

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

Eun-Ah Kim (✉ eun-ah.kim@cornell.edu)

Cornell University <https://orcid.org/0000-0002-9554-4443>

Jordan Venderley

Cornell University

Michael Matty

Cornell University

Krishnanand Mallayya

Cornell University

Matthew Krogstad

Argonne National Laboratory <https://orcid.org/0000-0002-8505-5945>

Jacob Ruff

Cornell University

Geoff Pleiss

Cornell University

Varsha Kishore

Cornell University

David Mandrus

Oak Ridge National Laboratory

Daniel Phelan

Argonne National Laboratory

Lekh Poudel

University of Tennessee

Andrew Wilson

NYU <https://orcid.org/0000-0002-2011-3315>

Kilian Weinberger

Cornell University

Puspa Upreti

Argonne National Laboratory

Michael Norman

Argonne National Laboratory <https://orcid.org/0000-0002-9459-078X>

Stephan Rosenkranz

Argonne National Laboratory <https://orcid.org/0000-0002-5659-0383>

Raymond Osborn

Argonne National Laboratory <https://orcid.org/0000-0001-9565-3140>

Physical Sciences - Article

Keywords: X-ray diffraction, machine learning, condensed matter physics

Posted Date: March 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-312345/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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

Jordan Venderley¹, Michael Matty¹, Krishnanand Mallaya¹, Matthew Krogstad², Jacob Ruff³,

Geoff Pleiss⁴, Varsha Kishore⁴, David Mandrus⁵, Daniel Phelan²,

Lekh Poudel^{6,7}, Andrew Gordon Wilson⁸, Kilian Weinberger⁴,

Puspa Upreti^{2,9}, M. R. Norman², Stephan Rosenkranz², Raymond Osborn², Eun-Ah Kim^{1*}

¹Department of Physics, Cornell University

²Materials Science Division, Argonne National Laboratory

³Cornell High Energy Synchrotron Source, Cornell University

⁴Department of Computer Science, Cornell University

⁵Department of Materials Science and Engineering, University of Tennessee

⁶Department of Materials Science and Engineering, University of Maryland

⁷NIST Center for Neutron Research, National Institute of Standard and Technology

⁸Courant Institute of Mathematical Sciences, New York University

⁹Department of Physics, Northern Illinois University

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

(Dated: March 9, 2021)

The information content of crystalline materials becomes astronomical when collective electronic behavior and their fluctuations are taken into account. In the past decade, improvements in source brightness and detector technology at modern x-ray facilities have allowed a dramatically increased fraction of this information to be captured. Now, the primary challenge is to understand and discover scientific principles from big data sets when a comprehensive analysis is beyond human reach. We report the development of a novel unsupervised machine learning approach, *XRD Temperature Clustering (X-TEC)*, that can automatically extract charge density wave (CDW) order parameters and detect intra-unit cell (IUC) ordering and its fluctuations from a series of high-volume X-ray diffraction (XRD) measurements taken at multiple temperatures. We apply *X-TEC* to XRD data on a quasi-skutterudite family of materials, $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$, where a quantum critical point arising from charge order is observed as a function of Ca concentration. We further apply *X-TEC* to XRD data on the pyrochlore metal, $\text{Cd}_2\text{Re}_2\text{O}_7$, to investigate its two much debated structural phase transitions and uncover the Goldstone mode accompanying them. We demonstrate how unprecedented atomic scale knowledge can be gained when human researchers connect the *X-TEC* results to physical principles. Specifically, we extract from the *X-TEC*-revealed selection rule that the Cd and Re displacements are approximately equal in amplitude, but out of phase. This discovery reveals a previously unknown involvement of $5d^2$ Re, supporting the idea of an electronic origin to the structural order. Our approach can radically transform XRD experiments by allowing in-operando data analysis and enabling researchers to refine experiments by discovering interesting regions of phase space on-the-fly.

[1] From the early days of X-ray diffraction (XRD) experiments, they have been used to access atomic scale information in crystalline materials. The primary challenge has always been how to interpret the angle dependent scattering intensities of the resultant diffraction patterns (Fig 1(a)). Bragg and Bragg's initial insights into how to interpret such data [1] enabled the direct determination of crystal structures for the first time, and they were duly awarded a Nobel prize. Since the phase of the X-ray photon is lost in the measurement, the most common approach to interpreting XRD data is to employ forward modeling using the increasingly sophisticated tools of crystallography developed over the past century. These have been remarkably successful in determining the structure of highly crystalline materials, from simple inorganic solids to complex protein crystals. However, subtle structural changes can be difficult to determine when they only result in marginal changes in intensities without any change in peak locations [2]. Furthermore, thermal and quantum fluctuations captured in diffuse scattering away from the Bragg peaks are beyond the reach of conventional crystallographic analysis. The information-rich diffuse scattering is typically weaker than Bragg scattering by several orders of magnitude and can be difficult to differentiate from background noise.

