Automated Discovery of Temperature Dependent Structural Change

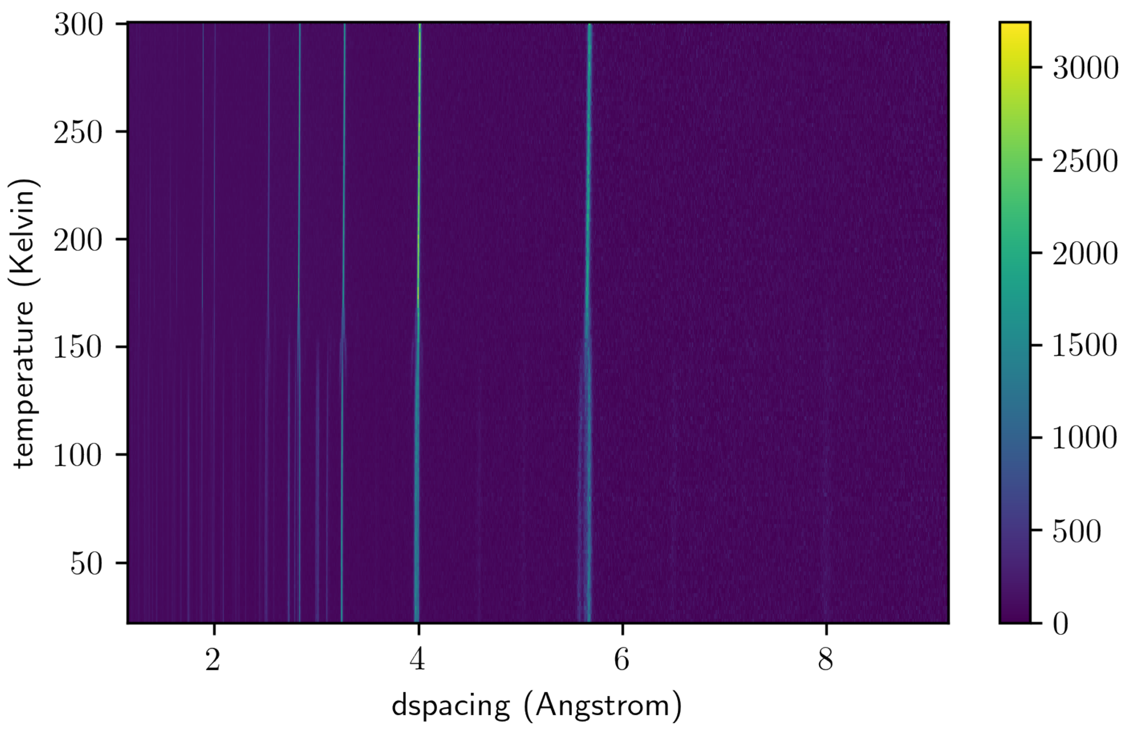
Yaohua Liu[1] and Yawei Hui[2]

1. Quantum Condensed Matter Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, liuyh@ornl.gov
2. Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, huiy@ornl.gov

Jupyter notebook available @ https://github.com/yaohualiu/NPD

Image credit: David Delano

Introduction



2D Time-of-Flight neutron powder diffraction data: I = I (d, T)

- Parametric experiments as a function of temperature.
- No error bars.

Four tasks:

