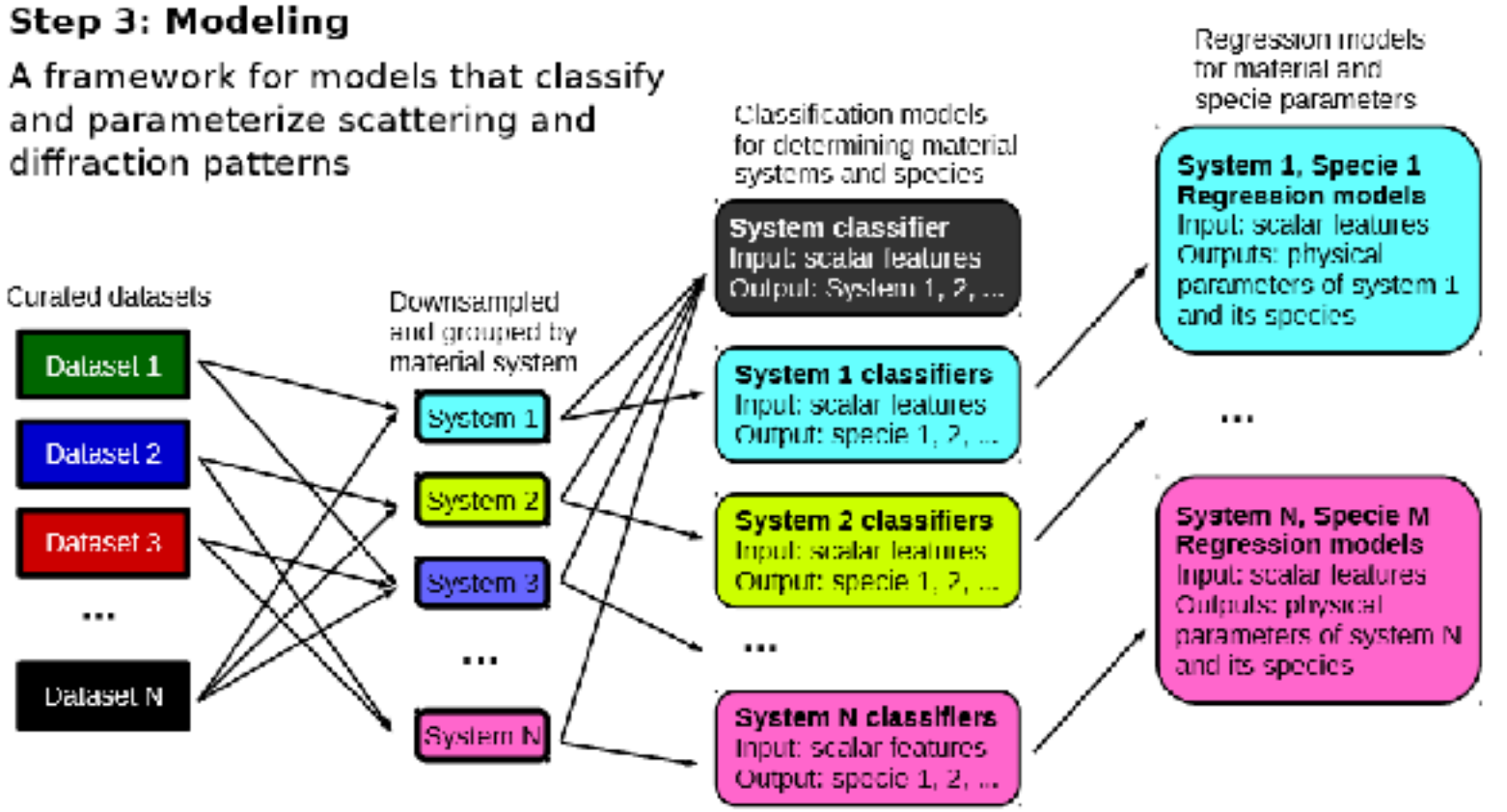
Data Fitting Approach

$$\begin{aligned} I_{\mathbf{s}}(q) = I_{\text{noise}} &= I_{\text{noise}} \\ &+ I_{\text{pre}}(q) &+ I_{0,\text{pre}} * \overline{\langle f^2_{\text{G-P}} \rangle}[r_{\text{g,pre}}, D_{\text{pre}}](q) \\ &+ I_{\text{dil}}(q) &+ I_{0,\text{dil}} * \overline{\langle f^2_{\text{sph}} \rangle}[\bar{r}_{\text{dil}}, \sigma_{\text{dil}}](q) \\ &+ I_{\text{cond}}(q) &+ I_{0,\text{cond}} * \overline{S_{\text{hs}}}[r_{\text{h,cond}}, \phi_{\text{cond}}](q) * \overline{\langle f^2_{\text{sph}} \rangle}[\bar{r}_{\text{cond}}, \sigma_{\text{cond}}](q) \end{aligned}$$

$$c(\mathbf{s}) = \frac{1}{\Omega_q} \int_q \mathrm{d}q \left[\frac{1}{\sigma_I(q)^2} \left[\log(I_{\text{exp}}(q)) - \log(I_{\mathbf{s}}(q)) \right]^2 \right]$$