[2] The massive data that modern facilities generate, spanning 3D reciprocal space volumes that include $\mathcal{O}(10^4)$ Brillouin zones (BZ) (Fig 1(a)), at rates of $\mathcal{O}(10^2)$ gigabytes per hour should capture the systematics of such subtle atomic scale information. Yet the sheer quantity of data presents a major challenge. Overcoming this challenge is of paramount importance especially in searching for an unknown order parameter and its fluctuations. Specifically, two types of orders and their fluctuations are targets of XRD (see the illustration for a one-dimensional system in Fig. 1(b-e)): those that change the size of the unit-cell, such as charge density waves (CDW), and those that involve intra-unit cell (IUC) distortions. XRD evidence of CDW order is the emergence of new superlattice peaks, which can be weak and fluctuating, often requiring a targeted search [3, 4]. XRD evidence of IUC order are even subtler changes in structure factors of Bragg peaks [5], unless there are changes in extinction rules. However, the ubiquity of electronic nematic order [6, 7] has turned the study of electronically driven IUC order into an increasingly important scientific objective. Electronically driven IUC order and related 'hidden order' phases typically have profound consequences for the electronic structure as revealed by various probes, yet are often accompanied by subtle structural distortions. Examples range from 3d oxides like cuprates, to 4d and 5d oxides like ruthenates and iridates, to 4f and 5f heavy fermion materials like URu₂Si₂. These small distortions can challenge conventional crystallographic structural refinement that only tracks Bragg peaks and deduce the structural symmetry by fitting all the atomic positions in a forward model. As an example of proposed CDW order, the quasi-skutterudite family, $(\text{Ca}_x\text{Sr}_{1-x})_3\text{X}_4\text{Sn}_{13}$ where X is a transition metal

66 ion like Co, Rh or Ir, exhibits marginal Fermi liquid behavior. Much like in cuprates and heavy fermion materials such
 67 as YbRh_2Si_2 , this order can be suppressed to very low temperatures, leading to a linear in temperature resistivity over
 68 a large range in temperature. In the pyrochlore, $\text{Cd}_2\text{Re}_2\text{O}_7$, a very subtle structural distortion is associated with large
 69 changes in the specific heat and susceptibility. This led Liang Fu [8] to propose the presence of spin nematic order,
 70 and some evidence for this was provided by subsequent non-linear optics measurements [9]. Moreover, the inversion
 71 breaking structural order itself is novel, whose candidate description by an E_u tensor could support pseudo-Goldstone
 72 fluctuations between its two components, $I4_122$ and $I\bar{4}m2$ (see Fig. 1(f)) [10]. Interestingly, both of these examples
 73 exhibit superconductivity at low temperatures, leading to the question of how superconductivity is related to these
 74 orders.

75 [3] To extract atomic scale information encoded in massive XRD data volumes, much needed is a versatile, in-
 76 terpretable, and scalable approach that can reveal order parameters and fluctuations associated with CDW orders
 77 and IUC orders: the vision behind *X-TEC*. For the analysis of complex experimental data, dimension reduction
 78 and machine learning techniques are increasingly employed [11–18], with an emphasis on supervised learning using
 79 hypothesis-driven synthetic data [11–13]. To date, most applications of unsupervised techniques to materials data
 80 have been limited to exploration of compositional phase diagrams of alloys [19–21]. However, an interpretable and
 81 unsupervised approach aiming at discovering interaction driven emergent phenomena in quantum materials such as
 82 order parameters and fluctuations can greatly benefit scientific progress. For versatility, we opted for an unsupervised
 83 approach guided by a fundamental principle of statistical mechanics: the balance between the energy (E) and entropy
 84 (S) resting on the temperature (T). A change in the collective state of a system occurs in the direction of minimizing
 85 the Helmholtz free energy F [22]:

$$F = E - TS. \quad (1)$$

86 When the temperature T is lowered below a certain threshold, the entropy S gives way to the ordered state domi-
 87 nated by the system Hamiltonian. Hence the temperature (T) evolution of the XRD intensity for reciprocal space
 88 point \vec{q} , $I(\vec{q}, T)$, must be qualitatively different if the given reciprocal space point \vec{q} reflects order parameters or their
 89 fluctuations. Tracking the temperature evolution of thousands of Brillouin zones to identify systematic trends and
 90 correlations in any comprehensive manner is impossible to achieve manually without selection bias (see Fig. 2(a)).
 91 *X-TEC* embodies the principle of Eq. 1 by clustering the ‘temperature series’ associated with a given \vec{q} , $I(\vec{q}, T)$, ac-
 92 cording to qualitative features in the temperature dependence, as in high-dimensional clustering approaches that learn

93 qualitative differences in the voice trains for speaker verification [23] (see Fig. 2(b)). *X-TEC* achieves interpretability
 94 and scalability by using a simple Gaussian mixture model (GMM) [24] at its core (see SM section II) and incorporates
 95 correlation among nearby \vec{q} points and within and across BZs using label smoothing (see SM section II-C) similar to
 96 how signals from different cameras can be correlated for computer vision [25] (Fig. 2(c)).

