⊕ Unsupervised Learning of X-ray

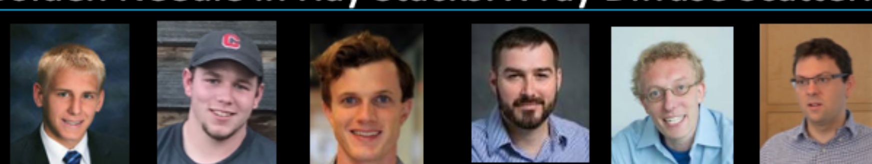
Lect 3.

Unsupervised ML of Diffuse Scattering (X-ray)

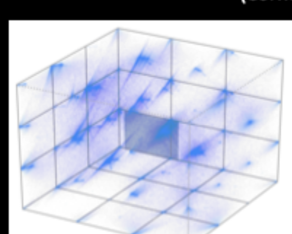
(T, p, δ)

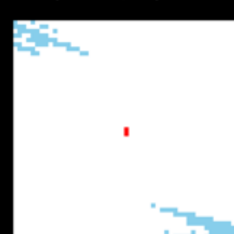
(T, p, δ)

Golden Needle in Hay Stacks: X-ray Diffuse Scattering



Jordan Venderley | Mike Matty | Geoff Pleiss (Cornell, CS) | Jacob Ruff (CHESS) | K. Winberger (Cornell, CS) | A. Wilson (Cornell, ORIE)



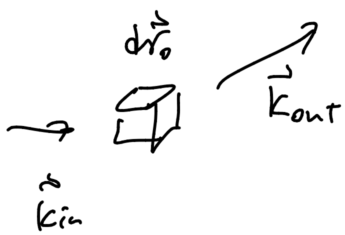


□ Diffraction Measurement = Density-Density Corr.

$$\Psi_{out}(\vec{r}) \propto A e^{i\vec{k}_{in}\cdot\vec{r}_0}\,\rho(\vec{r}_0)\, e^{i\vec{k}_{out}\cdot(\vec{r}-\vec{r}_0)}$$

$\Psi_{in}(\vec{r}_0)$

$= A e^{i\vec{k}_{in}\cdot\vec{r}_0}$



$$I(\vec{q}) = |\Psi_{out}(\vec{r})|^2 \propto A^2 \int d\vec{r}_0 \int d\vec{R}\, e^{i\vec{q}\cdot\vec{R}}\, \rho(\vec{r}_0)\,\rho(\vec{r}_0-\vec{R})$$

$\vec{q} = \vec{k}_{out} - \vec{k}_{in}$

If $\exists$ density wave: $\rho(\vec{r};T) = \rho_0 + \mathrm{Re}\left[\Delta_{\vec{Q}}(T)\, e^{i\vec{Q}\cdot\vec{r}}\right]$

$$I(\vec{Q};T) \propto |\Delta_{\vec{Q}}(T)|^2$$

What does diffraction measure?

- $I(\vec{q};T) \propto |\widetilde{\rho}(\vec{q};T)|^2$: Fourier amplitude of density
- Density Wave with wave vector $\vec{Q}$

