

1 **Harnessing Interpretable and Unsupervised Machine Learning to Address Big Data**  
2 **from Modern X-ray Diffraction**

3 Jordan Venderley<sup>1</sup>, Michael Matty<sup>1</sup>, Krishnanand Mallayya<sup>1</sup>, Matthew Krogstad<sup>2</sup>, Jacob Ruff<sup>3</sup>,  
4 Geoff Pleiss<sup>4</sup>, Varsha Kishore<sup>4</sup>, David Mandrus<sup>5</sup>, Daniel Phelan<sup>2</sup>,  
5 Lekh Poudel<sup>6,7</sup>, Andrew Gordon Wilson<sup>8</sup>, Kilian Weinberger<sup>4</sup>,  
6 Puspa Upreti<sup>2,9</sup>, M. R. Norman<sup>2</sup>, Stephan Rosenkranz<sup>2</sup>, Raymond Osborn<sup>2</sup>, Eun-Ah Kim<sup>1\*</sup>

7 <sup>1</sup>Department of Physics, Cornell University

8 <sup>2</sup>Materials Science Division, Argonne National Laboratory

9 <sup>3</sup>Cornell High Energy Synchrotron Source, Cornell University

10 <sup>4</sup>Department of Computer Science, Cornell University

11 <sup>5</sup>Department of Materials Science and Engineering, University of Tennessee

12 <sup>6</sup>Department of Materials Science and Engineering, University of Maryland

13 <sup>7</sup>NIST Center for Neutron Research, National Institute of Standard and Technology

14 <sup>8</sup>Courant Institute of Mathematical Sciences, New York University

15 <sup>9</sup>Department of Physics, Northern Illinois University

16 \*To whom correspondence should be addressed; E-mail: eun-ah.kim@cornell.edu.

17 (Dated: March 9, 2021)

SUPPLEMENTARY INFORMATION

I. X-RAY MEASUREMENTS

A schematic of the x-ray measurement scattering geometry is shown in Fig. 1 of the main article. Three-dimensional volumes of diffuse X-ray scattering were collected at Advanced Photon Source (APS) and CHESS. The APS data were measured on sector 6-ID-D using an incident energy of 87.1 keV and a detector distance of 638 mm, except for the high-resolution measurements on  $\text{Cd}_2\text{Re}_2\text{O}_7$ , which used an incident energy of 60.0 keV and a distance of 1406 mm. The raw images were collected on a Dectris Pilatus 2M with a 1-mm-thick CdTe sensor layer. The data were collected over a temperature range of 30 K to 300 K, with samples cooled by flowing He gas below 105 K and  $\text{N}_2$  gas above 105 K. The CHESS data on  $\text{TiSe}_2$  were measured on beamline A2 using an incident beam energy of 59 keV and a Dectris Pilatus 6M detector with a 1-mm-thick Si sensor layer. The data were collected over a temperature range of 90 K to 300 K, with samples cooled by flowing  $\text{N}_2$  gas. During the measurements, the samples were continuously rotated about an axis perpendicular to the beam at  $1^\circ\text{s}^{-1}$  over  $360^\circ$ , with images read out every 0.1 s. Three sets of rotation images were collected for each sample at each temperature to fill in gaps between the detector chips. The resulting images were stacked into a three-dimensional array, oriented using an automated peak search algorithm and transformed in reciprocal space coordinates using the software package CCTW (Crystal Coordinate Transformation Workflow), allowing  $S(\mathbf{Q})$  to be determined over a range of  $\sim \pm 15 \text{ \AA}^{-1}$  in all directions ( $\sim \pm 6 \text{ \AA}^{-1}$  for the high-resolution measurement on  $\text{Cd}_2\text{Re}_2\text{O}_7$ ). Further details are given in ref. 1.

II. X-TEC: XRD TEMPERATURE CLUSTERING

A. X-TECd analysis of  $\text{TiSe}_2$  CDW ordering

This section illustrates the steps of the *X-TEC* pipeline benchmarked on a well-known CDW material:  $\text{TiSe}_2$ <sup>2,3</sup>. Fig S1 shows the outcome of *X-TECd* applied to XRD data of bulk 1T- $\text{TiSe}_2$ , collected at the Cornell High Energy Synchrotron Source (CHESS). As a test case, we specifically explored non-Bragg trajectories with the number of clusters set to  $K = 2$ .

*X-TEC* starts by collecting XRD data on a single crystal encompassing many Brillouin zones in reciprocal space over a range of temperatures  $\{T_1, \dots, T_{dT}\}$  [see Fig. S1(a)]. The data is then put through a two-stage preprocessing to deal with two key challenges against working with comprehensive data: the volume and the dynamic range of the intensity scale. First, we threshold our data in order to simultaneously reduce its size and isolate its meaningful features. The volume is set by the  $\sim 10^9$  grid points in 3D reciprocal space grid  $\{\vec{q} = (q_x, q_y, q_z)\}$  for a single temperature and the 10-30 temperature measurements typically collected. However, the relevant peaks are sparse in  $\vec{q}$ -space for crystalline samples. We thus developed an automated thresholding algorithm (SM section II.B) which removes low intensity noise and reduces the number of  $\vec{q}$ -space points to be canvassed from the full grid to a selection of points  $\{\vec{q}_i\}$ , see Fig. S1(b). Removing the intensities associated with Bragg peaks further simplifies the search.

Second, we rescale the remaining temperature series  $\{I(\vec{q}_i, T_j), j = 1, \dots, d^T\}$  still exhibiting a formidable dynamic range [see Fig. S1(d), and Fig. 2(a) of main text] in order to compare trajectories at different scales, focusing on their temperature dependence rather than the absolute scale. Different rescaling schemes can be applied depending on the nature of the data. For  $\text{TiSe}_2$ , since Bragg peak intensities are removed, subtracting each non Bragg intensity trajectory by its temperature mean is sufficient to extract their distinct temperature dependencies. For other data sets that involve Bragg peaks in the analysis such as  $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$ , the considerable increase in the range of intensity requires a better rescaling scheme. For this, each intensity trajectory is assigned a z-score (divided by standard deviation after its average value is subtracted). With some datasets such as  $\text{Sr}_3\text{Rh}_4\text{Sn}_{13}$ , we find it useful to employ an alternative rescaling scheme that facilitates further variance-based thresholding as described in the following section (SM. II-B).

[5] We now cluster the resulting collection of preprocessed temperature trajectories,  $\tilde{\mathbf{I}}(\vec{q}_i) \equiv \{\tilde{I}(\vec{q}_i, T_j); j = 1, \dots, d^T\}$  for each  $\vec{q}_i$  to discover qualitatively distinct types of temperature dependence in the data. For this, we adopt a Gaussian mixture model (GMM)<sup>4</sup>. Our approach with *X-TECd* is to ignore correlations between different reciprocal space points ( $\vec{q}$ 's) and treat each temperature series  $\tilde{\mathbf{I}}(\vec{q}_i)$  as an independent point in the  $d^T$  dimensional Euclidean space  $\mathbf{R}^{d^T}$ . The GMM assumes that each point in the data set  $\{\tilde{\mathbf{I}}(\vec{q}_i)\}$  has been independently and identically generated by a weighted sum of  $K$  distinct multivariate normal distributions. The number of clusters,  $K$ , is the only parameter we set manually.

From Fig. S1(e), the contrast between the means of the yellow cluster and the teal cluster makes it evident that the yellow cluster represents the order parameter and the teal cluster represents the temperature at which it crashes down is the critical temperature.

69 The separation between the means exceeding the individual variance affirms the clustering to be a meaningful result.  
70 Interpretation of the *X-TEC* results is immediate upon locating the two clusters in reciprocal space, as shown in  
71 Fig S1(c), and inspecting the raw data. The locations of the yellow pixels identify the CDW wave vector to be  
72  $\vec{Q}_{CDW} = \{(\pi, 0, \pi), (\pi, \pi, \pi)\}$ , and equivalent momenta in the hexagonal basis. *X-TEC* thus detected the CDW  
73 transition with the correct transition temperature  $T_c = 200$  K and correct ordering wavevector  $\vec{Q}_{CDW}$ <sup>5</sup> without any  
74 prior knowledge.