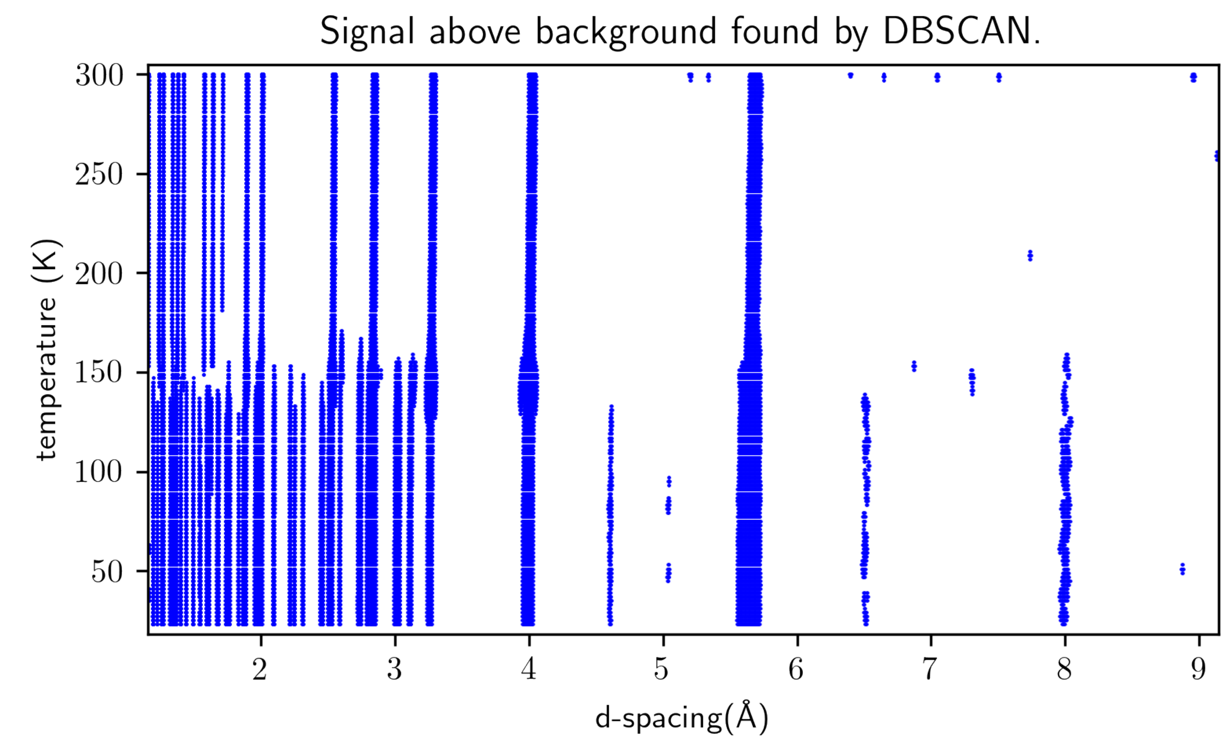
1. Find structural transition temperature(s).
2. Determine signal characteristics b/w 3.22 and 3.32 Å as a function of T.
3. Resolve information for all peaks at a given T.
4. Report a phase transition likelihood b/w two adjacent T's. Peak tables for all T's.

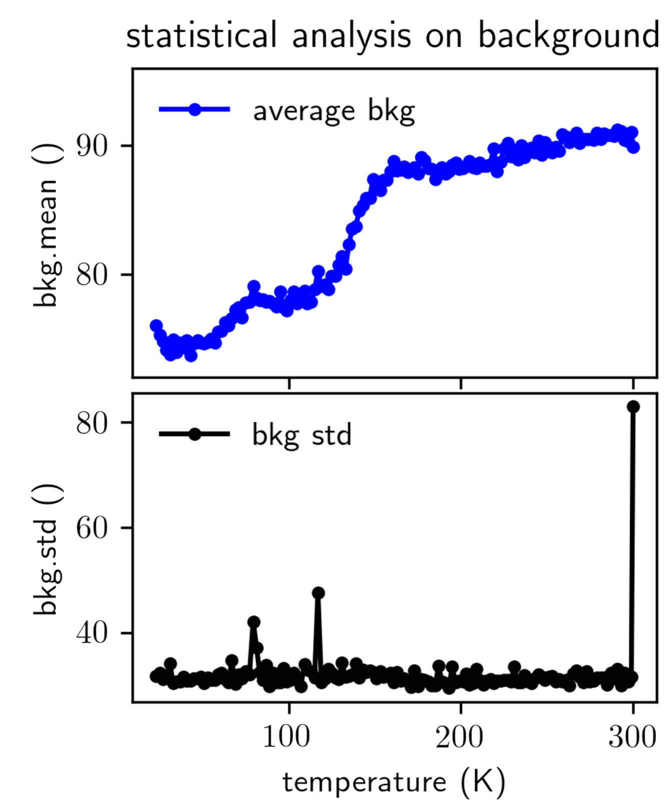
Key philosophy: Emphasizing the global information.

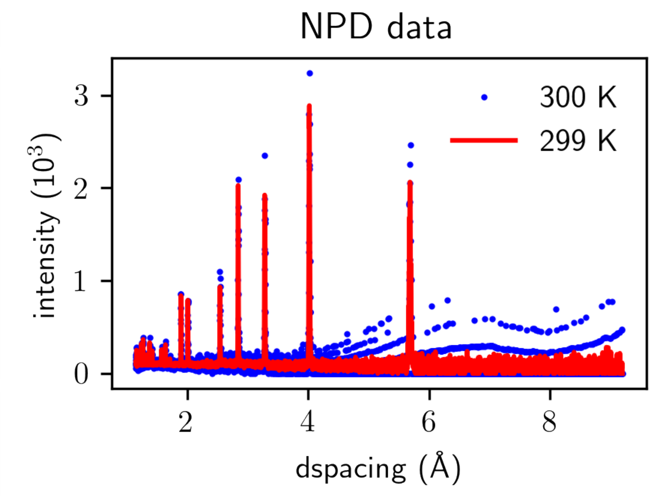
Key algorithms:

- Density-based spatial clustering of applications with noise (DBSCAN).
- Hierarchical peak finding approach using Wavelet Transformation and Le Bail fitting.