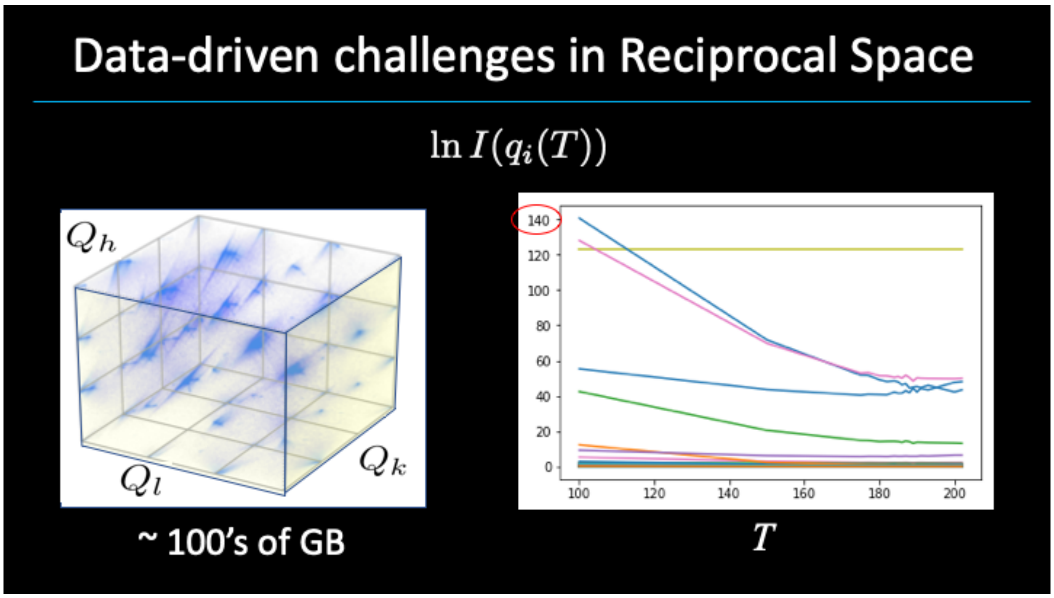
$$\rho(\vec{r};T) = \rho_0 + \mathrm{Re}\left[\Delta_{\vec{Q}}(T)\, e^{i\vec{Q}\cdot\vec{r}}\right]$$

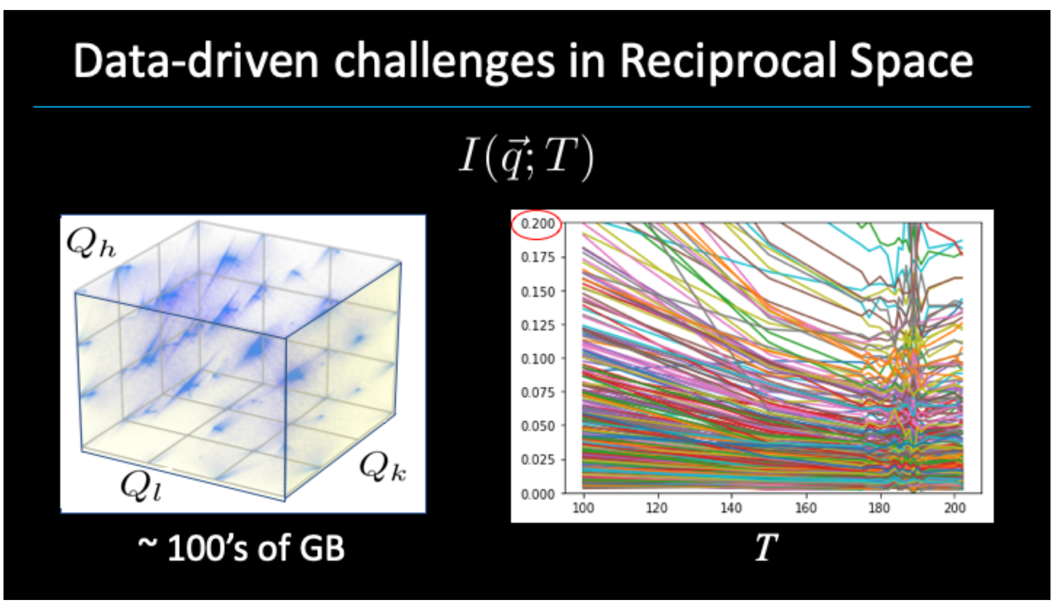
- $I(\vec{Q};T) \propto |\Delta_{\vec{Q}}(T)|^2$: the Order Parameter

✱ Challenge of Dealing with Real Data

: Noise and high intensity background

(atomic positions giving Bragg Peak)





Mission:

Months of inspection by eye $\Rightarrow$ Discover Models $\mu_k^*(T)$'s & thereby Cluster.