$$\mathbf{s}^* = \underset{\mathbf{s}}{\operatorname{argmin}}\, c(\mathbf{s})$$

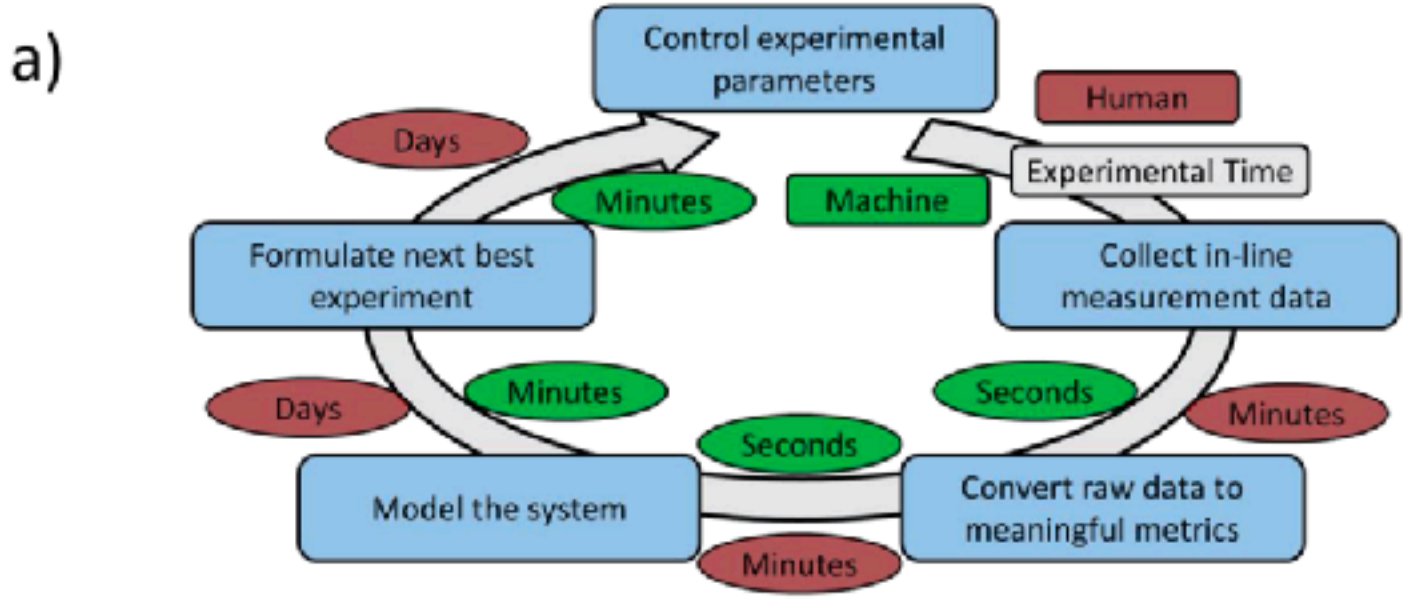
Using ML to improve SAXS Fitting

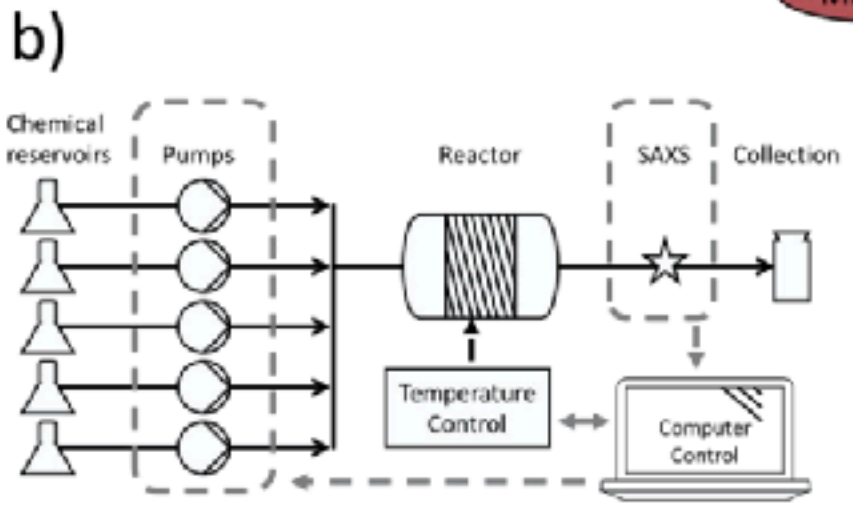


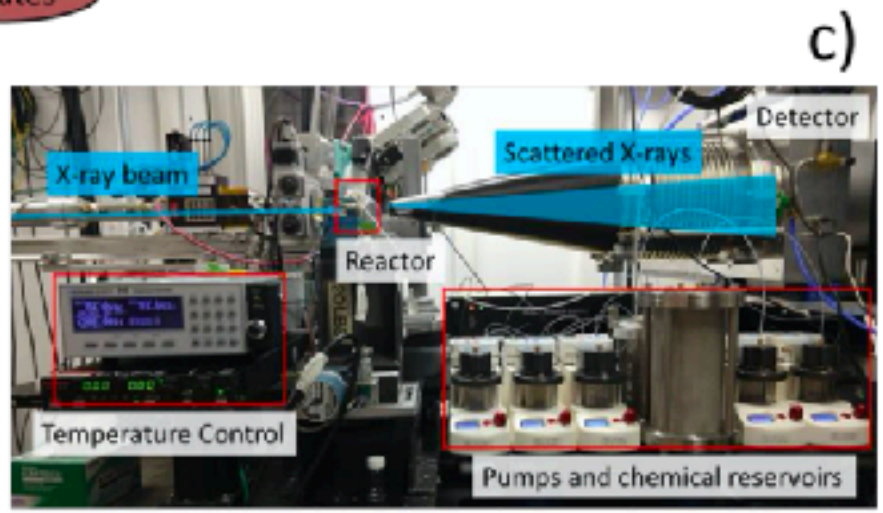
Code: xrsdkit

https://github.com/scattering-central/xrsdkit

Bayesian Optimization to find synthesis conditions



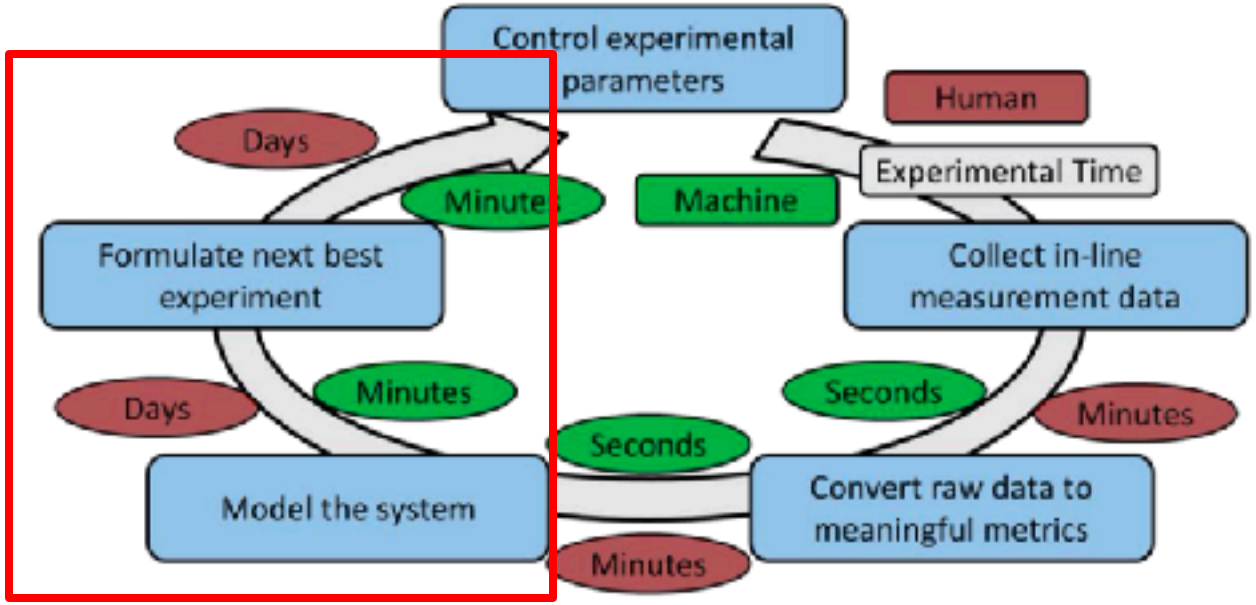




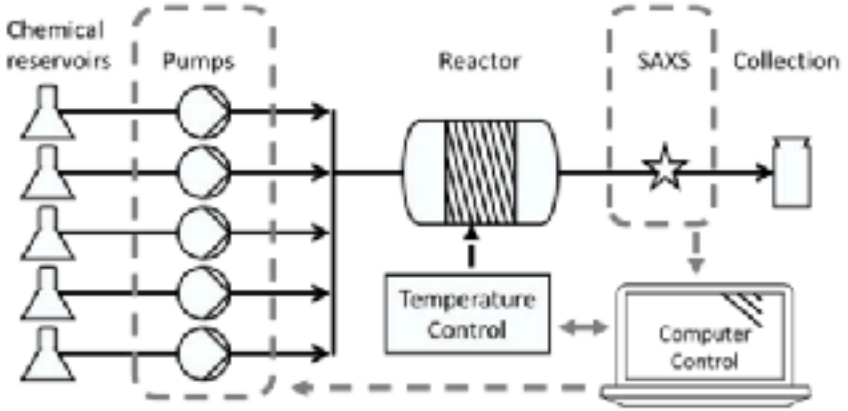
Bayesian Optimization to find synthesis conditions

SLAC

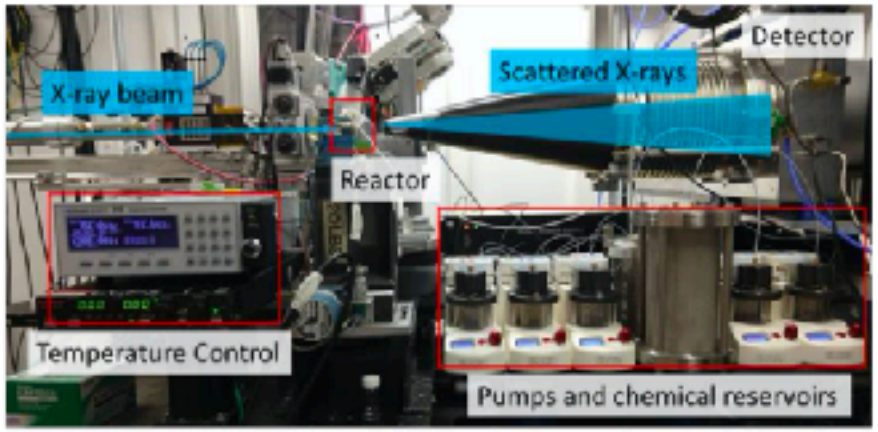
a)



b)



c)



Details of Bayesian Optimization

$$x = \begin{bmatrix} T_{reac} \\ F_{tot} \\ x_{ODE} \\ x_{TOP} \\ x_{oley} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} p_{\mathrm{dil}} \\ p_{\mathrm{cond}} \\ p_{\mathrm{cryst}} \\ I_0 \\ \tilde{I}_{0,\mathrm{dil}} \\ \bar{r}_{\mathrm{dil}} \\ \sigma_{\mathrm{dil}} \end{bmatrix}$$

1. Presence of dilute nanoparticle products (binary): $p_{\mathrm{dil}} \in \{-1, 1\}$
2. Presence of condensed (disordered) nanoparticle products: $p_{\mathrm{cond}} \in \{-1, 1\}$
3. Presence of crystalline (superlattice) nanoparticle products: $p_{\mathrm{cryst}} \in \{-1, 1\}$
4. Intensity at $q = 0$: I_0, in $[0, \infty)$
5. Fraction of I_0 attributable to dilute nanoparticle products: $\tilde{I}_{0,\mathrm{dil}} = I_{0,\mathrm{dil}}/I_0$, in $[0, 1]$
6. Mean radius of dilute nanoparticle products: $\bar{r}_{\mathrm{dil}}$, in $(0, \infty)$
7. Fractional standard deviation of radii of dilute nanoparticle products: σ_{dil}, in $(0, \infty)$

Individual GPs

Enforce

- Size
- Monodispersity
- Large Yield
- Desired Population (dilute)

$$g(\bar{y}, \sigma_y^2) = \prod_j [g_j(\bar{y}_j, \sigma_{yj}^2)]$$

$$x^* = \underset{(x)}{argmax}\, g(\bar{y}, \sigma_y^2)$$

$$g_j(\bar{y}_j, \sigma_{yj}^2) = P(|y_j - y_j^*| < |y_j^+ - y_j^*|)$$

Acquisition Function Details

$$\mathbf{y} = \begin{bmatrix} p_{\text{dil}} \\ p_{\text{cond}} \\ p_{\text{cryst}} \\ I_0 \\ \tilde{I}_{0,\text{dil}} \\ \bar{r}_{\text{dil}} \\ \sigma_{\text{dil}} \end{bmatrix}$$

Constraints Coded into Acquisition Functions

$$g_0(\bar{p}_{\text{dil}}, \sigma^2_{p\text{dil}}) = P(p_{\text{dil}} > 0)$$

$$g_1(\bar{p}_{\text{cond}}, \sigma^2_{p\text{cond}}) = P(p_{\text{cond}} < 0)$$

$$g_2(\bar{p}_{\text{cryst}}, \sigma^2_{p\text{cryst}}) = P(p_{\text{cryst}} < 0)$$

$$g_3(\bar{I}_0, \sigma^2_{I0}) = P(I_0 > 10)$$

$$g_4(\bar{\tilde{I}}_{\text{dil}}, \sigma^2_{\tilde{I}\text{dil}}) = P(\tilde{I}_{\text{dil}} > 0.9)$$

$$g_5(\bar{\bar{r}}_{\text{dil}}, \sigma^2_{\bar{r}\text{dil}}) = P(|\bar{r}_{\text{dil}} - \bar{r}^*_{\text{dil}}| < |\bar{r}^+_{\text{dil}} - \bar{r}^*_{\text{dil}}|)$$

$$g_6(\bar{\sigma}_{\text{dil}}, \sigma^2_{\sigma\text{dil}}) = P(\sigma_{\text{dil}} < 0.2)$$