T0: Data characteristics





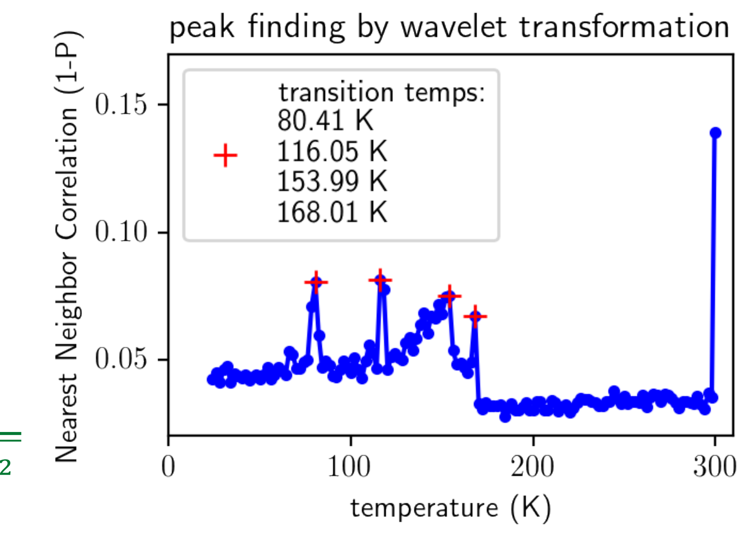


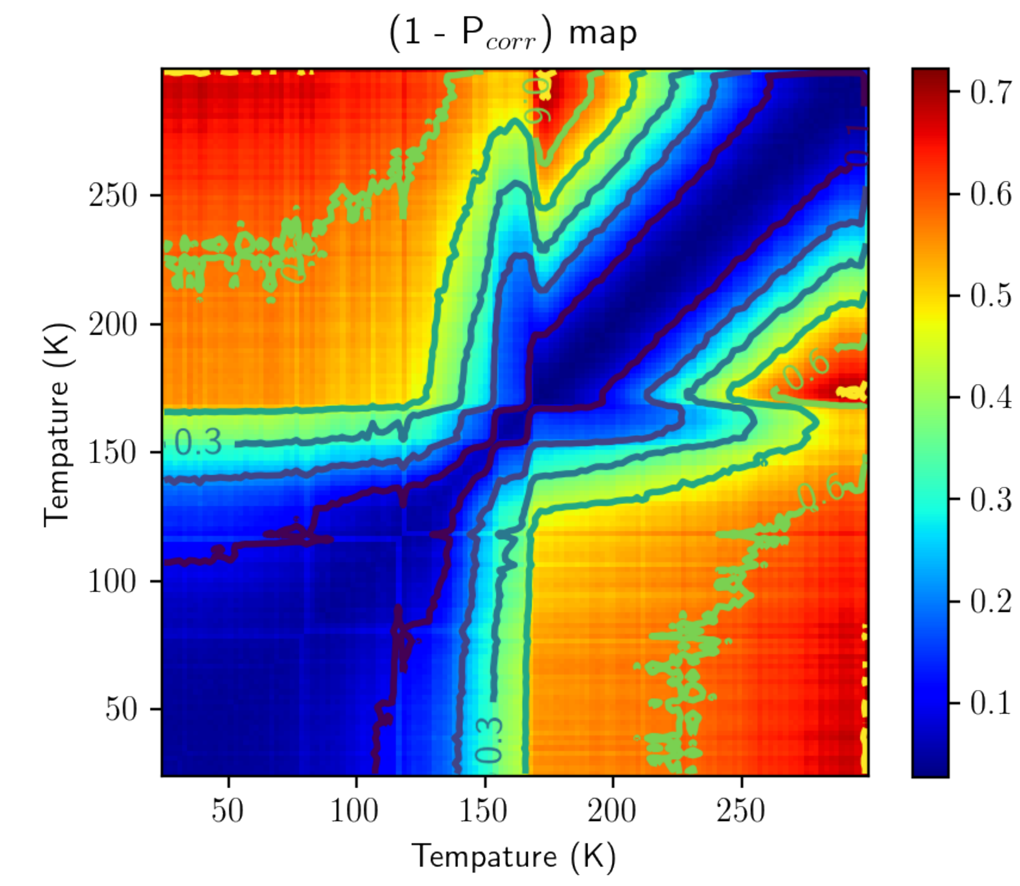
- Statistical analysis on the background found by DBSCAN shows an anomaly at 300 K, due to the low signal/noise ratio.