★: Demo: Gaussian Mixture Model for Clustering
⇒ simple code building on Scikit Learn.

https://github.com/iamc/ML-CM-2019/blob/master/Eun-Ah_Kim/Summer_School_GMMs_and_GPs.ipynb

① First half of the notebook demonstrates how to use the GMM function setup in Scikit learn to cluster using Gaussian Mixture Model.

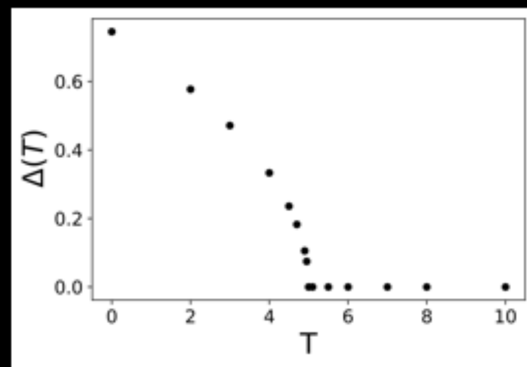
② Second half of the notebook Implements a Gaussian process which is a form of generative modeling literally following the theory in following pages.

✱ Gaussian Process Regression

A Traditional Generative Model

- Training Data

$$\mathbb{T} \equiv \{T_1, T_2, \cdots, T_m\}$$
$$\mathbf{\Delta} \equiv \{\Delta_1, \Delta_2, \cdots, \Delta_m\}$$



- Prediction given testing points $\mathbb{T}^* \equiv \{T_1^*, T_2^*, \cdots, T_n^*\}$

$$\mathbf{\Delta}^* =$$

A Traditional Generative Model

- Ginzburg-Landau Theory: $\Delta_{\vec{Q}}(T;\{r,s\}) = \sqrt{\dfrac{r(T-T_c)}{s}}$
- Training Data

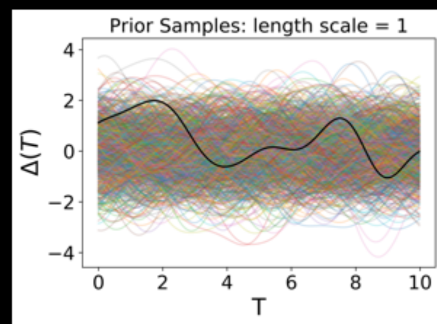
$$\mathbb{T} \equiv \{T_1, T_2, \cdots, T_m\}$$
$$\mathbf{\Delta} \equiv \{\Delta_1, \Delta_2, \cdots, \Delta_m\} \xrightarrow{\text{Regression}} \{r^*, s^*\}$$

- Prediction given testing points $\mathbb{T}^* \equiv \{T_1^*, T_2^*, \cdots, T_n^*\}$

$$\mathbf{\Delta}^* = \{\Delta(T_i^*; r^*, s^*)\}_{i=1,\cdots,n}$$

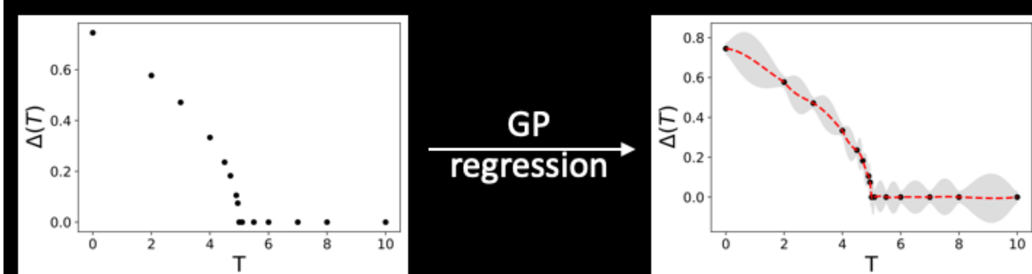
Generative Model: Guassian Process

- ~~Fit for the function~~
- Gaussian Probability Distribution over a Function Space