97 [4] The first step in the *X-TEC* pipeline is to preprocess the raw set of temperature series obtained for each
 98 reciprocal space point $\{\vec{q} = (q_x, q_y, q_z)\}$ in $\sim 10^9$ grid points in 3D reciprocal space grid over 10-30 temperatures. The
 99 challenges in working with the raw comprehensive data are in the volume and the dynamic range of the intensity scale
 100 (see Fig. 2(a)). Our preprocessing scheme (SM section II-B) reduces the number of \vec{q} -space points to be canvassed
 101 from the full grid to a selection of points $\{\vec{q}_i\}$ through a systematic thresholding. The trajectories that passed the
 102 thresholding, $\{I(\vec{q}_i, T_j), j = 1, \dots, d^T\}$, are then rescaled so that we can compare trajectories at different intensities
 103 scales, focusing on their temperature dependence rather than the absolute scale. We now cluster the resulting collection
 104 of preprocessed temperature trajectories, $\tilde{\mathbf{I}}(\vec{q}_i) \equiv \{I(\vec{q}_i, T_j); j = 1, \dots, d^T\}$ for each \vec{q}_i to discover qualitatively distinct
 105 types of temperature dependences in the data. There are two modes for clustering depending on whether the label
 106 smoothing is on or off: *X-TEC* smoothen (*X-TECs*) and *X-TEC* detailed (*X-TECd*). For *X-TECs*, we adopt the
 107 label smoothing approach that effectively correlates signals from different cameras for computer vision [25] to correct
 108 the independence assumption and enforce local smoothness across the cluster assignments of points with similar
 109 momenta within and across Brillouin zones. The algorithm first constructs a nearest neighbor graph in momentum
 110 space, connecting reciprocal space points that share similar momenta. For each point, the neighbors are weighted by
 111 their distance in momentum space and the weights normalized. Label smoothing averages the cluster assignments of
 112 a point with its (weighted) neighbors. We incorporate this smoothing step between the E- and M- step of the GMM.
 113 The human researcher makes the choice between *X-TECs* and *X-TECd* and trades between a simpler output and
 114 a simpler (i.e., more scalable) algorithm. *X-TECs* is best suited for detecting order parameters while *X-TECd* can
 115 reveal the nature of fluctuations in high-resolution data. Using *X-TECs* and *X-TECd* in tandem can reveal systematic
 116 correlations between order parameters captured by peak centers and fluctuations captured by diffuse scattering in an
 117 unprecedented manner.

118 [5] Outside of the label smoothing, *X-TEC* uses standard GMM on the temperature series, $\{\tilde{\mathbf{I}}(\vec{q}_i)\}$, treated as a
 119 point in the d^T -dimensional space. Once the human researcher sets the number of clusters K , *X-TEC* attempts to
 120 model each point in the data set $\{\tilde{\mathbf{I}}(\vec{q}_i)\}$ to be independently and identically drawn from a weighted sum of K distinct

121 multivariate normal distributions. The hyper-parameters to be learned are the mixing weights π_k , d^T -dimensional
 122 means \mathbf{m}_k , and $d^T \times d^T$ -dimensional covariances \mathbf{s}_k , $(\pi, \mathbf{m}, \mathbf{s}) \equiv \{(\pi_k, \mathbf{m}_k, \mathbf{s}_k); k = 1, \dots, K\}$. The associated model
 123 log-likelihood is

$$\log p(\{\tilde{\mathbf{I}}(\vec{q}_i)\}|\pi, \mathbf{m}, \mathbf{s}) = \sum_{\vec{q}_i} \log \left[\sum_{k=1}^K \pi_k \mathcal{N}(\tilde{\mathbf{I}}(\vec{q}_i)|\mathbf{m}_k, \mathbf{s}_k) \right]. \quad (2)$$

124 Here, $\mathcal{N}(\tilde{\mathbf{I}}(\vec{q}_i)|\mathbf{m}_k, \mathbf{s}_k)$ is the probability density for the k^{th} multivariate Gaussian with mean \mathbf{m}_k and covariance \mathbf{s}_k
 125 evaluated at $\tilde{\mathbf{I}}(\vec{q}_i)$, i.e.,

$$\mathcal{N}(\tilde{\mathbf{I}}(\vec{q}_i)|\mathbf{m}_k, \mathbf{s}_k) \equiv \frac{1}{(2\pi)^{d_T/2}} \frac{1}{\sqrt{\det \mathbf{s}_k}} e^{-\frac{1}{2} [\tilde{\mathbf{I}}(\vec{q}_i - \mathbf{m}_k)^\dagger \mathbf{s}_k^{-1} (\tilde{\mathbf{I}}(\vec{q}_i) - \mu_k)]}. \quad (3)$$