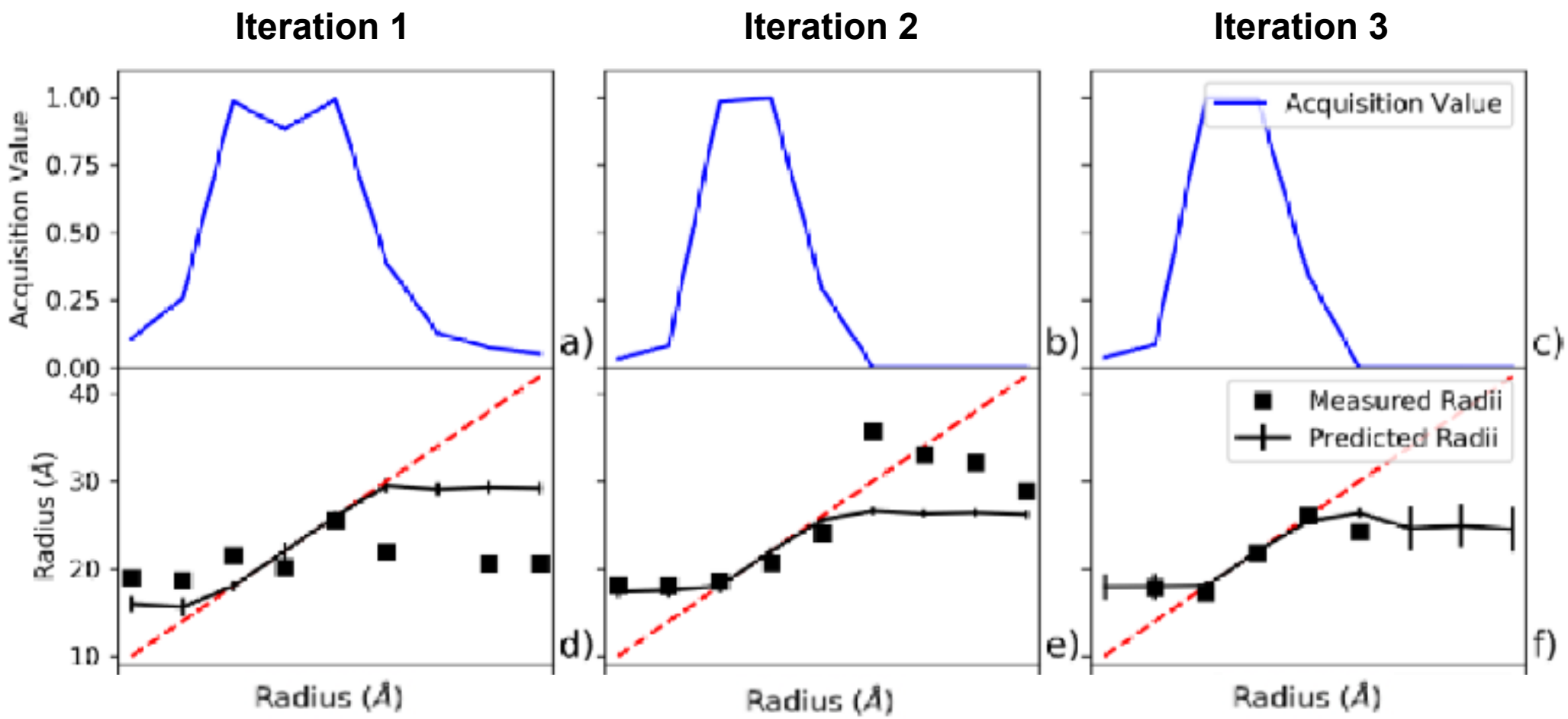
Closest training example

target

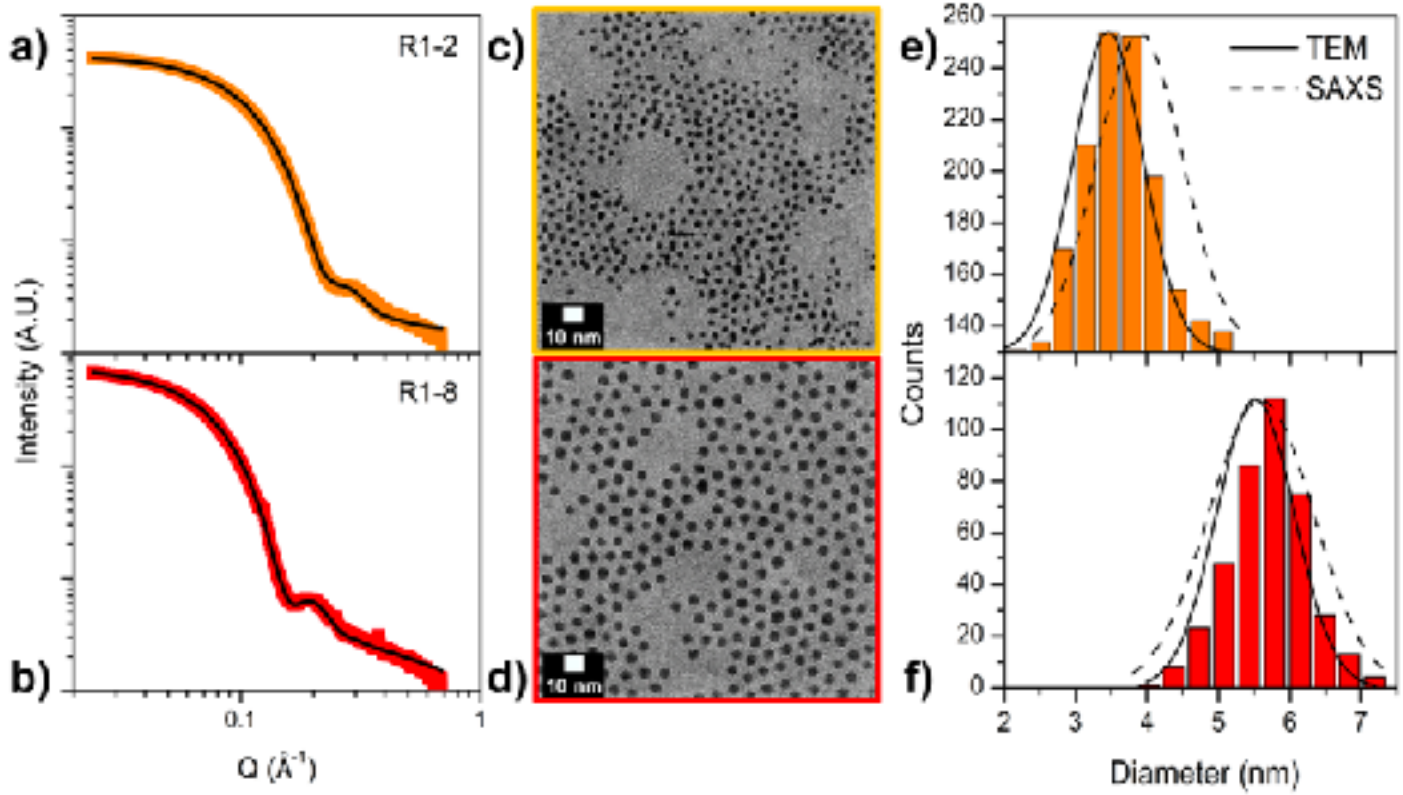
Automating Understanding of Synthesizability

SLAC

We don't want to synthesize a single catalyst, we want to synthesize a library



Sample Results for NP synthesis

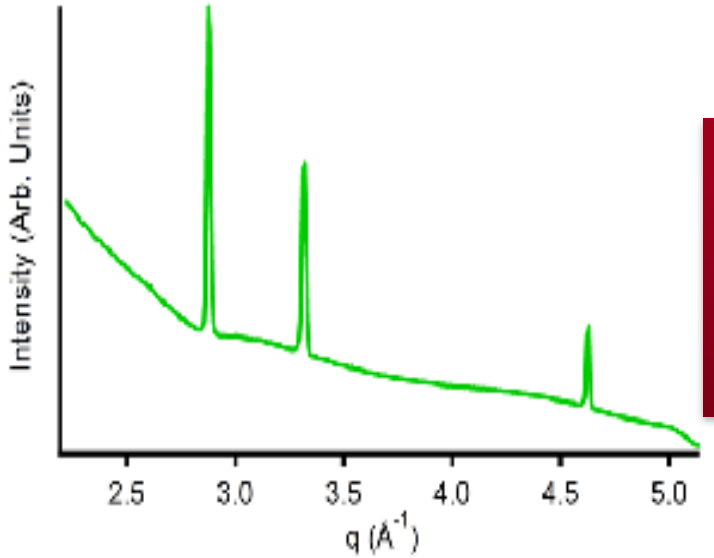


Reaction	SAXS [nm]	TEM [nm]	Flowrate [µL]	Temp. [°C]	$Pd(acac)_2$ [M]	TOP [M]	oleylamine [M]
R1-2	3.9 ± 0.6	3.5 ± 0.5	116	275	0.0168	0.0719	0.0036
R1-8	5.6 ± 0.7	5.5 ± 0.5	20	284	0.0198	0.0845	0.5117

Future Directions: Binary and Ternary Systems

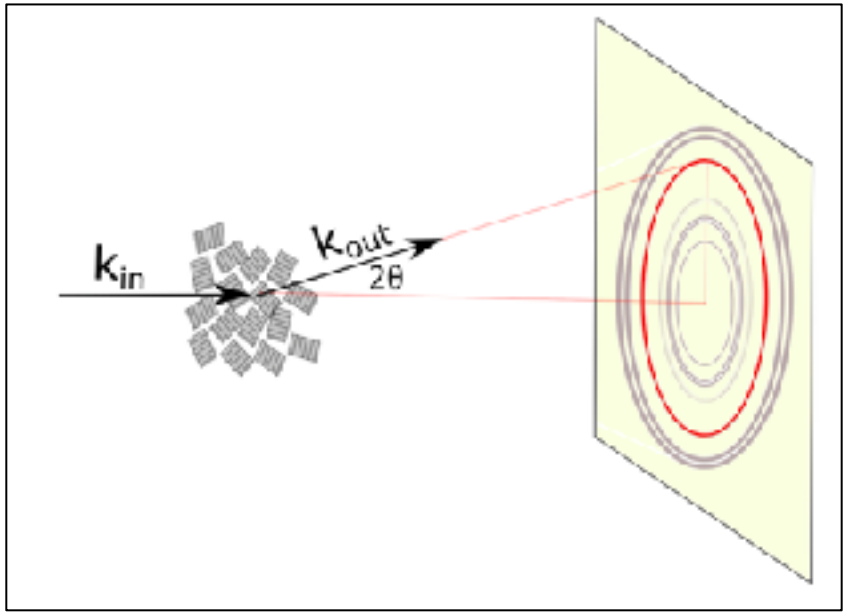
- Multimodal SAXS/WAXS experiments
  - Optimize for desired phase, composition and particle size
  - E.g. make NPs of Pt4Sn alloy
- Automate analysis of 1D powder/WAXS data using ML
  - Crystal System / Lattice parameters
  - Predict phases in data based on known priors