GP regression on

$$\mathbb{T} \equiv \{T_1, T_2, \cdots, T_m\}$$
$$\mathbf{\Delta} \equiv \{\Delta_1, \Delta_2, \cdots, \Delta_m\}$$



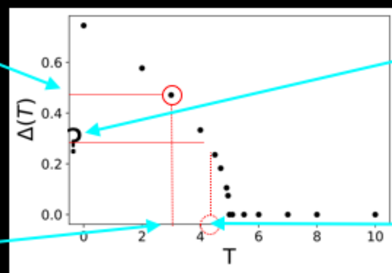
Goal: Find the conditional probability

Train Data

$$p(\mathbf{y}^*|\mathbf{y}, \mathbf{X}, \mathbf{X}^*) = ?$$

Test Data

$$\mathbf{y} = \begin{bmatrix} \Delta_1 \\ \vdots \\ \Delta_m \end{bmatrix} \qquad \mathbf{y}^* = \begin{bmatrix} \Delta(T_1^*) \\ \vdots \\ \Delta(T_{m'}^*) \end{bmatrix}$$



$$\mathbf{X} = \begin{bmatrix} T_1 \\ \vdots \\ T_m \end{bmatrix} \qquad \mathbf{X}^* = \begin{bmatrix} T_1^* \\ \vdots \\ T_{m'}^* \end{bmatrix}$$

Gaussian Process

- Assume

Train data

$$p(\mathbf{y}, \mathbf{y}^*|\mathbf{X}, \mathbf{X}^*) = \underbrace{\mathcal{N}\left(0, \begin{bmatrix} K(\mathbf{X},\mathbf{X}) + \sigma^2 I & K(\mathbf{X}^*,\mathbf{X}) \\ K(\mathbf{X},\mathbf{X}^*) & K(\mathbf{X}^*,\mathbf{X}^*) + \sigma^2 I \end{bmatrix}\right)}_{\text{Gaussian Distribution}}$$

Test data | Kernel (correlation between inputs) | Noise

- Integrate out y

$$p(\mathbf{y}^*|\mathbf{y}, \mathbf{X}, \mathbf{X}^*) = ?$$

Gaussian Process

- Assume

Train data

$$p(\mathbf{y}, \mathbf{y}^*|\mathbf{X}, \mathbf{X}^*) = \underbrace{\mathcal{N}\left(0, \begin{bmatrix} K(\mathbf{X},\mathbf{X}) + \sigma^2 I & K(\mathbf{X}^*,\mathbf{X}) \\ K(\mathbf{X},\mathbf{X}^*) & K(\mathbf{X}^*,\mathbf{X}^*) + \sigma^2 I \end{bmatrix}\right)}_{\text{Gaussian Distribution}}$$

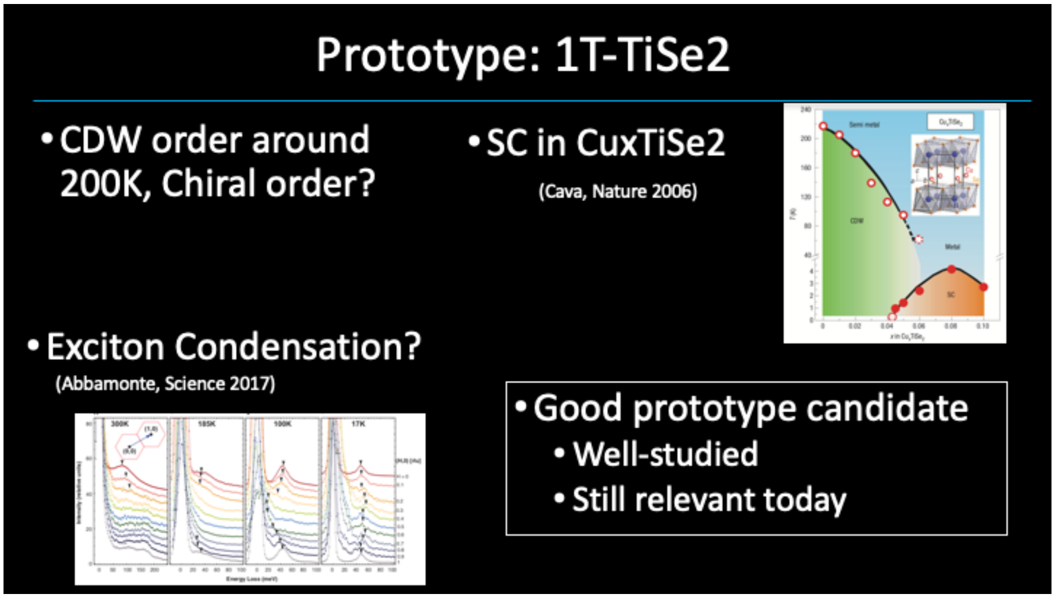
Test data | Kernel (correlation between inputs) | Noise