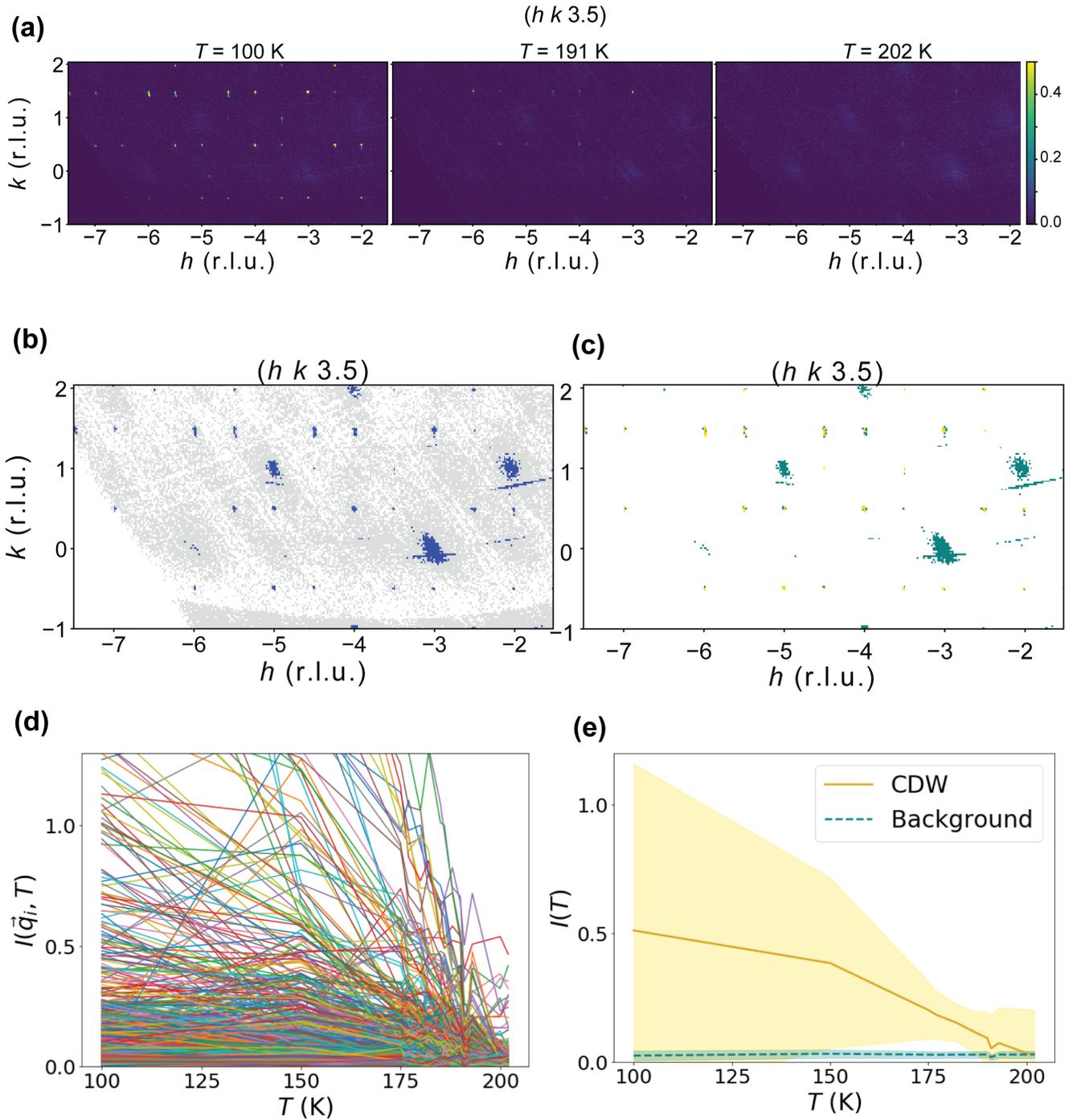


FIG. S1. *X-TECd* analysis of  $\text{TiSe}_2$ . (a) Two-dimensional slices of intensity of  $1\text{T-TiSe}_2$  on the  $l = 3.5$  plane at three temperatures. This plane contains super-lattice peaks at  $T < T_c = 200\text{K}$  (left) that disappears with the melting of the CDW order (right).  $(h, k, l)$  are in reciprocal lattice units (r.l.u.), and the color-map over saturates intensity (arb. unit.)  $> 0.5$ . (b) Thresholding described in SM. IIB removes the grey clusters in the reciprocal space of the plane shown in (a). Only the blue clusters belonging to a set  $\{\vec{q}_i\}$  are tracked using *X-TECd*. (c) *X-TECd* two-cluster ( $K = 2$ ) results assign the colors yellow and teal to the blue pixels of (b). The locations of yellow pixels identify with the CDW peaks, while teal pixels identify with the background scattering. (d) Raw non-Bragg intensity trajectories over  $d^T = 14$  temperature values,  $\{T_1 = 100\text{K}, \dots, T_{14} = 202\text{K}\}$ , of all  $\vec{q}_i$ -points in a  $3 \times 3 \times 3$  BZ's used for clustering. (e) Temperature trajectories of the two clusters from the GMM analysis of data in (d). Lines denote cluster means  $\mathbf{m}$  and shading represents covariance  $\mathbf{s}$  (one standard deviation) for the non-trivial CDW cluster (yellow) and the background cluster (teal), interpolated between the  $d^T = 14$  temperatures measured.

## B. Preprocessing

In this section we describe the technical details of *X-TEC*. A signature difficulty in the analysis of X-ray diffraction data is the existence of physics at several different intensity scales. This is only further exacerbated when probing low-intensity features where the signal-to-noise ratio can be small. If one is to employ thresholding as part of some preprocessing, it is imperative to be careful in order to avoid thresholding-out any important physics. Nevertheless, thresholding is extremely useful for mitigating the influence of noise and for reducing dataset size since most single crystal x-ray diffraction patterns are sparse. Consequently, we propose a new thresholding methodology for isolating the physically relevant regions of k-space.

A naive way to cluster the type of datasets offered by single crystal x-ray diffraction is to apply an i.i.d. assumption and directly try to cluster the associated trajectories,  $I(\vec{q}_i, T)$ , so that each  $q$ -point is classified according to its functional temperature dependence. However, such an attempt is immediately thwarted by the existence of a continuum of trajectories spanning over a large intensity range as seen in Fig. 2(a) so that getting any meaningful clustering is difficult. The standard way of dealing with this is to use feature scaling a.k.a. standardization in which one removes the mean for each trajectory and then normalizes it by dividing by its standard deviation. However, the dominant features of x-ray diffraction data are usually relatively well-localized peaks and most trajectories may be attributed to background fluctuations and thermal diffuse scattering. These trajectories have small, finite means and variances so that conventional standardization amplifies the underlying experimental error and noise, thereby spoiling any immediate attempt at clustering. On the other hand, failing to standardize makes it difficult to cluster over different energy scales since low-intensity variations can be washed out by larger ones. Thus some cutoff is needed in order to avoid clustering over noise while maintaining the ability to cluster over different energy scales.

In order to properly threshold our data, we exploit the statistical properties of our trajectories' average intensities,  $\log \overline{I(\vec{q}_i, T)}$ . Here, the average is performed over temperature so that a single average intensity is obtained for each  $\mathbf{q}$ . Several properties of our data make it advantageous to examine the statistics of  $\log \overline{I(\vec{q}_i, T)}$  rather than  $\overline{I(\vec{q}_i, T)}$ , most notably its positive semi-definiteness and large range. Since the dominant features our data are naturally sparse and the background trajectories are characterized by possessing small means and variances, we should expect the distribution of  $\overline{I(\vec{q}_i, T)}$  to be sharply peaked near some relatively small background value. Looking at the logarithm,  $\log \overline{I(\vec{q}_i, T)}$ , broadens this peak allowing us to resolve the finer structural details of this low-intensity background. To first order, we find the distribution of  $\log \overline{I(\vec{q}_i, T)}$  to be well-characterized by a bulk background contribution that is approximately normally distributed at low intensities with sparsely distributed high intensity contributions. This can be seen in when looking at the distribution of  $\log \overline{I(\vec{q}_i, T)}$  for a single unit-cell of  $\text{TiSe}_2$  in fig. S2. In order to separate these high intensity features from rest of the data, we take advantage of their sparsity relative to the background. Specifically, we minimize the Kullback-Leibler divergence  $D_{KL}$ , where for probability distributions  $p(x), q(x)$ :

$$D_{KL}(p(x)||q(x)) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)} \quad (1)$$

between the distribution of  $\log \left( \overline{I(\vec{q}_i, T)} \right)$  with a high intensity cutoff and a gaussian. Information theoretically, the Kullback-Leibler divergence quantifies the information loss associated with approximating the distribution  $p(x)$  by  $q(x)$ . In this context, the minimizing  $D_{KL}$  optimally chooses a high-intensity cutoff so that the distribution of the remaining  $\log \overline{I(\vec{q}_i, T)}$  looks closest to a normal distribution. This is illustrated by applying our procedure to a single unit-cell of  $\text{TiSe}_2$  in Fig. S2. Optimization is performed via gradient descent. Note that optimizing with this sliding cutoff is necessary and a Gaussian cannot be directly fitted because the distribution  $\log \overline{I(\vec{q}_i, T)}$  is heavy tailed. Directly fitting with a Gaussian yields a higher cutoff susceptible to missing important low-intensity features.

After thresholding, we find it convenient for the  $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$  data to rescale by dividing by the mean and subtracting one. This has the advantage over z-scoring of allowing us to implement another thresholding step in which we only cluster over high variance trajectories. In particular, it bolsters the model's ability to cluster distinct functional behaviors together because clusters can no longer be smoothly connected to the origin. However, in the case of the  $\text{TiSe}_2$ , we found it sufficient to simply subtract the mean.