T1: Structural transition temperature

- Represent NPD data at each temperature is represented as a vector $x = (x_1, x_2, \dots x_n)$.
- Calculate Pearson correlation coefficient b/w two temperatures,

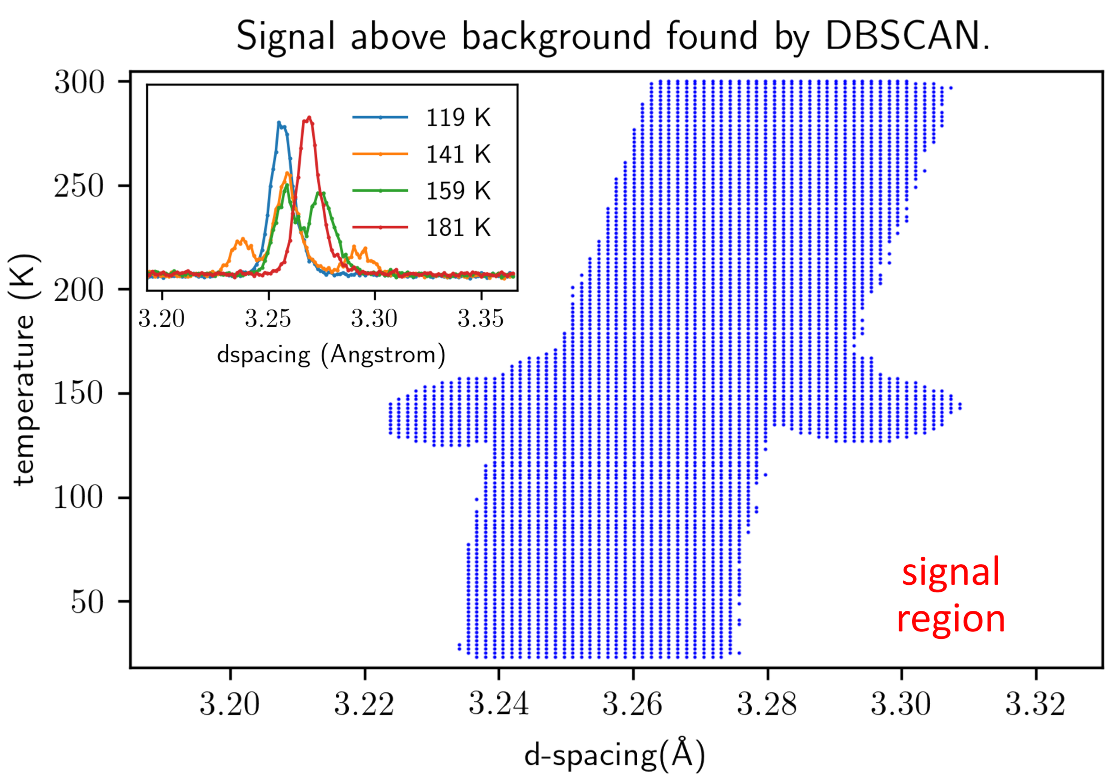
$$p = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