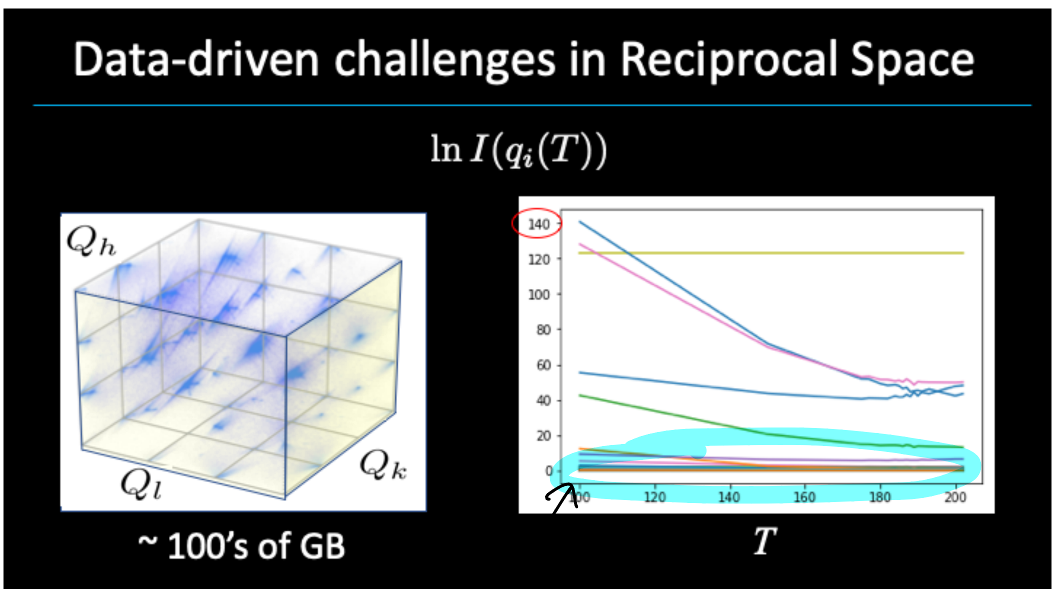
- Integrate out y $\boxed{p(\mathbf{y}^*|\mathbf{y}, \mathbf{X}, \mathbf{X}^*) = \mathcal{N}(\mu^*, \Sigma^*)}$

$$\mu^* = K(\mathbf{X}^*,\mathbf{X})(K(\mathbf{X},\mathbf{X}) + \sigma^2 I)^{-1}\mathbf{y}$$

$$\Sigma^* = K(\mathbf{X}^*,\mathbf{X}^*) + \sigma^2 I - K(\mathbf{X}^*,\mathbf{X})(K(\mathbf{X},\mathbf{X}) + \sigma^2 I)^{-1}K(\mathbf{X},\mathbf{X}^*)$$