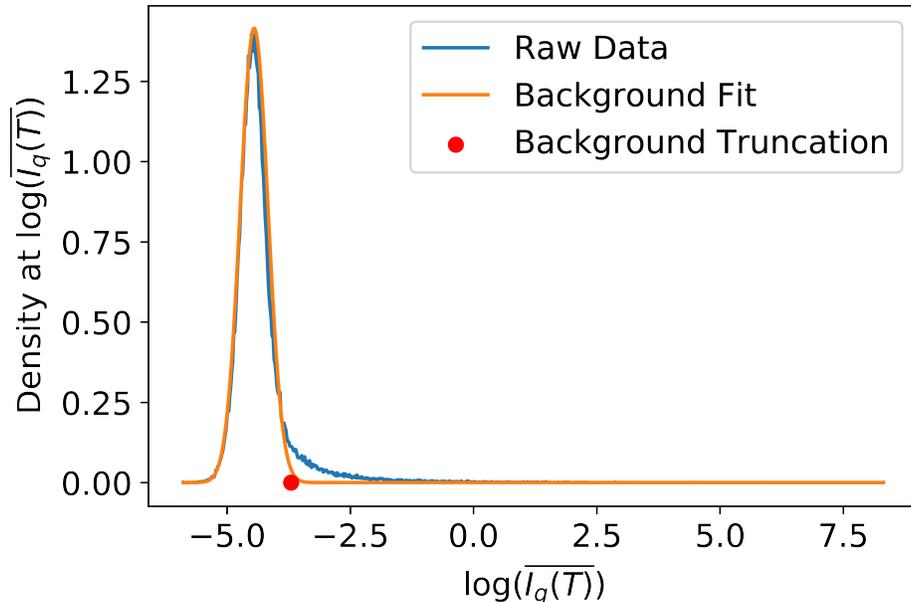


FIG. S2. Histogram (blue) of  $\log \overline{I(\vec{q}_i, T)}$  for a single unit-cell of  $\text{TiSe}_2$  with background fit (orange) and truncation point described in the main text.

119

### C. Label Smoothing

120  $X$ -TECs i.i.d. assumption ignores correlation between nearby momenta and between different Brillouin zones.  
 121  $X$ -TECs incorporate label smoothing as a first order approach for incorporating these correlations by allowing labels  
 122 to diffuse between neighboring points and between unit cells. It ultimately results in cleaner, smoother classifications  
 123 that better align with physicists intuition for order parameters.

124 Typical label smoothing is a semi-supervised method in which there exists a ground truth for certain points. These  
 125 labelings are then “clamped” and diffused through the rest of the system. Here, we lack a bona fide ground truth and  
 126 so instead incorporate label smoothing dynamically in between the E and M steps of our EM algorithm. Physically,  
 127 this adds a diffusive “force” to our update scheme that encourages a similar labeling of nearby points and points  
 128 differing by a reciprocal lattice vector. Convergence in this modified EM method occurs when an equilibrium is  
 129 reached between this diffusion and the GMM clustering.

130 Our label smoothing requires us to construct a weighted graph connecting similar momenta in order for diffusion  
 131 to occur. This may be done by computing the following kernel:

$$K(k, k') = \exp \left[ - \sum_i \sin^2 \left( \frac{Q_i \cdot (k - k')}{2} \right) / \ell^2 \right] \quad (2)$$

132 where the  $Q_i$  are the reciprocal basis vectors and  $\ell$  is the relevant length scale for the local correlations. The structure  
 133 of this kernel is shown in Fig. S3 where  $K(k, 0)$  is plotted as an intensity for a 2D slice.

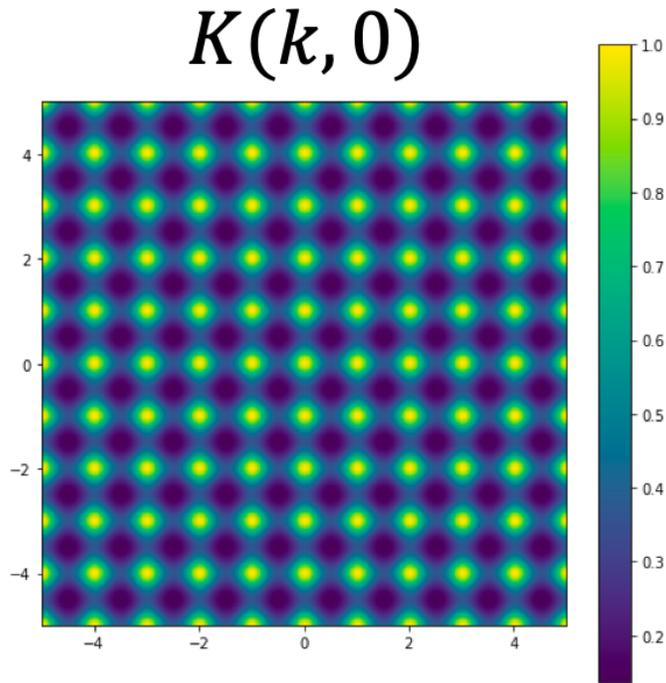


FIG. S3. Kernel,  $K(k, 0)$ , showing the similarity between the origin and momenta in a 2D.

134 This kernel is really just a weighted adjacency matrix. By incorporating a cutoff in the weights, we may exploit the  
 135 sparsity of our system for fast matrix-vector multiplication. When handling large datasets, this cutoff is essential since  
 136 the full kernel is too large to be stored in any reasonable amount of RAM. Define  $A$  to be the matrix associated with  
 137 this kernel after having normalized the rows i.e. it is row stochastic so that  $\sum_j A_{ij} = 1$ . Now define  $P$  to be the matrix  
 138 consisting of cluster probabilities calculated by the E-step. Specifically, let the first index correspond to the different  
 139 momenta and the second to the cluster probabilities so that  $P$  is also row stochastic. Then the product  $AP$  is also  
 140 row stochastic since  $\sum_{j^k} A_{ij} P_{jk} = \sum_j A_{ij} (1) = 1$ . So by multiplying  $P$  by  $A$ , we generate a new set of diffused cluster  
 141 probabilities. The strength of this diffusion can be controlled by the number of matrix multiplications. However, note  
 142 that we cannot simply apply  $A$  until  $A^n P$  converges, because the largest eigenvector of  $A$  is just the constant vector.  
 143 In practice, we find that even a single application of  $A$  between E- and M-steps is sufficient for obtaining smooth  
 144 labelings.

145

#### D. $X$ -TECd analysis of $\text{Sr}_3\text{Rh}_4\text{Sn}_{13}$

146 In Fig. 2 of the main text, we report the  $X$ -TECs results of  $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$  XRD data. Here we contrast the  
 147  $X$ -TECs results of  $\text{Sr}_3\text{Rh}_4\text{Sn}_{13}$  with the unsmoothed  $X$ -TECd analysis in Fig. S4. Without label smoothing, Fig. S4(b)  
 148 shows that nearby  $\vec{q}_i$  points are often assigned to different clusters. Label smoothing automatically harmonizes the  
 149 assignments in the vicinity of each peak at the cost of weakening the cluster separation.

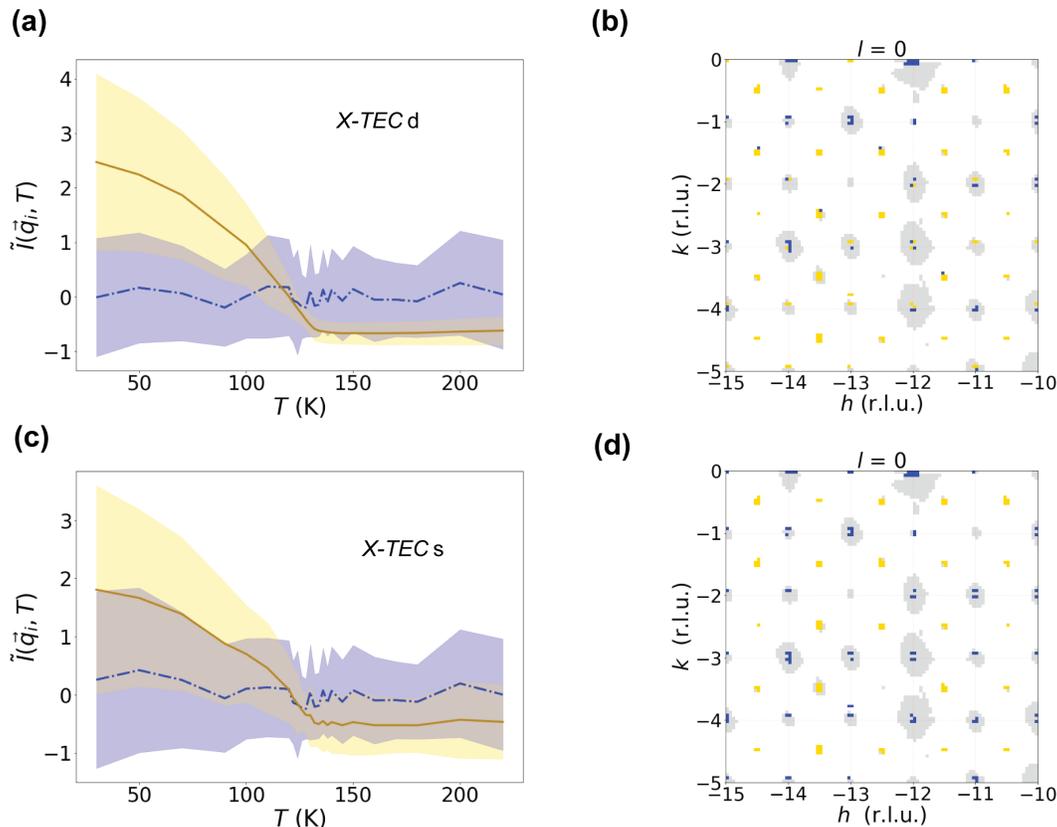


FIG. S4. Contrasting the unsmoothed  $X\text{-TECd}$  results of upper panels (a-b) with the label smoothed  $X\text{-TECs}$  results of lower panel (c-d) for the XRD data from  $\text{Sr}_3\text{Rh}_4\text{Sn}_{13}$ . **(a,c)** The cluster trajectories with clustering assignments color-coded as yellow and blue.  $\tilde{I}(\tilde{q}_i, T)$  are the rescaled intensities by dividing each intensity trajectory with its mean over temperature and subtracting one. The lines represent cluster means and the shaded region shows one standard deviation. **(b,d)** The corresponding yellow/blue cluster assignments in the  $(h, k, 0)$  plane. The low intensity background (white) and  $\tilde{I}(\tilde{q}_i, T)$  with low temperature variance (grey pixels) are excluded from GMM clustering by the preprocessing.