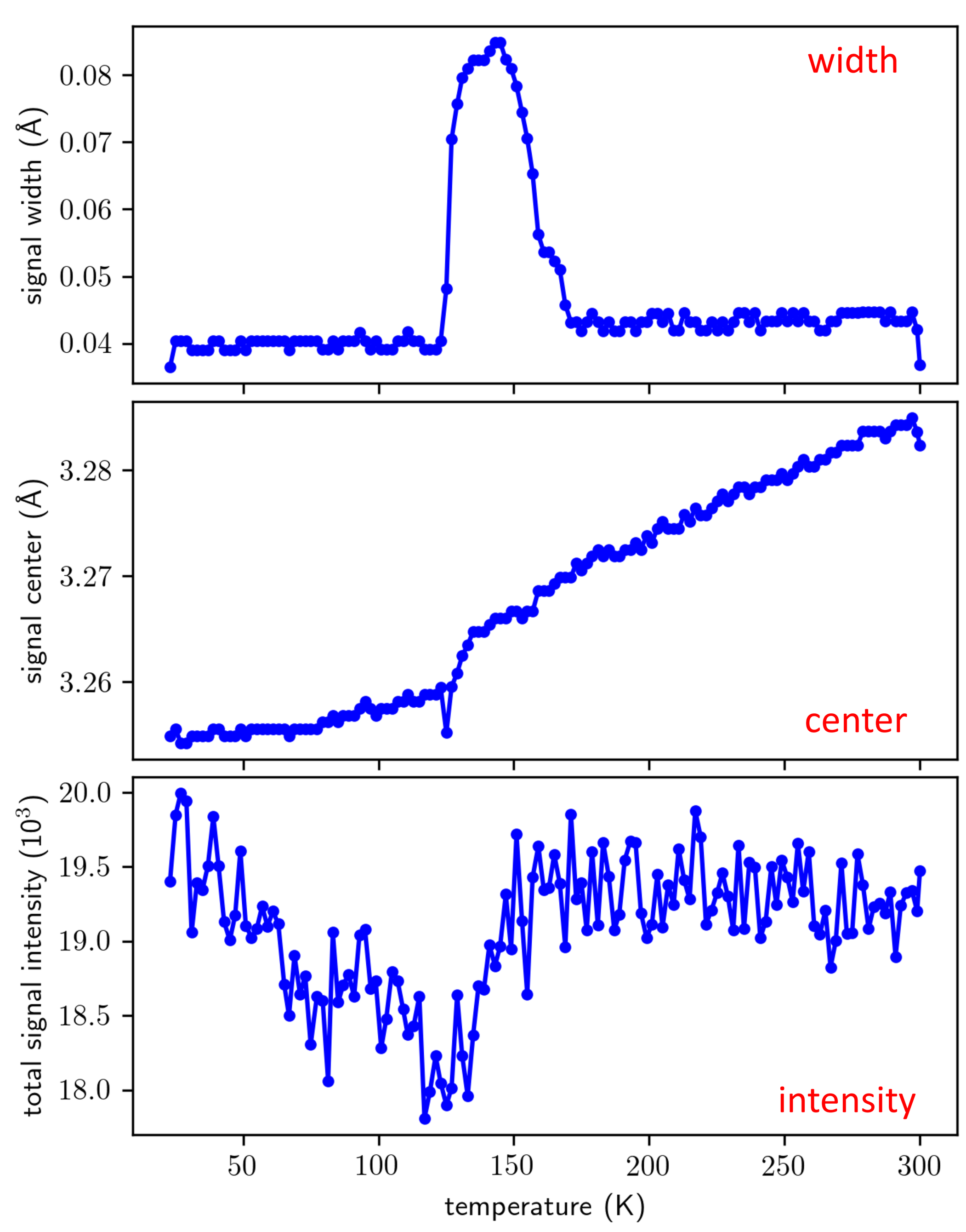




Visualization: abrupt color changes at potential phase transition temperatures.

T2: Signal between 3.22 and 3.32 Å as a function of T

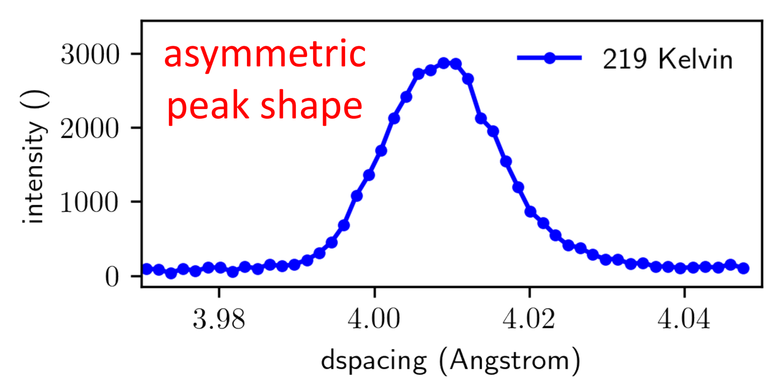




DBSCAN is used for finding the signal region.

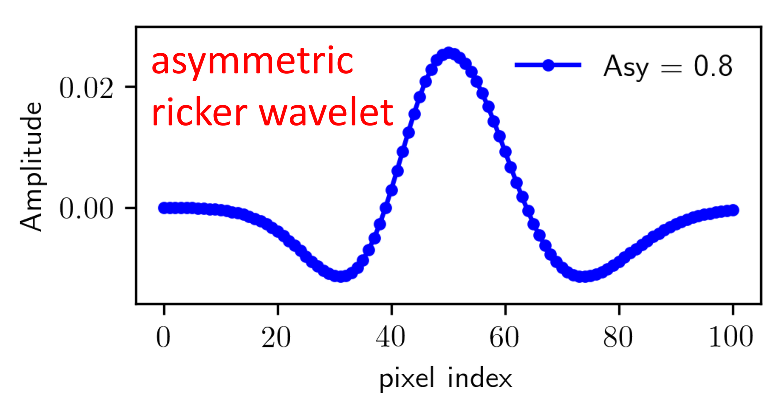
1. Identify the background region and signal above background region via DBSCAN.
2. Determine the signal center and width directly from the signal region.
3. Parameterize the background level using a polynomial function of temperature and d-spacing.
4. Remove the background from the raw data to calculate the signal intensity.