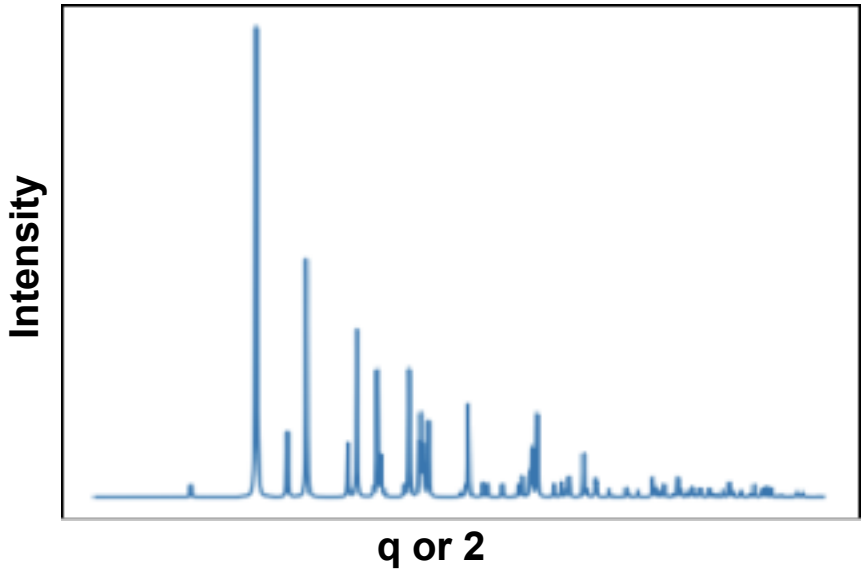
Wide-angle X-ray scattering (WAXS) characterizes



- Phase
- Composition
- Degree of Crystallinity
- Strain
- Crystallite Size

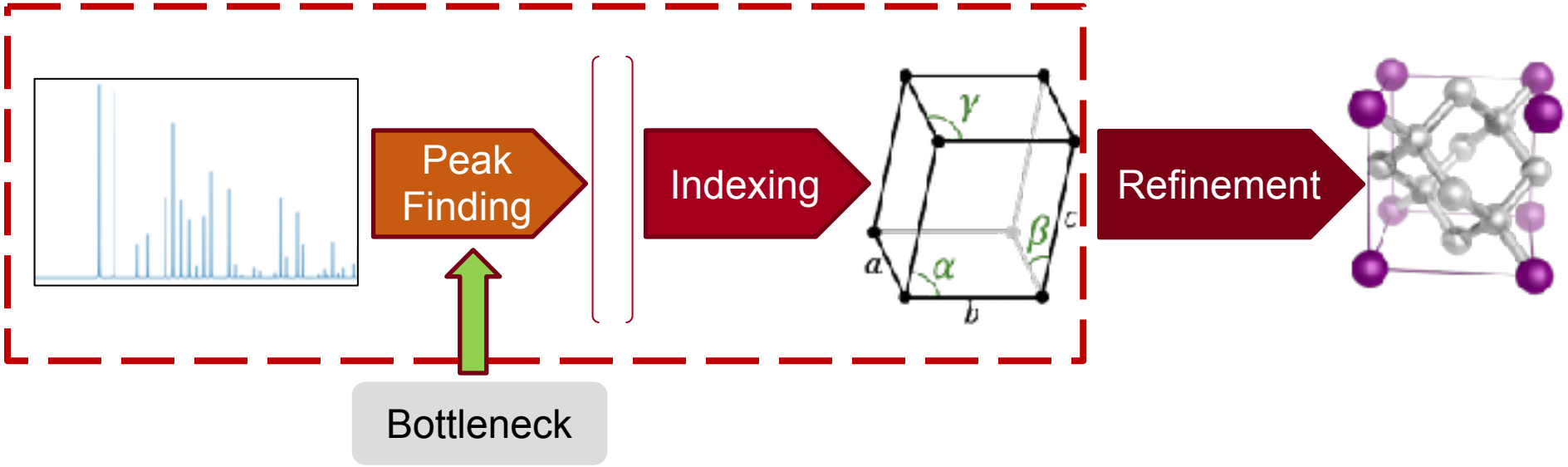
Introduction to WAXS / Powder Diffraction

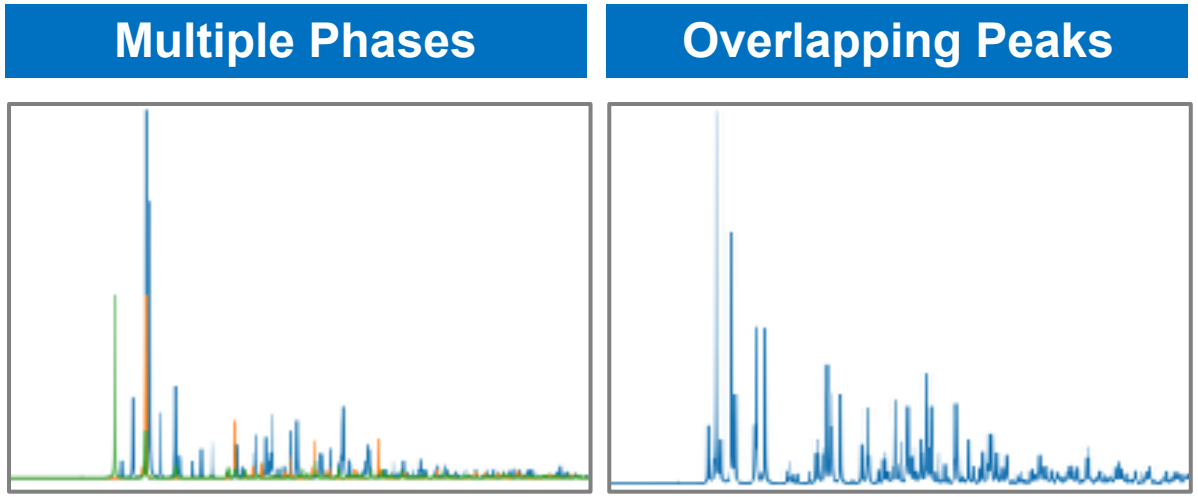




- **Peak locations**: crystal system / lattice parameters
- **Peak intensities:** space group / position of atoms, texturing
- **Peak shape:** sample / instrument effects
- **Phase Identification:** fingerprinting

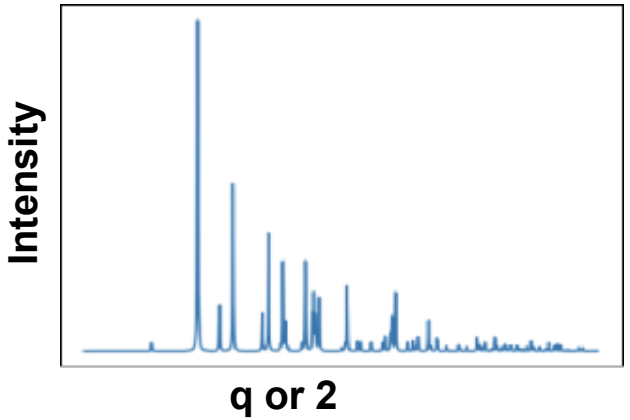
Conventional Analysis Pipeline

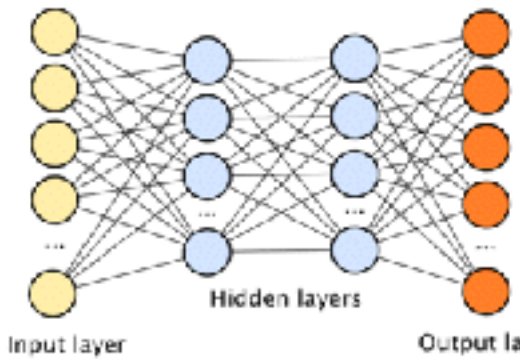


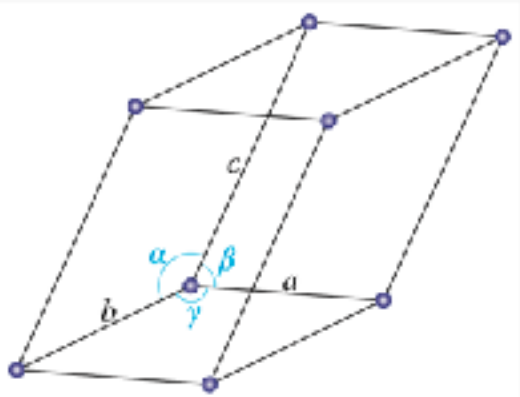


Problem Formulation

Goal: Develop machine learning models to directly **estimate lattice parameters** from powder diffraction data







Previous Work

- [1]Automatic space group and crystal system prediction
- [2]Lattice parameter prediction for neutron diffraction data

Potential Advantages

- Robust to Noise
- No Intervention Needed
- Fast

1. Park, W. B. et al., IUCrJ (2017).
2. Doucet, M. et al., Machine Learning: Science and Technology 2.2 (2020).

Previous ML Analysis of Powder Diffraction Data

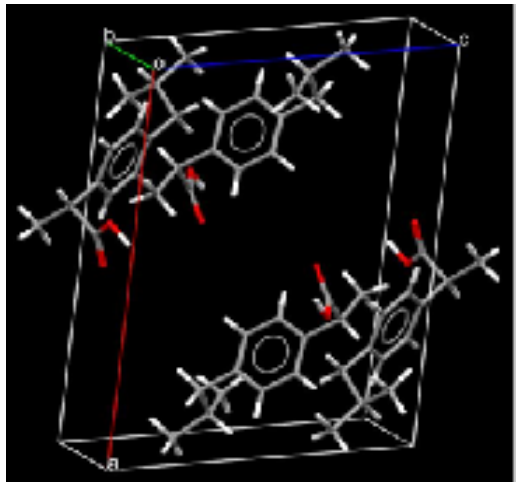
- Determination of **Space Group** and **Crystal System** directly from diffractogram
  - 230 Space Groups, 7 Crystal Systems

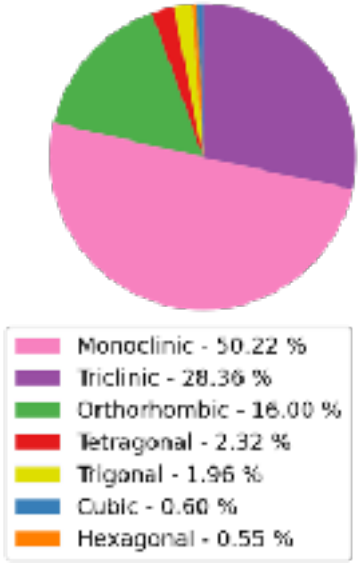
	Method	Accuracy	Type	Output	Dataset Size	Datasets
Park et al. 2017	CNN	81.14%	Sim	230 space groups	150000	ICSD
Oviedo et al. 2018	CNN + physics augmentation	89%	Sim/Exp	7 Crystal Systems	164 + 115	ICSD/Exp
Vecsei et al. 2018	FCNN	57%	Sim/Exp	230 space groups	128404 + 800	ICSD/RRUFF
Suzuki et al. 2018	RF - 10 peak positions	83.62%	Sim	230 space groups	188607	ICSD