150

### E. Derivation of EM algorithm for GMM and general proof of convergence.

151 We follow derivations in Refs. 4 and 6. First recall Jensen's inequality: for convex function  $f$  and random variable  
 152  $X$ ,  $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$  where for strictly convex functions, equality holds iff  $X = \mathbb{E}[X]$  almost surely. Let  $\ell(\theta)$ , denote  
 153 the model log-likelihood and  $X$  be our dataset with  $x_i \in X$ . Then

$$\begin{aligned}
 \ell(\theta) &= \log p(X; \theta) = \sum_i \log p(x_i; \theta) = \sum_i \log \sum_{z_i} p(x_i, z_i; \theta) \\
 &= \sum_i \log \sum_{z_i} q_i(z_i) \frac{p(x_i, z_i; \theta)}{q_i(z_i)} \geq \sum_{i, z_i} q_i(z_i) \log \frac{p(x_i, z_i; \theta)}{q_i(z_i)} \equiv \tilde{\ell}(q, \theta)
 \end{aligned} \tag{3}$$

154 where  $q_i(z_i)$  is some distribution over a random variable  $z_i$  (in our case this will be the cluster assignment) s.t.  
 155  $\sum_{z_i} q_i(z_i) = 1$  and we have used Jensen's inequality. In order for this bound to be tight,  $X = \mathbb{E}[X] \implies q_i(z_i) =$   
 156  $p(z_i|x_i; \theta)$ . Tightness of this bound implies that improving  $\tilde{\ell}(q, \theta)$  necessarily improves  $\ell(\theta)$  but since theta is unknown,  
 157 we will have to make a guess,  $\theta_t$ , and improve it iteratively. This iterative prescription is known as expectation  
 158 maximization (EM). It consists of an E-step, where  $q_i^t \leftarrow p(z_i|x_i; \theta_t)$  and an M-step  $\theta^{t+1} \leftarrow \underset{\theta}{\operatorname{argmax}} \tilde{\ell}(q^t, \theta)$ .

We now derive the EM algorithm for the GMM. The E-step follows directly from the model likelihood and Bayes'

theorem:

$$w_i^k \equiv p(z_i = k | x_i; \pi_k, \mu_k, \Sigma_k) = \frac{\pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)}{\sum_k \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)} \quad (4)$$

$$\mathcal{N}(x_i | \mu_k, \Sigma_k) \equiv \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det \Sigma_k}} e^{-\frac{1}{2}(x_i - \mu_k)^\dagger \Sigma_k^{-1} (x_i - \mu_k)}$$

159 For the M-step, we must find  $\{\pi, \mu, \Sigma\}$  that optimizes our lower log-likelihood bound:

$$\tilde{\ell}(\{w_i^k, \pi_k, \mu_k, \Sigma_k\}) = \sum_{i,k} w_i^k \log \left[ \frac{\pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k)}{w_i^k} \right] + \lambda (1 - \sum_k \pi_k) \quad (5)$$

160 where  $\lambda$  is a Lagrange multiplier constraining the mixing weights to sum to unity.

161 Solving for the mixing weights:

$$\begin{aligned} 0 = \partial_{\pi_j} \tilde{\ell} &= \sum_{i,k} w_i^k \frac{1}{\pi_k} \delta_{jk} - \lambda \sum_k \delta_{jk} \implies \lambda = \frac{1}{\pi_j} \sum_i w_i^j \\ \lambda = \lambda \sum_k \pi_k &= \sum_{i,k} w_i^k = \sum_i 1 \equiv m \\ \implies \pi_j &= \frac{1}{m} \sum_i w_i^j \end{aligned} \quad (6)$$

162 Solving for the mean:

$$\begin{aligned} 0 = \partial_{\mu_l} \tilde{\ell} &= 2 \sum_i w_i^l \Sigma_l^{-1} (x_i - \mu_l) \\ \implies \mu_l &= \frac{1}{\sum_i w_i^l} \sum_i w_i^l x_i \end{aligned} \quad (7)$$

163 Solving for the covariance is a little trickier. First note the following matrix identities for symmetric invertible  
164 matrix  $A$ :

$$\begin{aligned} \partial(\log(\det A)) &= \text{Tr}(A^{-1} \partial A) \\ \partial A^{-1} &= -A^{-1} (\partial A) A^{-1} \end{aligned} \quad (8)$$

165 Now, when solving for the covariance we promote the covariance cluster index to an upper index so that the lower  
166 indices refer to the matrix elements:

$$\begin{aligned} 0 = \partial_{\Sigma_{mn}^l} \tilde{\ell} &= \sum_{i,k} w_i^l \partial_{\Sigma_{mn}^l} \left[ \log \det \Sigma^k + (x_i - \mu_k)^\dagger (\Sigma^k)^{-1} (x_i - \mu_k) \right] \\ &= \sum_{i,k} w_i^l \left[ \delta_{lk} \text{Tr} \{ (\Sigma^{k-1})_{rs} \delta_{sm} \delta_{tn} \} - \delta_{lk} \sum_{ps} (x_i - \mu_k)_p^\dagger \left\{ \sum_{qr} (\Sigma^{k-1})_{pq} \delta_{mq} \delta_{nr} (\Sigma^{k-1})_{rs} \right\} (x_i - \mu_k)_s \right] \\ &= \sum_i w_i^l \left[ \Sigma_{nm}^{l-1} - \sum_{p,s} (x - \mu_l)_p^\dagger \Sigma_{pm}^{l-1} \Sigma_{ns}^{l-1} (x - \mu_l)_s \right] \\ &= \sum_i w_i^l \left[ \Sigma^{l-1} - \Sigma^{l-1} (x_i - \mu_l) (x_i - \mu_l)^\dagger \Sigma^{l-1} \right] \\ 0 = \sum_i w_i^l &\left[ \Sigma^l - (x_i - \mu_l) (x_i - \mu_l)^\dagger \right] \\ \implies \Sigma_l &= \frac{1}{\sum_i w_i^l} \sum_i w_i^l (x_i - \mu_l) (x_i - \mu_l)^\dagger \end{aligned} \quad (9)$$

167 Note that all quantities derived about have the same form as one would expect from standard regression but with  
168 each data point  $x_i$  having a cluster weight  $w_i^k$ .

### III. $\text{Cd}_2\text{Re}_2\text{O}_7$ ANALYSIS

#### A. Specific Heat Measurements

In the main text, the heat capacity ( $C_p$ ) of  $\text{Cd}_2\text{Re}_2\text{O}_7$  was displayed in Fig. 3(b). The data shown in that figure was processed by a standard method in relaxation calorimetry (“pseudostatic method”) in which the heat capacity is assumed to be constant throughout the heating and cooling segments of an applied heat pulse during which  $\Delta T \ll T$ . However, in the presence of a 1st order transition, the shape and magnitude of a peak in  $C_p$  at the phase transition temperature can be modified, while the hysteresis can be lost, when using the pseudostatic method. For this reason, we have also used the “scanning method” for which  $C_p$  is numerically determined at every point in the warming and cooling segments, which yields a more accurate peakshape and hysteresis for a 1st order transition at the cost of noise and absolute accuracy. A more detailed description of pseudostatic and scanning analysis can be found in Ref. 7. Fig. S5 shows the temperature dependence of  $C_p$  in the vicinity of the  $\sim 113$  K phase transition when analyzed using the scanning method. A small but resolvable thermal hysteresis was observed between the peaks in  $C_p$  from the heating and cooling curves, which is suggestive of a latent heat and hence a first-order character. We do note, however, that the peak height and width of the peak in  $C_p$  did not differ substantially between these two methods, as would also be anticipated for a first-order transition, and for this reason the analysis of  $C_p$  alone is not definitive in identifying the order of the transition.