T3: Intensity, center & width for all peaks at a given T

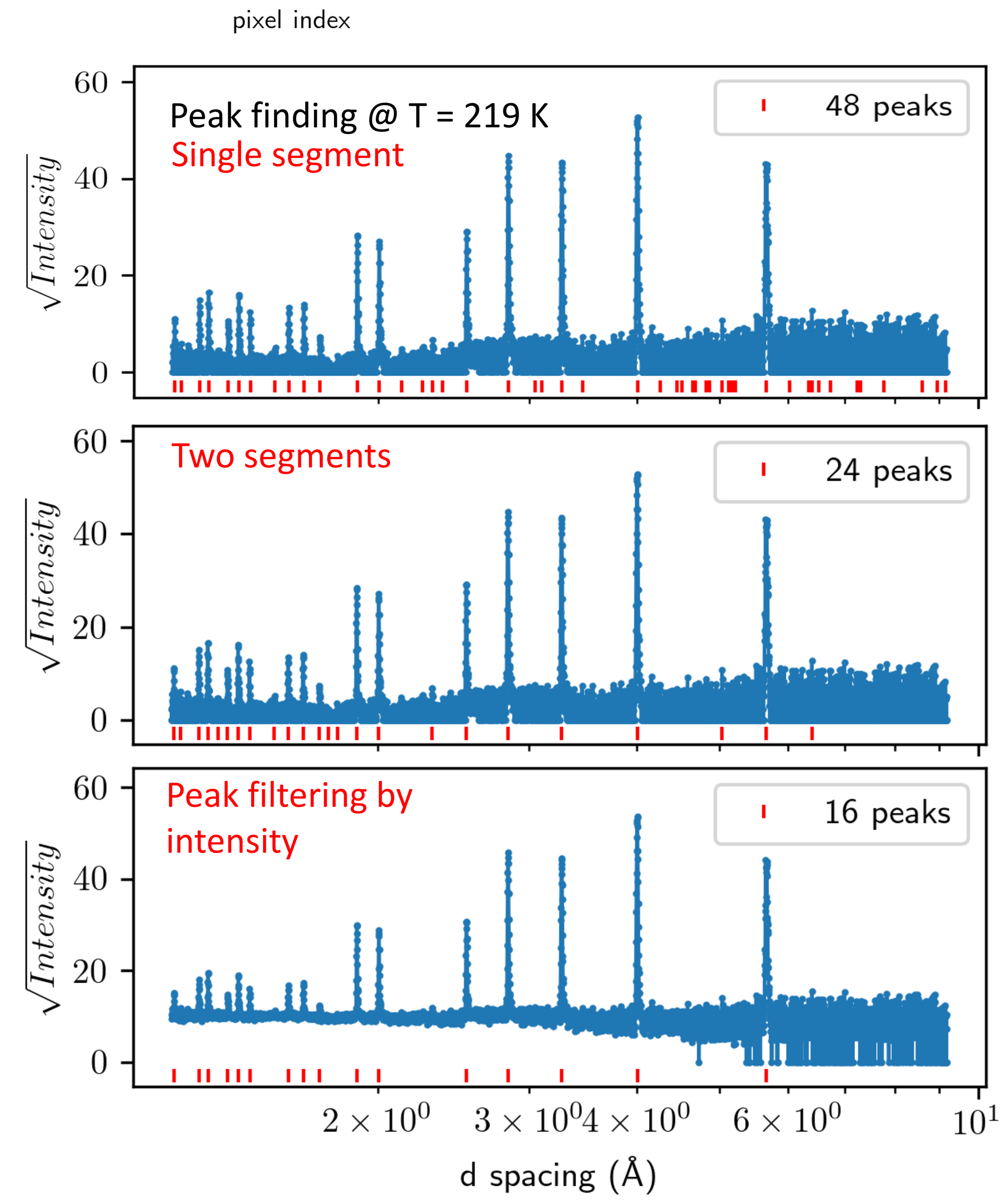


Using asymmetric wavelet and multiple-segment algorithm improves the peak prediction accuracy.