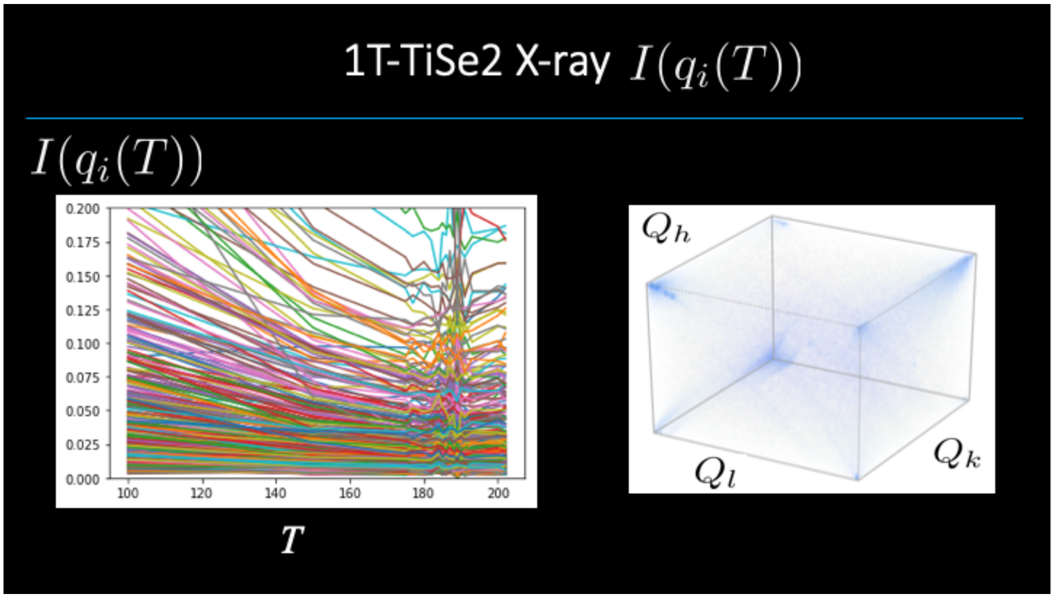
* Using Gaussian Mixture Model & GP mixture model to "discover" CDW peak locations in X-ray data.





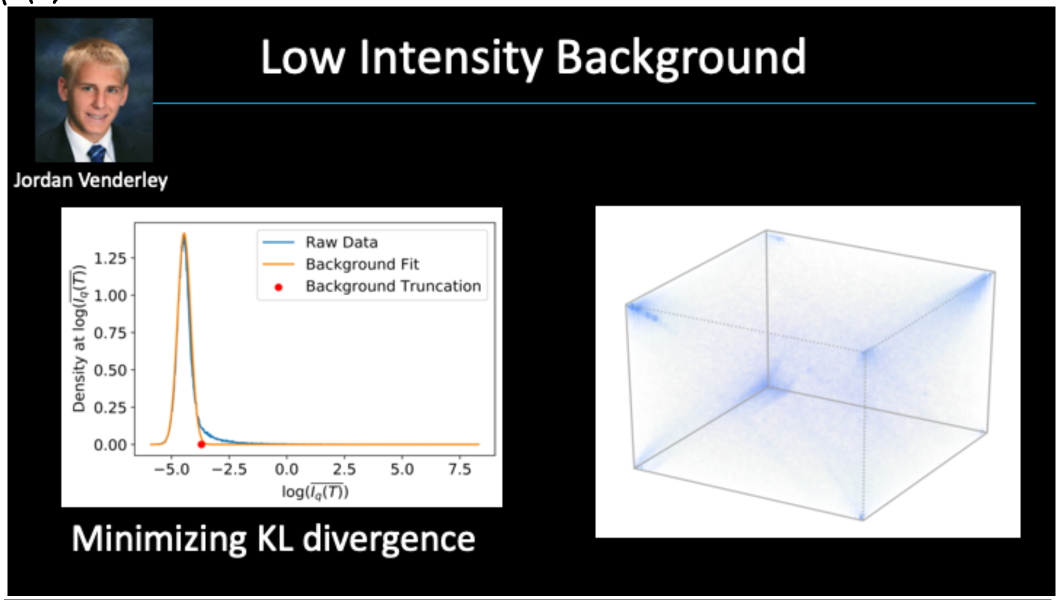
What matters for collective behavior is down in intensity by $10^5 \sim 10^6$