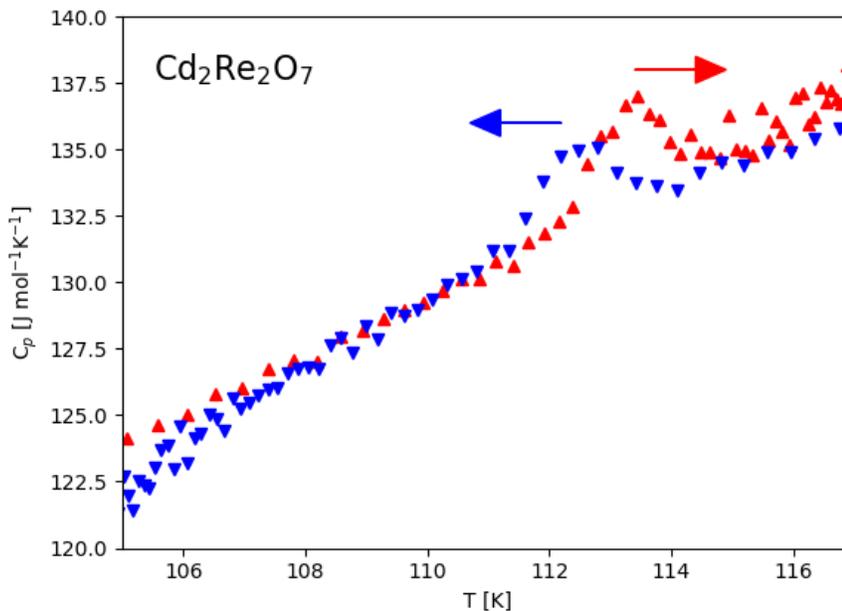


FIG. S5. Specific heat of  $\text{Cd}_2\text{Re}_2\text{O}_7$ , measured using the scanning method described in the text.

#### B. Low resolution $\text{Cd}_2\text{Re}_2\text{O}_7$ XRD data

This section discusses the  $X$ -TEC analysis with a lower resolution XRD data of  $\text{Cd}_2\text{Re}_2\text{O}_7$ . We first performed scans using an x-ray energy of 87 keV, which contained scattering spanning nearly 15,000 Brillouin zones. A first pass of  $X$ -TECs<sup>8</sup> for two clusters ( $K = 2$ ) readily finds a cluster whose intensity rises sharply below  $T_{s1} = 200$  K [the purple cluster in Fig. S6(a)]. The crisp clustering results with tight variance around the means reflect the amplification of the meaningful trend upon using data from a large number of BZ’s. By examining the  $X$ -TEC cluster assignments, we find the purple cluster to exclusively consist of peaks with  $\vec{Q} = (H, K, L)$ , with all indices even, exactly one of which is not divisible by four, using the cubic indices of Phase I [see Fig. S6(b)]. Peaks that are equivalent in the cubic phase have different temperature dependence in Phase II, implying that the sample is untwinned, something that is confirmed by our high-resolution data. This means that the presence of  $(00L)$  peaks with  $L = 4n + 2$  below  $T_{s1}$  in phase II unambiguously rules out all the tetragonal space groups compatible with the pyrochlore structure, apart

196 from  $I\bar{4}m2$  and  $I\bar{4}$ . According to an earlier group theoretical analysis<sup>9</sup>, of these two, only the former is compatible  
 197 with a single second-order phase transition, so our data is strong confirmation of previous conclusions that, at  $T_{s1}$ ,  
 198  $I\bar{4}m2$  phase is selected out of two-dimensional  $E_u$  representation<sup>10,11</sup>.

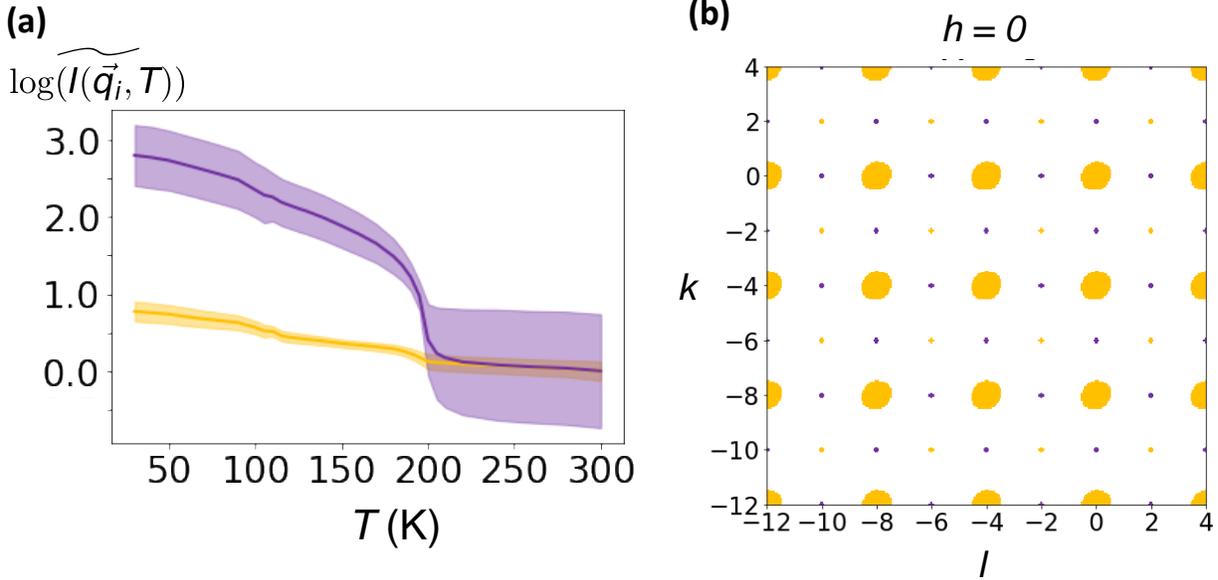


FIG. S6. Two-cluster  $X$ -TECs results on the lower resolution data spanning 15,000 BZ's of  $\text{Cd}_2\text{Re}_2\text{O}_7$ . (a) Cluster means (solid lines) and standard deviations (shaded areas) for the two clusters are shown in purple and yellow, interpolated between  $d^T = 30$  temperature points of measurement. The data is peak averaged prior to the  $X$ -TEC preprocessing to suppress fluctuation signal and isolate the transition at  $T_{s2}$ .  $\log(\widetilde{I(\vec{q}_i, T)})$  denotes rescaled data by logging the peak averaged intensity, and subsequently subtracting its mean over temperature. (b) The cluster assignments of thresholded  $\vec{q}_i$  points that belong to the two clusters in (a) in a portion of the  $h = 0$  plane.

### C. Preprocessing and clustering setup details

Here we specify different preprocessing steps and clustering choices for the analysis of  $\text{Cd}_2\text{Re}_2\text{O}_7$  high resolution data presented in Fig. 3 and 4 of main text, as well as the lower resolution data in Fig S6.

#### • $X$ -TECs (peak averaged) on cubic forbidden peaks of high resolution data: Fig 3(c), Fig 4(a)

1. We begin by selecting a  $50 \times 50 \times 50$  region around each known peak center and thresholding as described in SM IIB.
2. We then floodfill from the peak centers and average all resulting trajectories to form a single, averaged trajectory per peak.
3. We rescale the data by z-scoring it.
4. We exclude all the cubic allowed Bragg peaks, and restrict the temperature range to [30 K, 150 K] so that  $X$ -TEC can focus on better resolving the distinct cluster trajectories across  $T_{s2}$ . See Fig S7 for the same analysis, but including all Bragg peaks and over the full temperature range.
5. We cluster the peak-averaged trajectories using  $K = 2$  clusters. We found two clusters to be the minimum number necessary to separate all distinct behaviors and that there was no advantage to using more than two.
6. The dashed lines in Fig. 3(c) and symbols in Fig 4(a) show the cluster averaged intensity trajectory of the two clusters. The cluster averaged trajectory is shown for the full temperature range: [30 K, 300 K], although the clustering assignments were obtained from trajectories  $\leq 150$  K.

#### • $X$ -TECd (peaks opened) on cubic forbidden peaks of high resolution data: Fig 3(c-d), Fig: 4(c)

1. We select a  $50 \times 50 \times 50$  window around each known peak center and threshold as described in SM II(a).

- 219 2. We only include peaks that are forbidden in the cubic phase. The temperature range is restricted to [30  
220 K, 150 K] like in the peak averaged *X-TEC*s analysis.
- 221 3. We rescale the data by z-scoring it.
- 222 4. We cluster the data using  $K = 3$  clusters.
- 223 5. The resulting cluster averaged intensity trajectory for the full temperature range: [30 K, 300 K] is shown  
224 as solid lines in Fig. 3(c).
- 225 6. Cluster averaging the absolute intensity trajectories washes out the characteristic temperature dependence  
226 of the diffuse halos. This can be remedied by cluster averaging over z-scored intensities. This is reported  
227 in Fig. 4(c). To confirm that z-scored intensity indeed represents the behavior of diffuse halos, see SM  
228 Fig. S9 for the absolute intensity trajectories of the diffuse halos in two manually selected peaks.

229 • ***X-TEC*s (peak averaged) on low resolution data: Fig S6**

- 230 1. In order to reduce noise, we first construct an average BZ mask by thresholding every BZ as described in  
231 SM IIB and then averaging the thresholded BZs together.
- 232 2. We then manually select a cutoff value for the averaged BZ that maintains all the peaks while removing  
233 as much background as possible, and set each  $\vec{q}$ -point in the average BZ with value greater than the cutoff  
234 to 1, and the rest to 0 to form the mask.
- 235 3. We multiply each BZ by the average BZ mask to remove noise and emphasize the peaks.
- 236 4. Beginning from the known peak centers, we floodfill to pick out all  $\vec{q}$ -points belonging to each peak.
- 237 5. We perform peak averaging by averaging the trajectories of all  $\vec{q}$ -points belonging to each peak and replacing  
238 them with the single, averaged trajectory.
- 239 6. We rescale the data by taking the log of one plus each peak-averaged trajectory, and subsequently sub-  
240 tracting the mean.
- 241 7. Finally we cluster using  $K = 2$  clusters. We subtract the minimum value of the cluster means when plotting  
242 to emphasize the order-parameter like behavior of the purple cluster in Fig. S6(a). Here *X-TEC* analyses  
243 the data for the full temperature range [30 K, 300 K].