Simulation Details

➢ **~ 1 million patterns simulated** from CIF structure files from Inorganic Crystal Structure Database and Cambridge Structural Database

Cambridge Structural Database

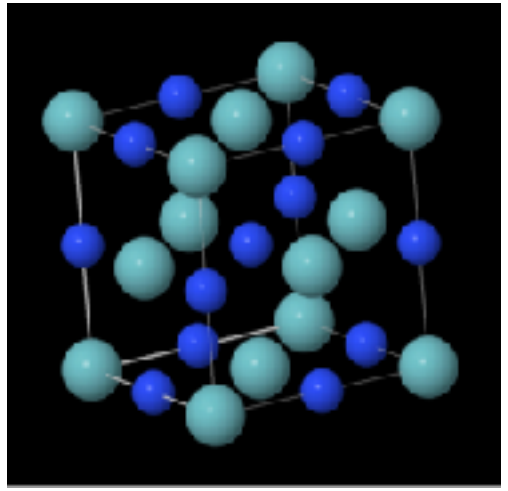


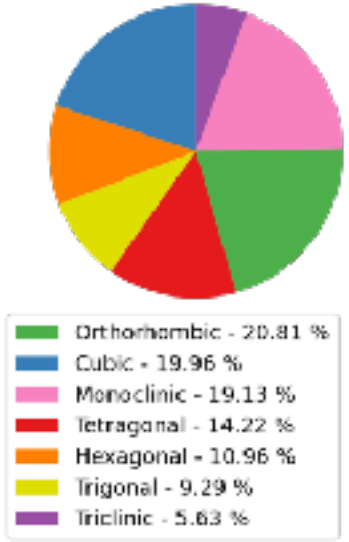


Ibuprofen

~ Large Unit Cells
~ Low symmetry

Inorganic Crystal Structure Database





NaCl

~ Small Unit Cells
~ High symmetry

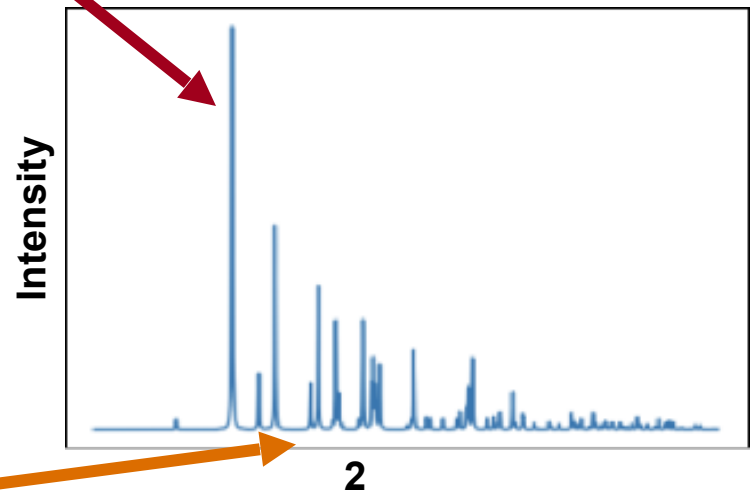
Simulation Details

$$I = \alpha(m_{hkl})\left(|F_{hkl}|^2\right)\frac{\left(1+\cos^2 2\theta\right)}{\left(\sin\theta \sin 2\theta\right)}$$

Multiplicity

Crystal Structure Factor

Lorentz-Polarization



$$\lambda = 2d_{hkl}\sin\theta$$

Performance of Baseline Models on Simulated Data

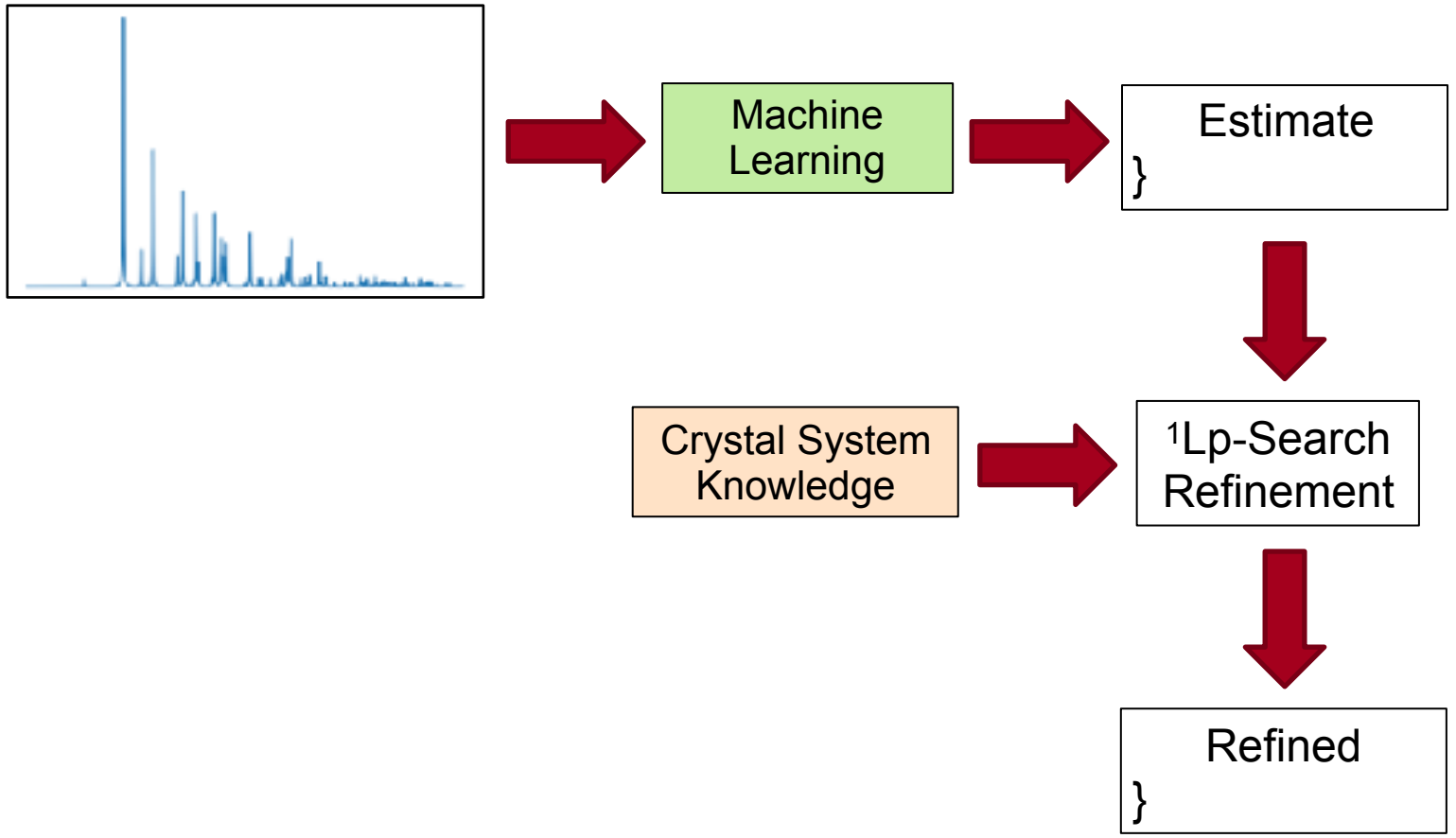
SLAC

ML Models: 1D Convolutional Neural Networks trained and tested on simulated intensity arrays for each crystal system

Crystal System	Independent Parameters	Mean Percentage Error (%)
Cubic		7.4
Hexagonal		6.4
Trigonal		14.8
Tetragonal		11.6
Orthorhombic		10.0
Monoclinic		11.8
Triclinic	,	3.1

Are these predictions good enough?

Machine Learning + Refinement Pipeline



ML can provide the **initial guess** required for refinement

1. Coelho, A. A., Journal of Applied Crystallography (2017).

Monte Carlo Pawley Refinement (Lp Search)

SLAC

Pawley Refinement

- Minimize difference between calculated and observed (experiment) diffractogram
- Method does not rely on the structure factor

$$M = \sum_i W_i \left\{ y_i^{obs} - \frac{1}{c} y_i^{calc} \right\}^2$$

Parameters	Interpretation
Lattice Parameters	Dictates peaks via Bragg Equation
$I(hkl)$	Intensities for each (hkl)
$2\theta_{zero}$	Zero-offset (instrument)
U,V,W	Peak-width
η	Other detector, peak shape parameters

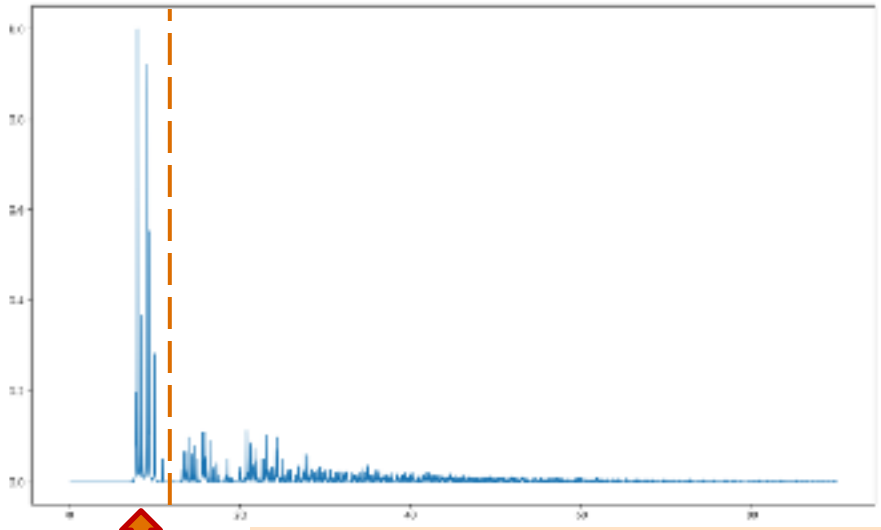
Lp Search is a Monte Carlo minimization of the objective:

- Not used so much due to large parameter range
- Lattice parameter ranges for {}
  - 3 <
  - < 120
- Space Group