Just thresh-holding low intensity is not sufficient

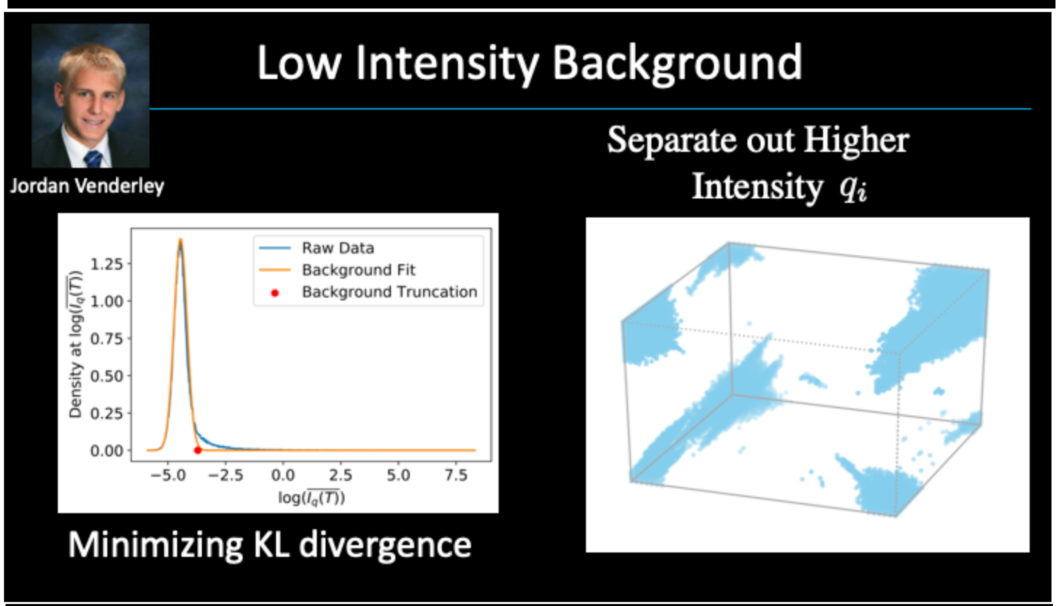


Demo: The full preprocessing, separating Bragg peak from the rest and clustering in action

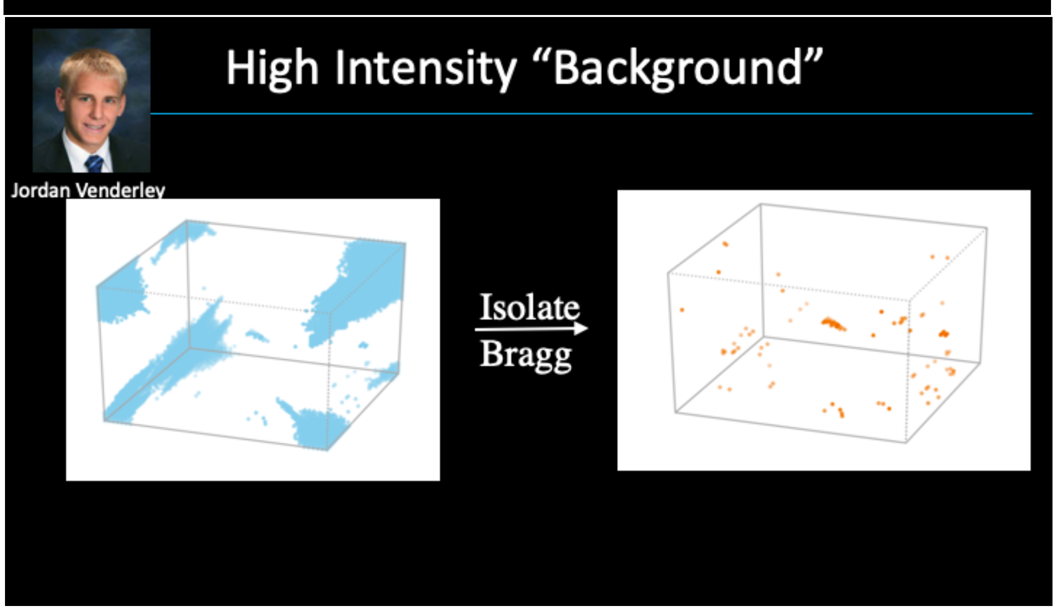
* Results.