244 **D. Structure Factor Analysis**

245 Fig 3 and 4 of main text discuss the two-clustering *X-TEC*s analysis after excluding the cubic allowed peaks. By  
246 including all Bragg peaks, four clusters are sufficient for *X-TEC*s to identify all the distinct trajectories. Fig. S7 shows  
247 the cluster means (z-scored intensity trajectories) for all four clusters identified by *X-TEC*s (peak averaged) analysis.  
248 Two of these sub-clusters (yellow and green symbols) can be identified with the behavior of the cubic excluded peaks.  
249 It should be noted that these clusters represent the average temperature dependence of all the peaks assigned to  
250 their respective clusters, so there can be large variations within each cluster. However, the ML analysis has identified  
251 distinctive behavior in each cluster that we have verified by manual inspection of a number of peaks. All four clusters  
252 show similar temperature dependence close to the transition at  $T_{s1} = 200$  K, but strikingly different behavior at the  
253 lower transition at  $T_{s2} = 113$  K. The yellow cluster trajectory show a sudden drop while the green cluster peaks show  
254 a sudden increase in intensity across  $T_{s2}$ . The magenta and brown lines show a sharp spike in intensity at  $T_{s2}$ , before  
255 falling back to their values just above the transition. We do not currently have an explanation for this remarkable  
256 behavior.

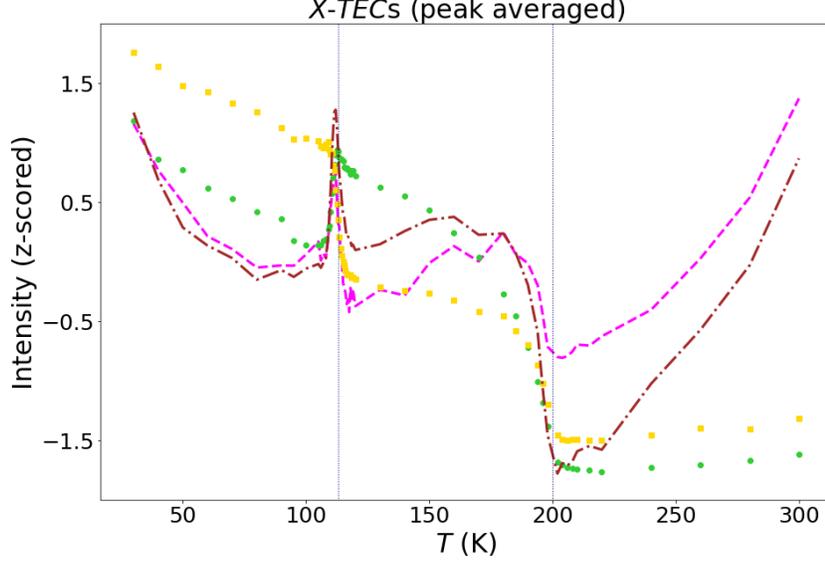


FIG. S7. Four-cluster  $X$ -TECs (peak averaged) results on the high resolution measurements of  $\text{Cd}_2\text{Re}_2\text{O}_7$  retaining all Bragg peaks. Two of these sub-cluster trajectories (yellow and green symbols) identify with the cubic forbidden trajectories shown in Fig. 3(c) and Fig. 4(a) of main text. The other two sub-cluster trajectories (magenta and brown lines) arise from peaks that are not forbidden in the high-temperature cubic phase. The temperatures of the two structural phase transitions are shown as dotted lines.

257 The structural phase transition at  $T_{s1}$  is from the cubic pyrochlore structure, with space group  $Fd\bar{3}m$ , to a distorted  
 258 tetragonal structure, with space group  $I\bar{4}m2$ . This space group allows distortions of the cadmium and rhenium cations  
 259 along the  $z$  direction and either the  $x$  or  $y$  direction depending on the Wyckoff positions, using the  $I\bar{4}m2$  unit cell,  
 260 which is rotated by  $45^\circ$  from the cubic unit cell, *i.e.*,  $x$  is parallel to the (110) direction of the high-temperature cubic  
 261 structure. There are associated displacements of the oxygen ions, but the x-ray measurements are not sensitive to  
 262 them.

263 Analytic calculations of the structure factors for the Bragg peaks in terms of the allowed  $x$  and  $z$  distortions fall into  
 264 four groups that correspond well to the four ML clusters. For example, the two groups whose intensities are forbidden  
 265 in the high-temperature cubic phase (yellow and green) have the following form (H,K,L in following equations are in  
 266 tetragonal indices):

$$F_1(H, K, L) \propto (-1)^{n_3} \sum_{M=\text{Cd,Re}} \{f_M [(-1)^{n_1} \cos(2\pi H \delta x_M) e^{-2\pi i L \delta z_M} - (-1)^{n_2} \cos(2\pi K \delta x_M) e^{2\pi i L \delta z_M}]\} \quad (10)$$

267 where  $n_1 = \frac{1}{2}H$ ,  $n_2 = \frac{1}{2}K$ , and  $n_3 = \frac{1}{4}(L - 2)$ .

$$F_2(H, K, L) \propto (-1)^{n_3} \sum_{M=\text{Cd,Re}} \{f_M [(-1)^{n_1} \sin(2\pi H \delta x_M) e^{-2\pi i L \delta z_M} + (-1)^{n_2} \sin(2\pi K \delta x_M) e^{2\pi i L \delta z_M}]\} \quad (11)$$

268 where  $n_1 = \frac{1}{2}(H - 1)$ ,  $n_2 = \frac{1}{2}(K - 1)$ , and  $n_3 = \frac{1}{4}L$ .

269 It can be seen that, for small values of  $H$  and  $K$ ,  $F_1(H, K, L)$  are mostly sensitive to distortions along the  $z$ -axis,  
 270 whereas for small values of  $L$ ,  $F_2(H, K, L)$  is mostly sensitive to in-plane distortions along  $x$  or  $y$  (where  $\delta x = \delta y$ ).  
 271 The assignments of individual peaks in the  $X$ -TEC analysis show that the (H,K,L) values of the green cluster are  
 272 indeed dominated by in-plane distortions whereas the yellow cluster peaks are dominated by  $z$ -axis distortions. This  
 273 suggests that the distinctive temperature dependences of peaks in the green and yellow clusters can be used to derive  
 274 information about the relative distortions along  $x$  and  $z$ . If we assume that the temperature dependence of  $\delta x$  and  
 275  $\delta z$  follows that expected for an order parameter with a common critical exponent,  $\beta$ , from 200K down to 120K, the  
 276 peak intensities would vary as  $(T - T_c)^{2\beta}$ .

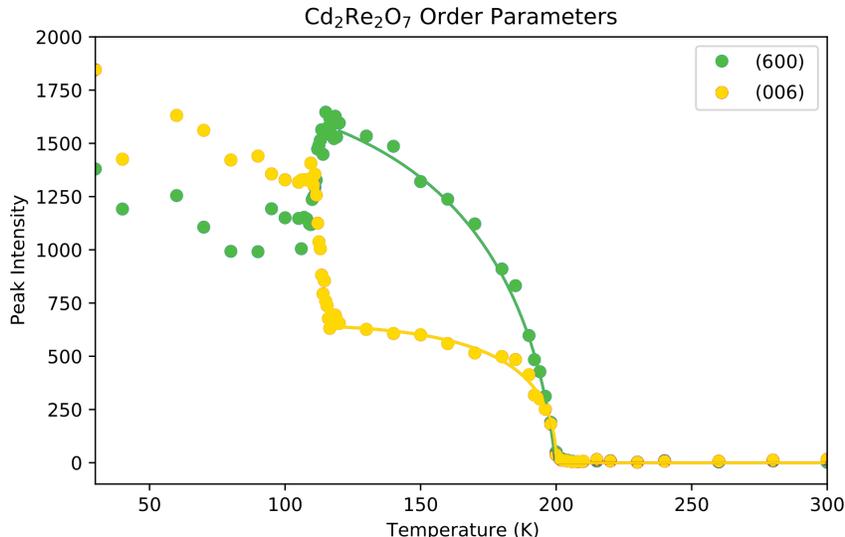


FIG. S8. Temperature dependence of the 006 (yellow) and 600 ( $3\bar{3}0$ ) in tetragonal indices (green) Bragg peaks using the indices of the high-temperature cubic phase. The green and yellow solid lines are fits between 120 K and 300 K to the structure factors in equations 10 and 11, respectively, assuming that the distortions,  $\delta x$  and  $\delta z$  for the Cd and Re ions, vary as  $(T - T_c)^\beta$ , with  $\beta = 0.25$  and  $T_c = 200$  K.