Automatic Unit-Cell Refinement

Chemical Formula: $^{1}C_{48}H_{62}Er\ N_7O_2Si_2$ **Crystal System**: Hexagonal

Lattice Parameters: {a = 13.11, b = 13.11, c = 57.64, = 120}



Peaks for c lattice parameter

CNN Estimate

{a = 13.71, b = 13.71, c = 57.74}

After Refinement

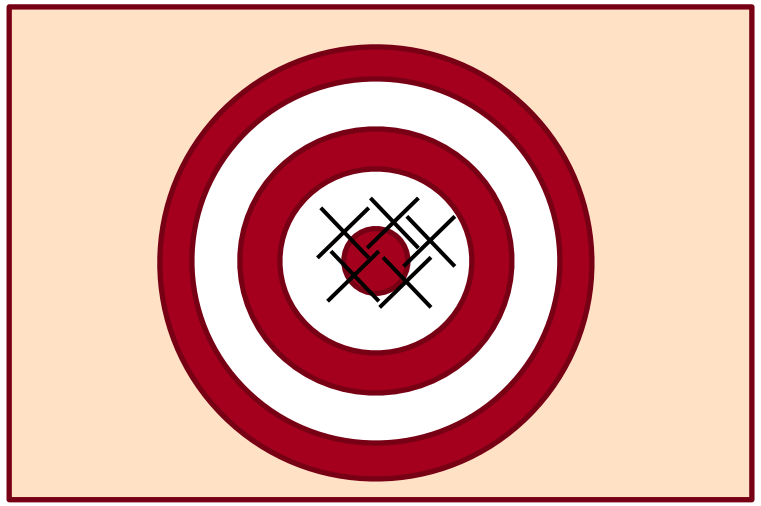
✓ {a = 13.11, b = 13.11, c = 57.64}

Possible to **fully automate** the analysis of challenging powder diffraction patterns

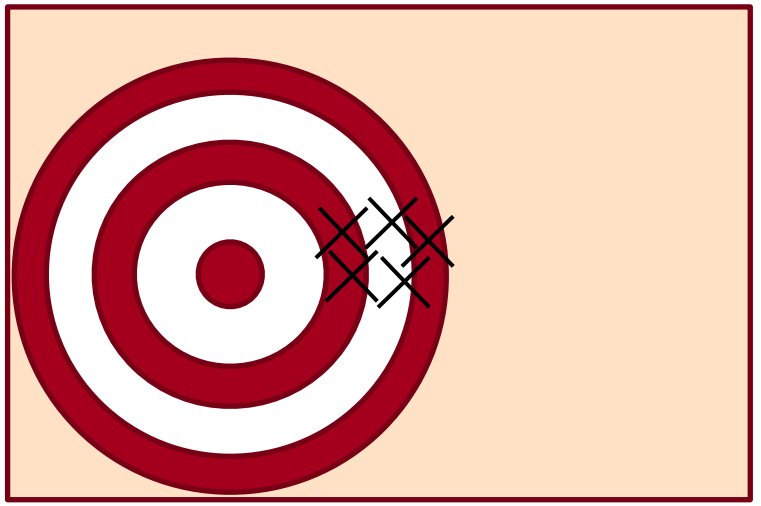
1. Huang, Z. et al., Inorganic chemistry (2018).

Modelling Realistic Non-idealities

- There is normally exists a generalization gap between training on simulated data and predicting on experimental data



Train on Simulated / Test on Simulated



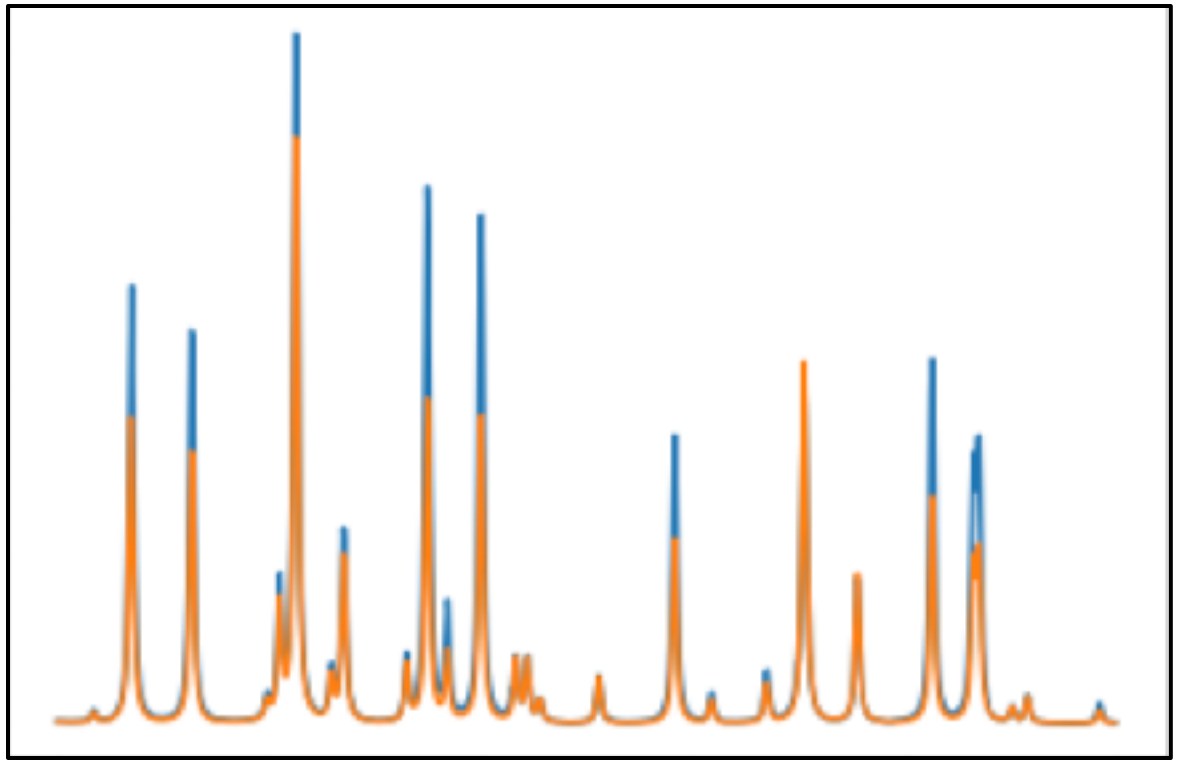
Train on Simulated / Test on Experimental

- Experimental data has extra non-idealities that may need to modelled
  - ➢ Peak Intensity Variation
  - ➢ Zero-Offset Error
  - ➢ Peak Broadening
  - ➢ Impurity Phases
  - ➢ Baseline Noise