Asymmetric Bragg peak shape.
✓ Using an asymmetric ricker wavelet profile



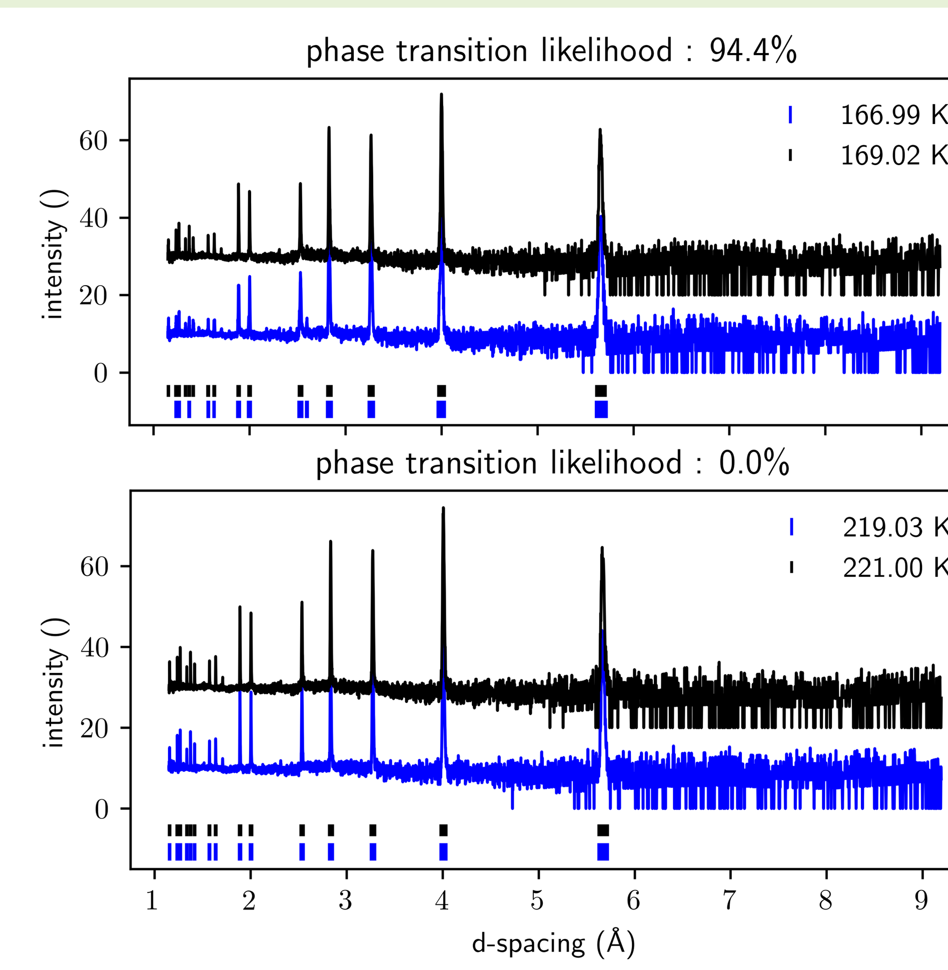
Low noise level at low-d and high noise level at high-d.
✓ Two-segment peak finding algorithm with different criteria



Le Bail fitting is used to generate the peak table.

1. Background is parameterized, bkg = bkg (d, T), and then removed.
2. Peak width is parameterized as a polynomial function of d-spacing.
3. Peak positions from the wavelet transformation are tightly constrained.
4. Peak amplitudes are loosely constrained.
5. Low intensity peaks (< 1.5 * bkg.std(T)) are removed from the initial peak list.

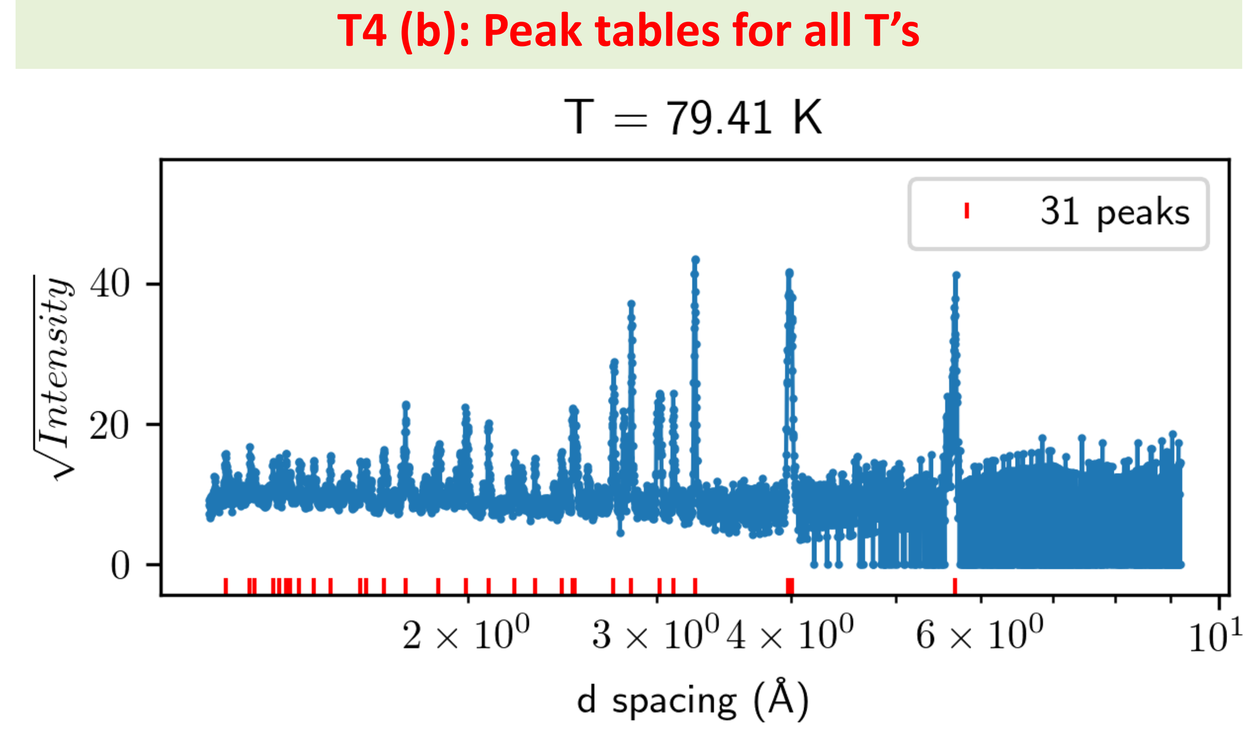
T4 (a): Phase transition likelihood b/w two adjacent T's