As an example, Figure S8 compares the 006 and 060 Bragg peaks using the indices of the cubic phase. These are the peaks that have been assigned to the yellow and green clusters of  $X$ -TECs shown in Fig. 4(a). Equations 10 and 11 show that the 006 (yellow) peak is only sensitive to  $\delta z_{\text{Cd}}$  and  $\delta z_{\text{Re}}$ , whereas the 060 (green) peak is only sensitive to in-plane distortions. The fit to the 006 peak yields relative  $z$ -axis distortions that are equal and opposite, *i.e.*,  $\delta z_{\text{Re}} = -\delta z_{\text{Cd}}$ , illustrated in Fig. 4(b). The out-of-phase distortions are the reason for the flattening of the peak intensity of the 006 peak between 180 K and 120 K, confirming the conclusions based on the fits to the cluster means in Fig. 4(a). On the other hand, the 060 peak follows the scaling law from 200 K to 120 K, showing either that  $\delta x_{\text{Re}}$  has the same sign as  $\delta x_{\text{Cd}}$  or that one of the distortions is much larger than the other. This is an example where the temperature dependence of the peak intensities below a structural phase transition yields information on the relative internal distortions, which have proved to be too subtle for conventional crystallographic refinement until now.

### E. Temperature Dependence of Diffuse Scattering

Fig. 3 in the main article showed two-cluster ( $K = 2$ ) assignments from  $X$ -TECd analysis of the high-resolution data with cubic excluded peaks. This reveals differences between clusters in the diffuse scattering halo around the Bragg peaks, which represent fluctuations in the order parameter at the  $\Gamma$  point. Peaks in the blue cluster, which are also assigned to the blue cluster in Fig. S8, display weak diffuse scattering halos in the range  $T_{s2} < T < T_{s1}$  while peaks in the red cluster displayed much stronger diffuse scattering halos in this region, as seen in Fig. 3(c). This is clearly illustrated in Fig. S9, where the temperature dependence of the diffuse scattering near the  $0\bar{6}0$  and  $0\bar{4}\bar{6}$  Bragg peaks are compared. Both show strong critical scattering at  $T_{s1}$ , but the diffuse contribution is much stronger in peaks assigned to the red cluster, which are most sensitive to the out-of-phase  $z$ -axis fluctuations of the Re and Cd sublattices. We attribute these strong  $z$ -axis fluctuations to the Goldstone modes seen in Raman scattering<sup>12</sup>, which are fluctuations between the two nearly degenerate  $E_u$  modes. This interpretation is justified by the fact that the Goldstone modes are fluctuations from the  $I\bar{4}m2$  ground state to  $I4_122$  symmetry, in which  $\delta z$  is constrained to be 0. It is therefore not surprising that  $z$ -axis fluctuations are dominant.

The  $X$ -TEC trajectories shown in Fig. 4(c) are the cluster average from  $z$ -scored intensities. The red cluster shows a substantial increase in diffuse scattering just above  $T_{s2}$ , whereas the blue cluster shows a weaker peak just below  $T_{s2}$ . Although 4(c) shows the  $z$ -scored intensities and not their absolute values, note its similarity with Fig. S9. This indicates that the  $X$ -TECd cluster averages of  $z$ -scored intensities give the typical diffuse trajectories.

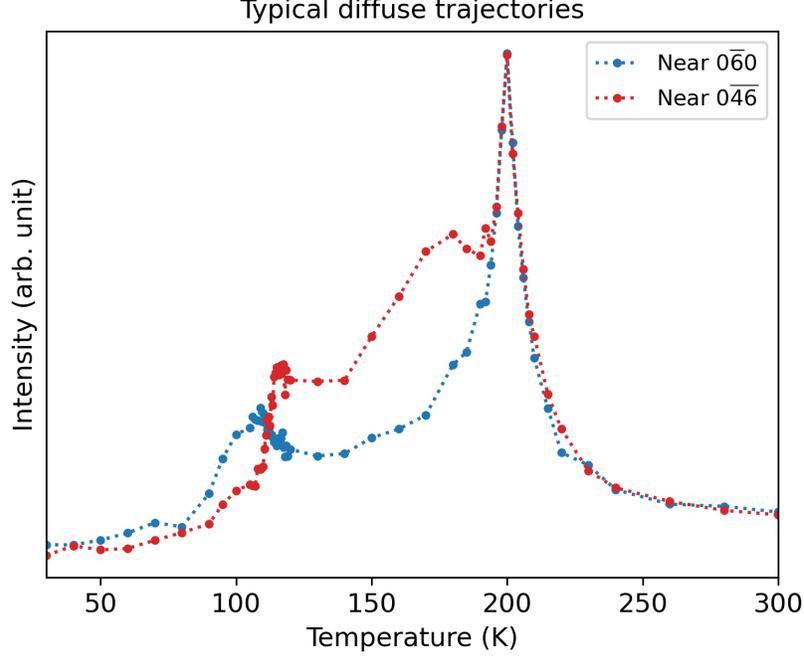


FIG. S9. Temperature dependence of the diffuse scattering near the  $\bar{0}\bar{6}0$  and  $\bar{0}4\bar{6}$  Bragg peaks shown in Fig. 3 (e-f) (blue and red circles, respectively). The temperature dependences are taken from  $0.02 \times 0.02 \times 0.02$  rlu bins centered at  $hkl = (-0.05, -6.0, 0.05)$  and  $hkl = (-6.0, -3.95, -0.05)$ .

### F. Mode energies and intensities from Landau theory

The Landau free energy in an  $E_u$  model for  $\text{Cd}_2\text{Re}_2\text{O}_7$  is<sup>9</sup>

$$F = a_1 Q^2 + a_2 Q^4 + a_3 Q^6 + a_4 Q^8 + Q^6 [b_1 + cQ^2] \frac{1}{2} [1 + \cos(6\phi)] \quad (12)$$

with  $Q$  the order parameter amplitude and  $\phi$  the phase angle. For  $I\bar{4}m2$ ,  $\phi = 30(2n+1)$  and for  $I4_122$ ,  $\phi = 60n$  (the different angles represent different domains).  $F$  vanishes at  $T_{s1}$  and the anisotropy (last term) would vanish at  $T_{s2}$  if it were not for the first-order jump in  $Q$ . The Goldstone (phase) mode energy is given by<sup>13</sup>

$$\omega_G^2 = \chi_G^{-1} / M \quad (13)$$

where the inverse Goldstone susceptibility is

$$\chi_G^{-1} = \frac{1}{Q^2} \frac{\partial^2 F}{\partial \phi^2} = 18Q^4 |b_1 + cQ^2| \quad (14)$$

with  $M$  some ion effective mass, where in the second expression we have taken into account the value of  $\phi$  in the two phases (which leads to the modulus). The Higgs mode energy is given by

$$\omega_H^2 = \chi_H^{-1} / M \quad (15)$$

where

$$\chi_H^{-1} = \frac{\partial^2 F}{\partial Q^2} = 2a_1 + 12a_2 Q^2 + 30a_3 Q^4 + 56a_4 Q^6 + 30Q^4 [b_1 + \frac{28}{15}cQ^2] \frac{1}{2} [1 + \cos(6\phi)] \quad (16)$$

The first-order transition from  $I\bar{4}m2$  (phase II) to  $I4_122$  (phase III) is given by the condition<sup>9</sup>

$$a_1 = 2a_2(b_1/c) - 3a_3(b_1/c)^2 + (4a_4 + c/2)(b_1/c)^3 \quad (17)$$

314  $Q^2$  is given by the cubic equation ( $\frac{\partial F}{\partial Q} = 0$ )

$$-2a_1 = 4a_2Q^2 + 6a_3Q^4 + 8a_4Q^6 + 6Q^4[b_1 + \frac{4}{3}cQ^2]\frac{1}{2}[1 + \cos(6\phi)] \quad (18)$$

315 Finally, the soft mode energy above  $T_{s1}$  ( $\omega_s$ ) is gotten by setting  $Q=0$  in  $\chi_H^{-1}$ . In practice, the effective  $M$  is unknown  
 316 (involving Cd, Nb and O masses), so all mode energies will be multiplied by the same constant in order to agree with  
 317 Raman data<sup>12</sup> for the Higgs energy at  $T=0$  (85 cm<sup>-1</sup>).

318 We now have all we need to calculate the order parameter, the phase boundary, and mode energies. What about  
 319 the mode intensities? The basic idea can be seen from the work of Fleury<sup>14</sup> and Shapiro<sup>15</sup>. The energy integrated  
 320 intensity (appropriate for the diffuse scattering collected from high energy x-rays) is given by<sup>15</sup>

$$I_q = \frac{1}{\pi} \int [n(\omega) + 1] \Im[\omega_q^2 - \omega^2 - i\omega\Gamma_q]^{-1} d\omega \quad (19)$$

321 where  $n(\omega)$  is the Bose factor,  $\omega_q$  is the mode energy for a given  $q$ , and  $\Gamma_q$  is the lifetime broadening. Assuming we  
 322 can replace  $n(\omega)$  by  $T/\omega$ , this integral reduces to

$$I_q = T/\omega_q^2 \quad (20)$$

323 This expression is obviously divergent for  $q=0$  at  $T_{s1}$ . To correct for this, we recognize that the data are collected  
 324 over a small range in  $q$ . We assume the  $q$  dependence of the mode energy goes like

$$\omega_q^2 = \omega_0^2 + \alpha^2 q^2 \quad (21)$$

325 where  $\alpha$  results from the gradient terms in the Landau energy. Integrating over  $q$ , we obtain

$$T \int \frac{q^2 dq}{\omega_q^2} \propto T[1 - \tilde{\omega}_0 \tan^{-1} \frac{1}{\tilde{\omega}_0}] \quad (22)$$

326 where  $\tilde{\omega}_0 = \frac{\omega_0}{\alpha q_c}$  with  $q_c$  the momentum cut-off. Since  $\alpha$  is unknown (no mode dispersions have been measured for  
 327 this material), we set  $\alpha q_c$  to the lower bound of the Raman data (6 cm<sup>-112</sup>) for all modes.