Modelling Realistic Non-idealities

SLAC

Intensity Variation

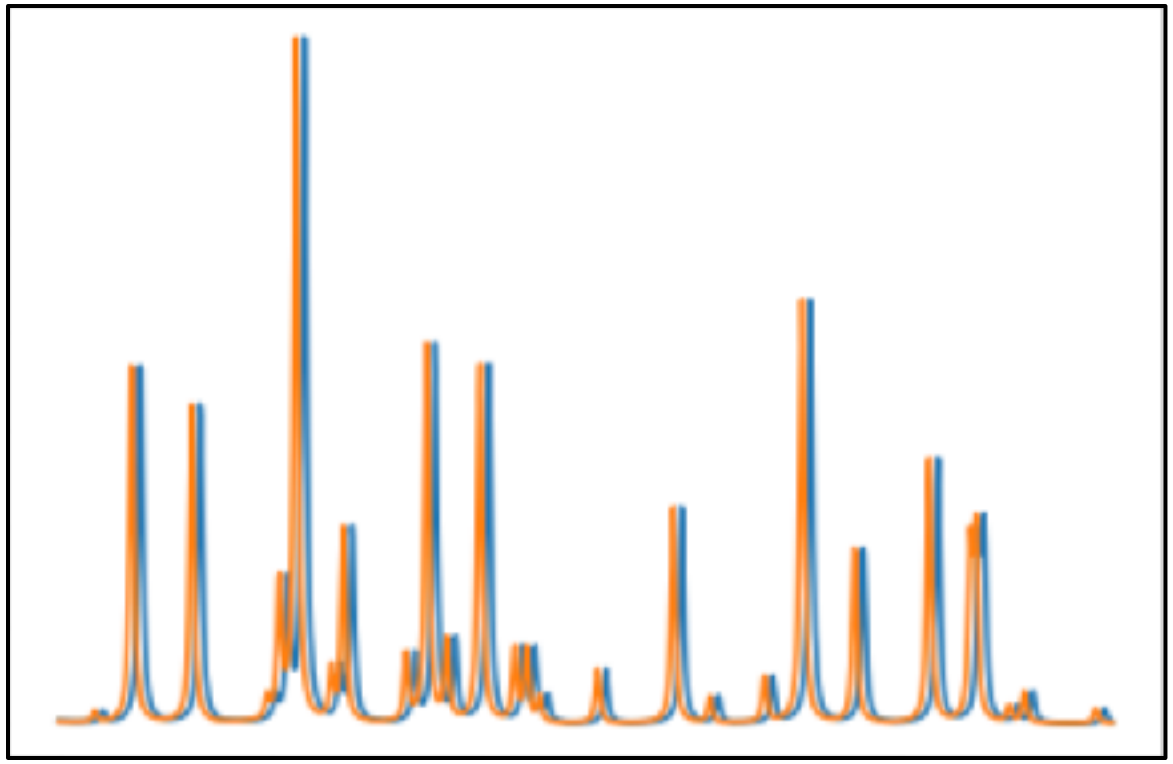


Non-ideality models preferential orientation effects

Modelling Realistic Non-idealities

SLAC

Zero-Offset Error

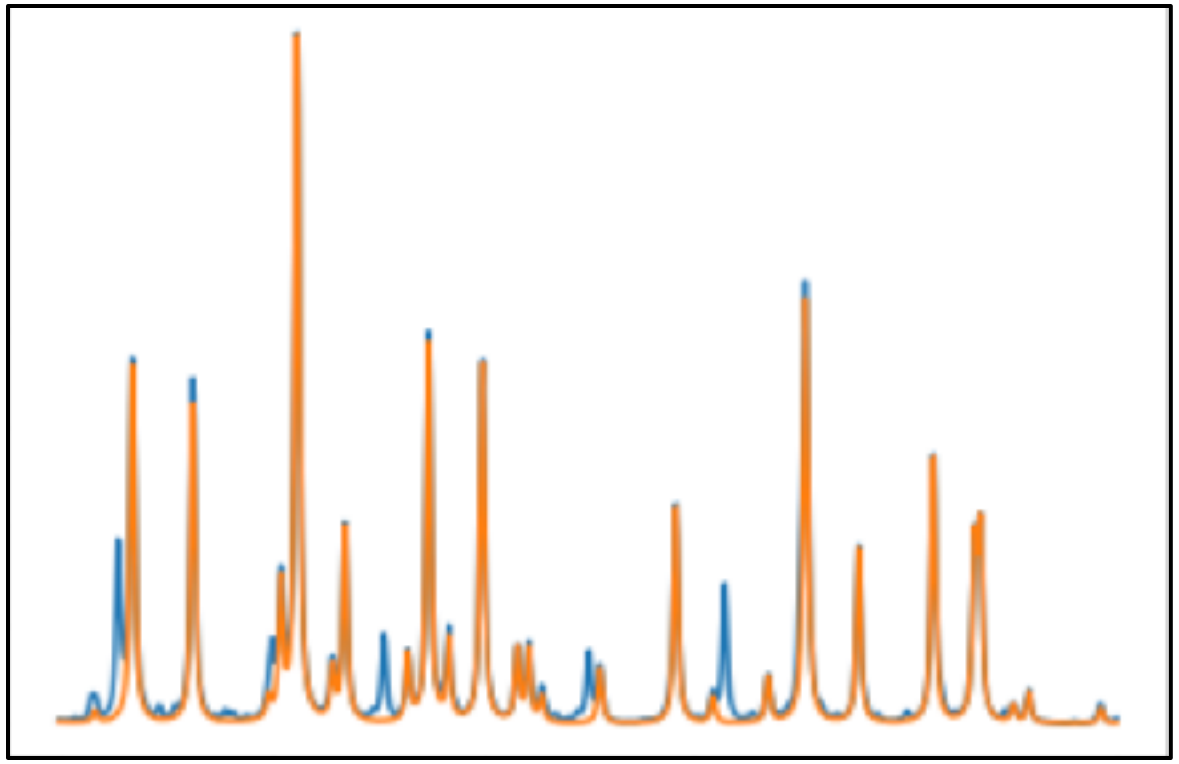


Non-ideality models detector drift

Modelling Realistic Non-idealities

SLAC

Impurity Phases

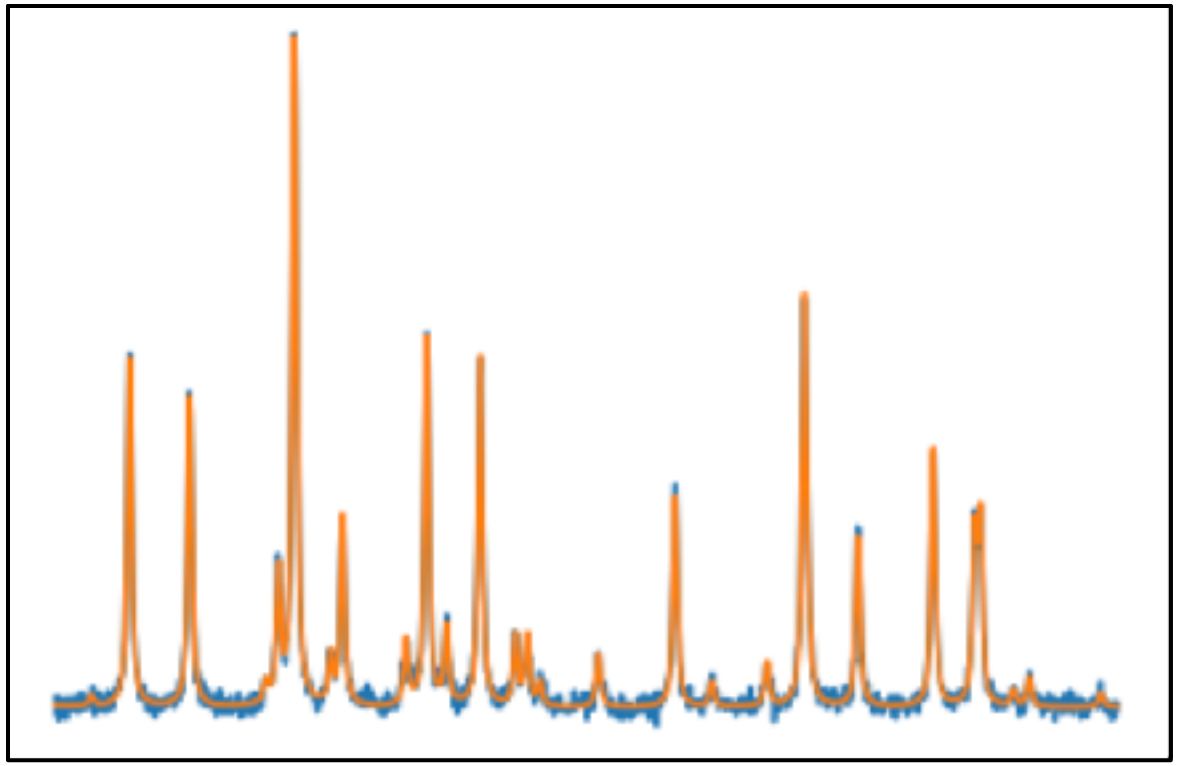


Non-ideality models the addition of extra impurity phases

Modelling Realistic Non-idealities

SLAC

Baseline Noise

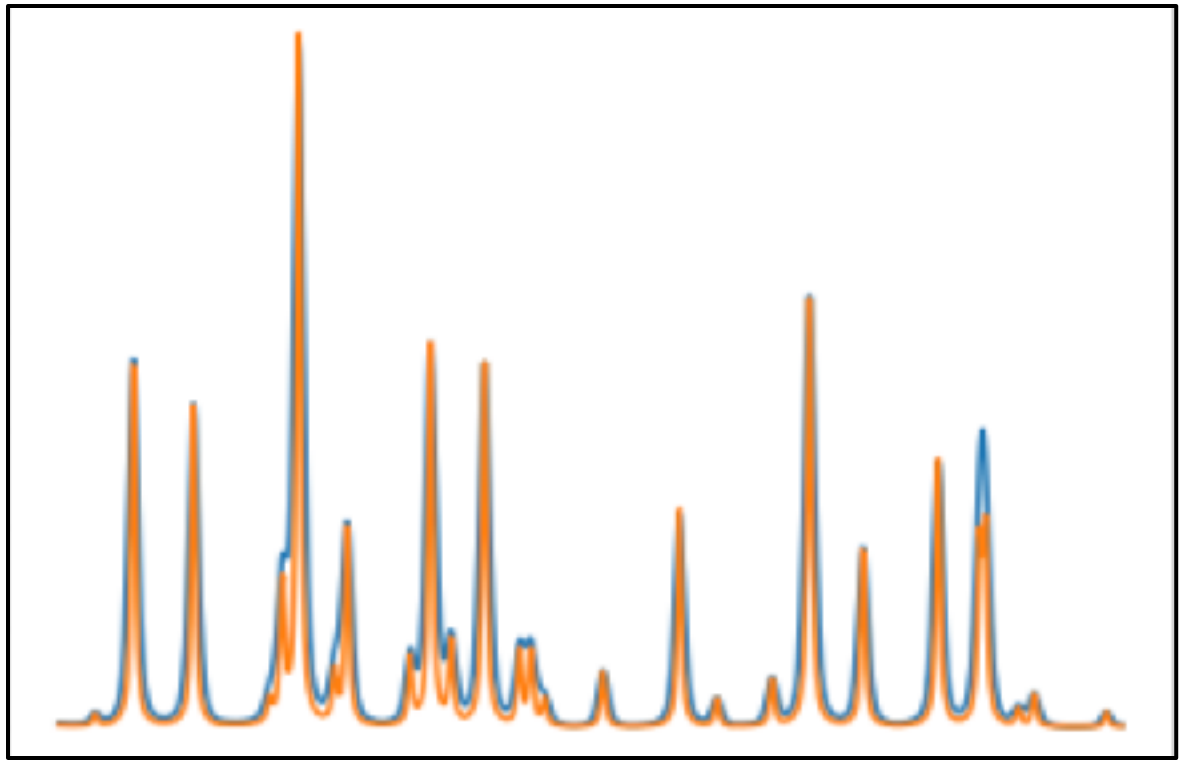


Non-ideality models various sources of baseline noise

Modelling Realistic Non-idealities

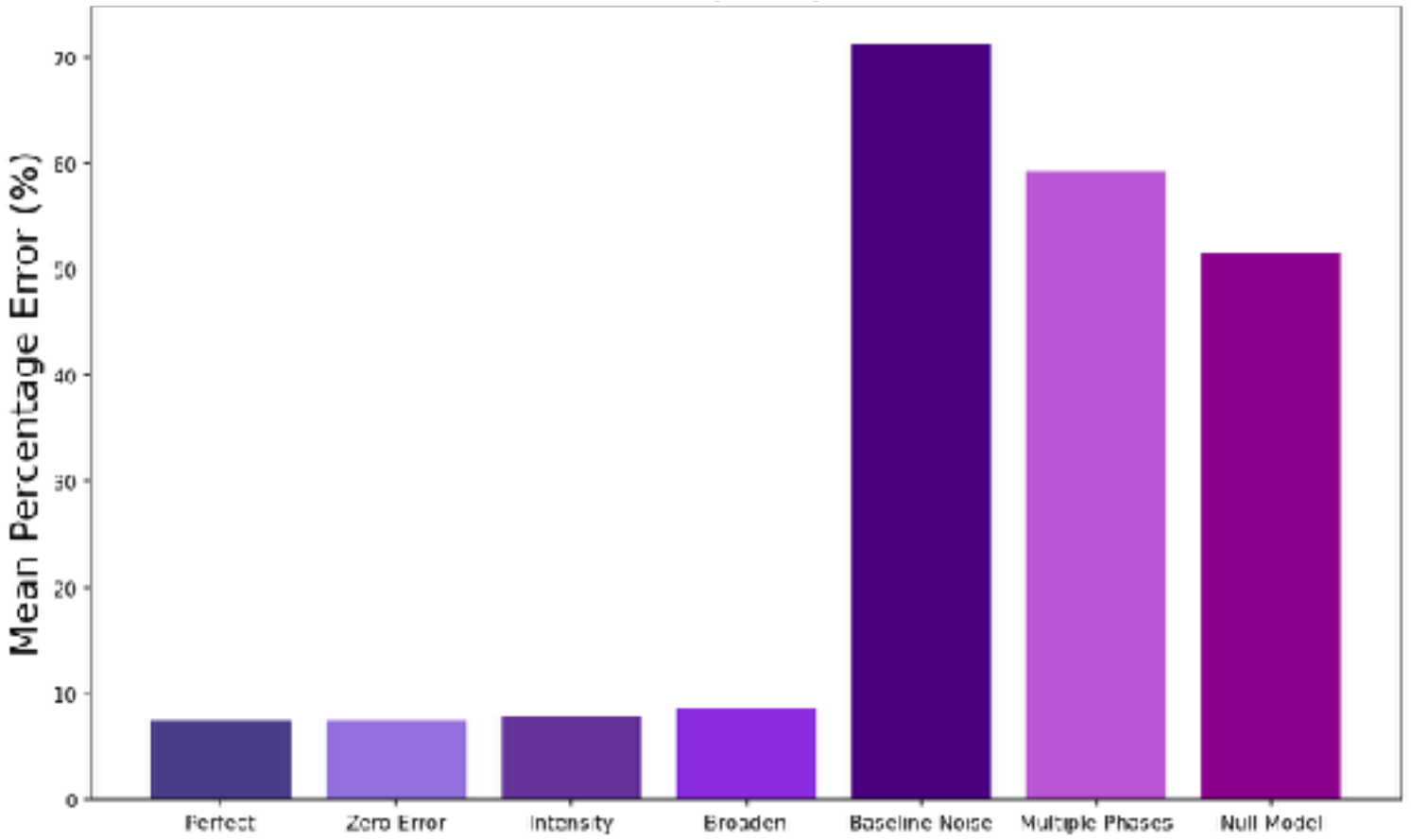
SLAC

Peak Broadening



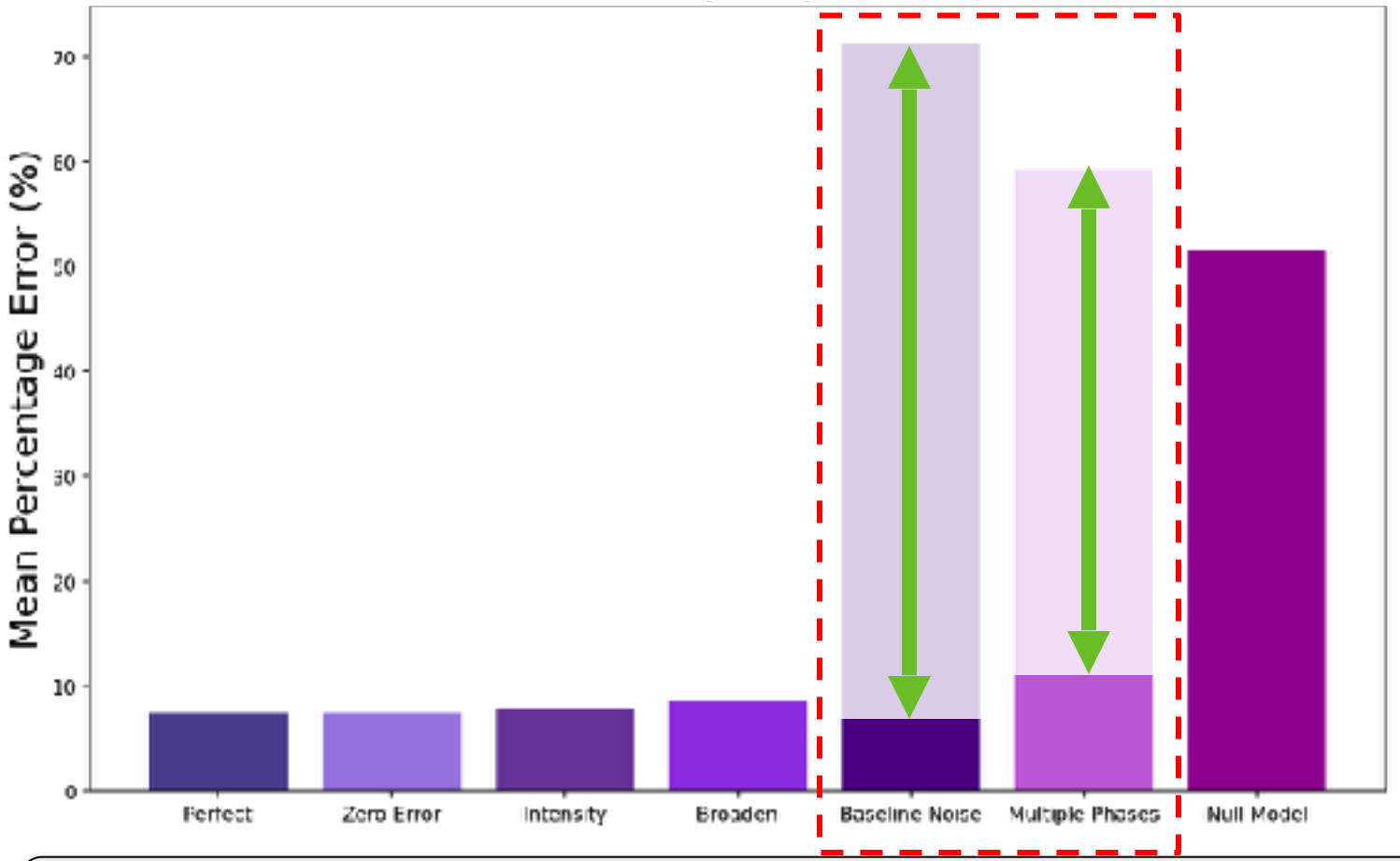
Non-ideality models broadening due to size and microstrain effects

Effect of Realistic Non-idealities



Baseline CNN models are **unstable** against some non-ideal conditions

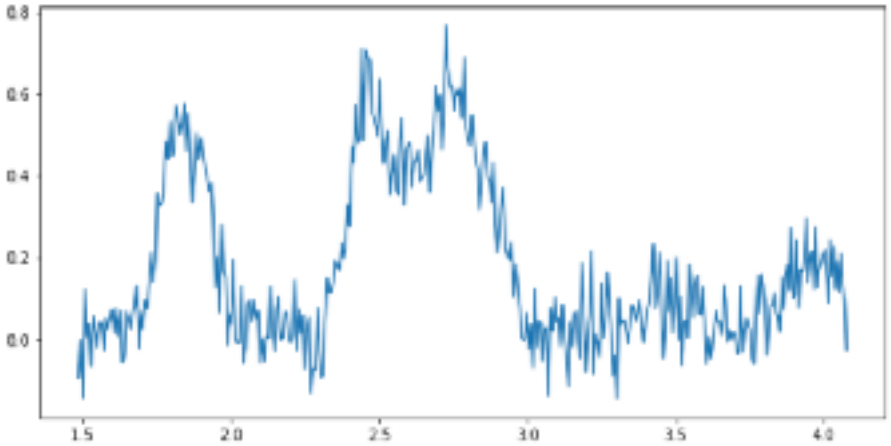
Stabilization against Realistic Non-idealities

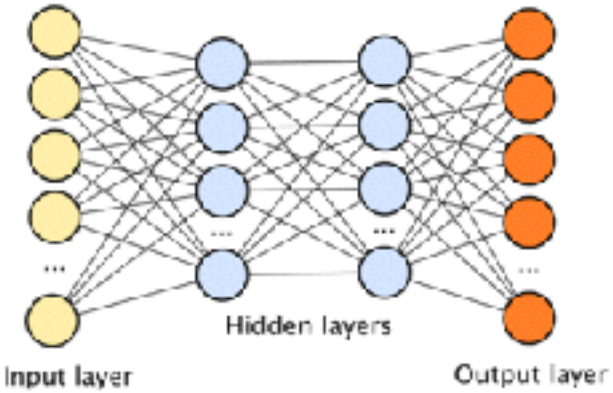


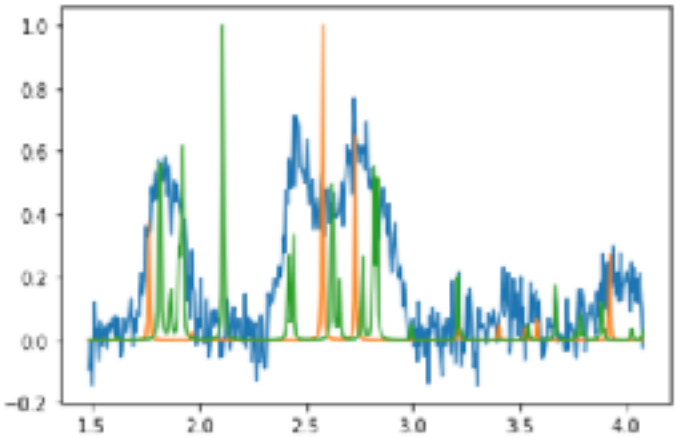
Using non-idealities **during training** greatly improves performance

Multiphase analysis: Predicting phases and phase fractions

Phase identification/fraction using CNNs







['PtSn', 'PtSn4']

+

Prior knowledge of possible phases

Future Work