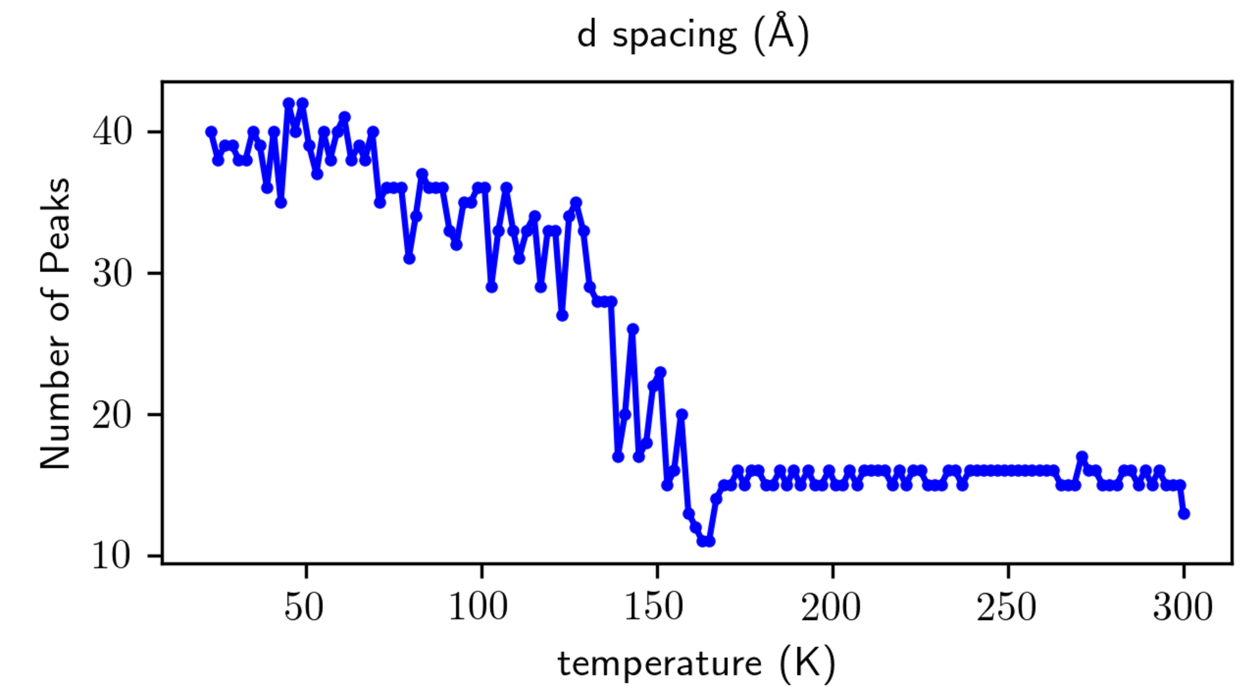


If limited information was available, e.g. only two datasets:

1. Find regions of signal above background via DBSCAN.
2. Calculate shared and unshared signal bins b/w two temperatures.
   - More shared signal bin ➔ lower likelihood
   - More unshared signal bin ➔ higher likelihood

T4 (b): Peak tables for all T's





Oak Ridge National Laboratory

Work is supported by Division of Scientific User Facilities of the Office of Basic Energy Sciences, US Department of Energy.

U.S. Department of Energy | Office of Science