328 Now for the matrix elements. That is, how do the x-rays couple to the modes? We assume unit coupling to the  
 329 Higgs and soft modes, the Higgs mode below  $T_{s1}$  being the analog of the soft mode above  $T_{s1}$ . But for the Goldstone  
 330 mode, which only exists below  $T_{s1}$ , we set the coupling constant to  $Q^{214}$ . So, above  $T_{s1}$  we have

$$T[1 - \tilde{\omega}_s \tan^{-1} \frac{1}{\tilde{\omega}_s}] \quad (23)$$

331 and below  $T_{s1}$  we have

$$T[1 - \tilde{\omega}_H \tan^{-1} \frac{1}{\tilde{\omega}_H} + Q^2(1 - \tilde{\omega}_G \tan^{-1} \frac{1}{\tilde{\omega}_G})] \quad (24)$$

332 To evaluate, we choose parameters as in Fig. 3b of<sup>9</sup>, with  $b_1=0.3$ . We then do the following normalizations.  $a_1$   
 333 is some constant times  $T - T_{s1}$ . This constant is adjusted so that  $T_{s1}$  is 200 K and  $T_{s2}$  is 113 K. Then the mode  
 334 energies are normalized as stated above (so that the Higgs mode energy is equal to 85 cm<sup>-1</sup> at  $T=0$  as observed by  
 335 Raman<sup>12</sup>). Finally, the intensities are normalized by  $T_{s1}$ . In Fig. S10, the resulting mode energies and intensities are  
 336 shown. Note the small jump in the Higgs energy and the dip in the Goldstone energy at  $T_{s2}$ . Also that the Goldstone  
 337 intensity completely dominates outside the critical region associated with  $T_{s1}$ . As an aside, the Raman data cut-off  
 338 at about 6 cm<sup>-1</sup> as noted above. The prediction is that the Goldstone mode energy should rise above this value at  
 339 low  $T$ . We suggest then that the Raman mode seen at 30 cm<sup>-1</sup> below  $T_{s2}$  could be the Goldstone mode. This in turn  
 340 implies that the central peak in the intensity from Raman has more contributions to it than the Goldstone one, and  
 341 this would presumably be due to elastic scattering from impurities and static short-range structural disorder.

342 Finally, some caveats. First, the behavior well below  $T_{s2}$  cannot be taken too seriously since Landau theory is  
 343 not valid at low  $T$  where  $Q(T)$  flattens as a function of  $T$  (as observed for the Higgs mode by Raman). Nor for the  
 344 intensities where the  $T/\omega$  approximation for  $n(\omega)$  is not valid. Second, the theory is for a pure  $E_u$  model. In reality,  
 345 the secondary mode  $A_{2u}$  (corresponding to distortions along the  $\langle 111 \rangle$  trigonal axis orthogonal to  $E_u$  distortions)  
 346 will play some role, and its coupling to  $E_u$  is also an anisotropy term in the Landau energy (it does not exist for  
 347  $I4_122$ )<sup>9</sup>. Finally, the critical exponent near  $T_{s1}$  is the mean field one. In reality, experiment finds  $\beta=1/4$ , not  $1/2$ .  
 348 Despite these caveats, Fig. S10 is remarkably similar to the Raman data, and the XRD data reported in this paper.

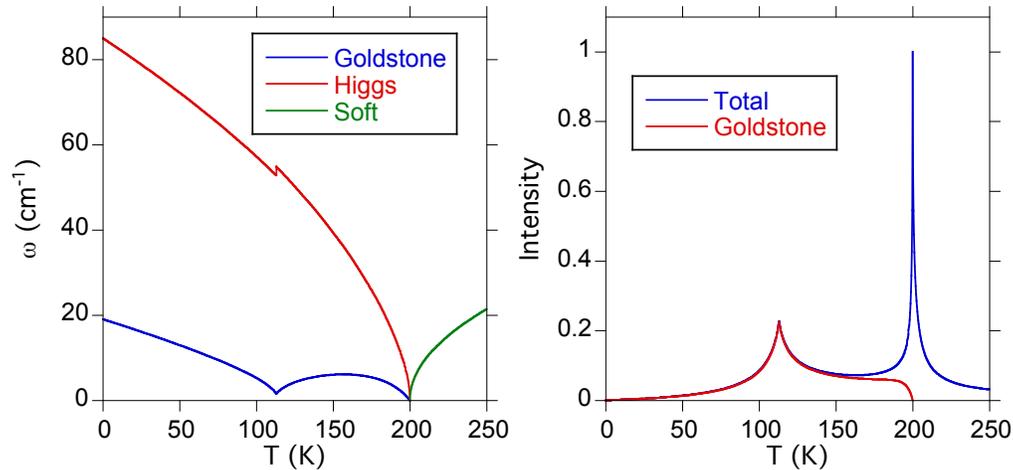


FIG. S10. (a) Landau mode energies as a function of  $T$  for  $\text{Cd}_2\text{Re}_2\text{O}_7$ . Note the first order jump in the Higgs energy and the dip in the Goldstone energy at  $T_{s2}$  (113 K). (b) Landau mode intensities as a function of  $T$ . Outside of the critical region near  $T_{s1}$  (200 K), the intensity is dominated by the Goldstone intensity. Note the resemblance of the calculated intensity to the XRD diffuse scattering intensity presented in this paper (Fig. 4(c)).

349 This brings into question the interpretation of the pump-probe measurements in Ref.<sup>16</sup> which claims that a structural  
 350 soft mode does not exist for  $\text{Cd}_2\text{Re}_2\text{O}_7$ .

- 
- 351 <sup>1</sup> M. J. Krogstad, S. Rosenkranz, J. M. Wozniak, G. Jennings, J. P. C. Ruff, J. T. Vaughey, and R. Osborn, *Nature Materials*  
 352 **19**, 63 (2020), 1902.03318.
- 353 <sup>2</sup> F. J. D. Salvo, D. E. Moncton, and J. V. Waszczak, *Phys. Rev. B* **14**, 4321 (1976).
- 354 <sup>3</sup> J. A. Wilson and A. D. Yoffe, *Adv. Phys.* **18**, 193 (1969).
- 355 <sup>4</sup> K. P. Murphy, *Machine learning : a probabilistic perspective* (MIT Press, Cambridge, Mass. [u.a.], 2013).
- 356 <sup>5</sup> F. J. Di Salvo, D. E. Moncton, and J. V. Waszczak, *Phys. Rev. B* **14**, 4321 (1976).
- 357 <sup>6</sup> A. Ng, CS229 Lecture notes (2017).
- 358 <sup>7</sup> T. M. Gillard, D. Phelan, C. Leighton, and F. S. Bates, *Macromolecules* **48**, 4733 (2015).
- 359 <sup>8</sup> Here we simply averaged peaks due to the volume of the data.
- 360 <sup>9</sup> I. A. Sergienko and S. H. Curnoe, *J. Phys. Soc. Jpn* **72**, 1607 (2003).
- 361 <sup>10</sup> J. C. Petersen, M. D. Caswell, J. S. Dodge, I. A. Sergienko, J. He, R. Jin, and D. Mandrus, *Nature Physics* **2**, 605 (2006).
- 362 <sup>11</sup> J.-i. Yamaura, K. Takeda, Y. Ikeda, N. Hirao, Y. Ohishi, T. C. Kobayashi, and Z. Hiroi, *Phys. Rev. B* **95**, 020102 (2017).
- 363 <sup>12</sup> C. A. Kendziora, I. A. Sergienko, R. Jin, J. He, V. Keppens, B. C. Sales, and D. Mandrus, *Phys. Rev. Lett.* **95**, 125503  
 364 (2005).
- 365 <sup>13</sup> Q. N. Meier, A. Stucky, J. Teyssier, S. M. Griffin, D. van der Marel, and N. A. Spaldin, *Phys. Rev. B* **102**, 014102 (2020).
- 366 <sup>14</sup> P. A. Fleury, *Annual Review of Materials Science* **6**, 157 (1976).
- 367 <sup>15</sup> S. M. Shapiro, J. D. Axe, G. Shirane, and T. Riste, *Phys. Rev. B* **6**, 4332 (1972).
- 368 <sup>16</sup> J. W. Harter, D. M. Kennes, H. Chu, A. de la Torre, Z. Y. Zhao, J.-Q. Yan, D. G. Mandrus, A. J. Millis, and D. Hsieh,  
 369 *Phys. Rev. Lett.* **120**, 047601 (2018).