- **Nanoparticle Optimization**
  - Other systems beyond Pd
  - Extension to multicomponent systems using SAXS/WAXS
  - Analysis of other optimization methods beyond Bayesian optimization

- **Powder Diffraction Anaysis**
  - Incorporation of more realistic physical models for preferred orientation, microstrain, crystallite size and temperature dependence
  - Adding atomic / geometrical features based on plausible compositions
  - Capturing additional information in the 2D area data
  - Using prior information to analyze multiphase data

Acknowledgements

Lenson Pellouchoud

Anthony Fong

Malcolm Davidson

Carena Church

Ekaterina Tcareva

Liheng Wu

Kyle Peterson

Bryce Meredig

Kevin Stone

Daniel Ratner

Richard Walroth

Vivek Thampy

Evan Reed

Mike Dunne

Chris Tassone

U.S. DEPARTMENT OF ENERGY | Energy Efficiency & Renewable Energy

ADVANCED MANUFACTURING OFFICE

SLAC NATIONAL ACCELERATOR LABORATORY

Supplementary Slides

Learning Nanoparticle Design Rules

SLAC

- TOP, oleylamine, oleic acid are ligands which can bind and control nucleation and growth
  - TOP slows nucleation kinetics by increasing Pd precursor
  - Oleylamine can enhance nucleation kinetics
  - ODE is just a diluting solvent

- Higher Temperatures + Lower Flow Rates give bigger NPs
  - Large T accelerates growth kinetics
  - Low Flow-rate increase growth duration