Low Intensity Background
Jordan Venderley
Raw Data
Background Fit
Background Truncation
Minimizing KL divergence

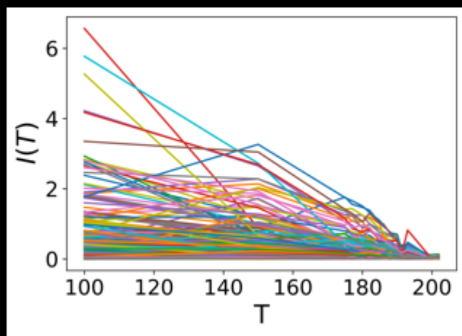


Low Intensity Background
Jordan Venderley
Separate out Higher Intensity q_i
Raw Data
Background Fit
Background Truncation
Minimizing KL divergence



High Intensity "Background"
Jordan Venderley
Isolate Bragg

Gaussian Mixture Model Clustering



- Hyperparameters to be learned:

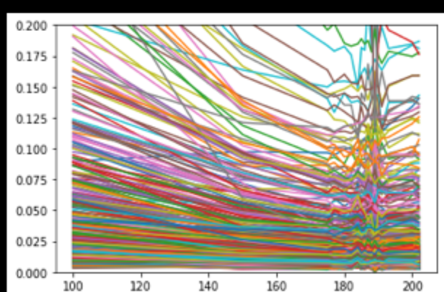
$$\{(\pi_1, \mu_1, \Sigma_1), \cdots, (\pi_K, \mu_K, \Sigma_K), \}$$

- Maximize the log-likelihood

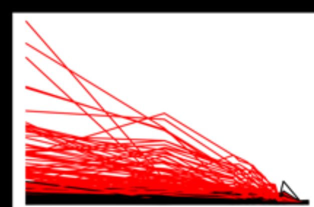
$$\log p(\mathbf{I}|\mu, \Sigma, \pi) = \sum_{\mathbf{q}} \log \left[\sum_{k=1}^{K} \pi_k \mathcal{N}(\vec{I}_{\mathbf{q}}; \mu_k, \Sigma_k) \right]$$

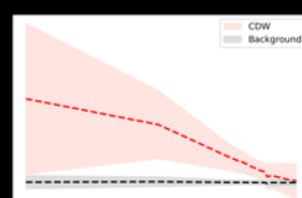
Results

$$\left[\sum_{k=1}^{K} \pi_k \mathcal{N}(\vec{I}_{\mathbf{q}}; \mu_k, \Sigma_k) \right]$$

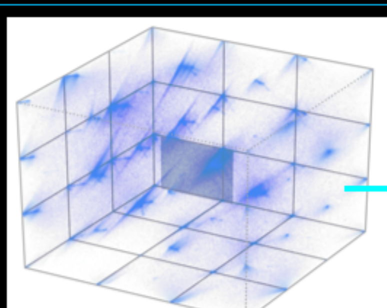


Clustering

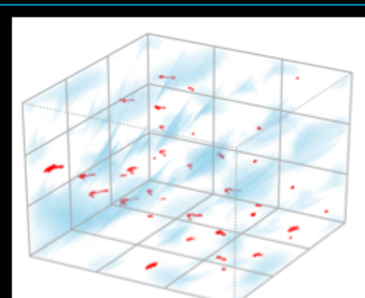




Results



TiSe$_2$ data (150K)



Charge ordering:
(π, π, π) $(\pi, 0, \pi)$