126 The probability, w_i^k , that the temperature series labeled by \vec{q}_i belongs to the k^{th} cluster is

$$w_i^k = \frac{\pi_k \mathcal{N}(\tilde{\mathbf{I}}(\vec{q}_i)|\mathbf{m}_k, \mathbf{s}_k)}{\sum_k \pi_k \mathcal{N}(\tilde{\mathbf{I}}(\vec{q}_i)|\mathbf{m}_k, \mathbf{s}_k)}, \quad (4)$$

127 according to Bayes' theorem (see SM section IIc). *X-TEC* learns the hyper-parameters $(\pi, \mathbf{m}, \mathbf{s})$ using a stepwise
 128 expectation maximization (EM) algorithm [26]. Much like mean-field theory familiar to physicists, the EM algorithm
 129 iteratively searches for the saddle point of the lower bound of the log-likelihood

$$\tilde{\ell}(\{w_i^k, \pi_k, \mathbf{m}_k, \mathbf{s}_k\}) = \sum_{i,k} w_i^k \log \left[\frac{\pi_k \mathcal{N}(\tilde{\mathbf{I}}(\vec{q}_i)|\mathbf{m}_k, \mathbf{s}_k)}{w_i^k} \right] + \lambda(1 - \sum_k \pi_k), \quad (5)$$

130 where λ is a Lagrange multiplier. The cluster assignment of a given reciprocal space point \vec{q}_i is then determined by
 131 the converged value of the clustering expectation $\arg \max_k \{w_i^k\}$.

132 [6] We first employ *X-TECs* to target a putative CDW quantum critical point and illustrate *X-TECs* in action.
 133 Electrical resistivity and heat capacity experiments on the quasi-skutterudite family $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$ indicate a
 134 quantum critical point at a composition of $x = 0.9$ under ambient pressure (see Fig. 2(g)) [27], driven by the volume
 135 change of replacing the larger Sr ion by the smaller Ca one, with superconductivity emerging at low temperatures.
 136 Given that this is associated with a linear in temperature resistivity as in the cuprates, the question has arisen about the
 137 nature of this order, the quantum critical fluctuations associated with it, and their connection to the superconductivity.
 138 Although CDW order was proposed [27], this has never been proven, so we use *X-TECs* to investigate this. The x-
 139 ray measurements on $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$ were taken on Sector 6-ID-D at the Advanced Photon Source using a
 140 monochromatic x-ray energy of 87 keV. Images are collected on a fast area detector (Pilatus 2M CdTe) at a frame
 141 rate of 10 Hz while the sample is continuously rotated through 360° at a speed of 1° per second (Fig. 1a). These

142 rotation scans are repeated twice to fill in gaps between the detector chips, so a single measurement represents an
 143 uncompressed data volume of over 100 GB collected in 20 minutes. This allows comprehensive measurements of
 144 the temperature dependence of a material in much less than a day. Using a cryostream, we are able to vary the
 145 temperature from 30 K to 300 K. The rotation scans sweep through a large volume of reciprocal space (Fig. 1a);
 146 when the data are transformed into reciprocal space coordinates, the 3D arrays are typically reduced in size by an
 147 order of magnitude. More details of both the measurement and data reduction workflow are given in Ref. [28], see
 148 also SM I. In the past, we would have analyzed such data by selecting a few superlattice peaks, with the assumption
 149 that they are representative of the whole, and fitting their temperature dependence. This may be justified in many
 150 cases, but in doing so, we would be ignoring over 99% of the data, limiting the statistical precision available from such
 151 comprehensive data sets and potentially missing other components of the order parameter. Here we apply *X-TECs* to
 152 around 200 GB of XRD data on four compounds, ($x = 0, 0.1, 0.6, 0.65$) and map out the phase diagram as a function
 153 of temperature and doping with no prior knowledge regarding the order parameter given to *X-TECs*.

154 [7] *X-TECs* extracts order parameter clusters from the entire 200 GB of XRD data for the four compounds within
 155 minutes. In Figs. 2(d) we present cluster means and variances of the two-cluster ($K = 2$) clustering results for
 156 undoped $\text{Sr}_3\text{Rh}_4\text{Sn}_{13}$. The temperature dependence of the learned means of the yellow cluster and the blue cluster
 157 makes it evident that the yellow cluster represents the order parameter and the temperature at which it crashes
 158 down is the critical temperature: $T_c \approx 130$ K. The clustering results can be interpreted by locating the cluster
 159 assignments in reciprocal space, as shown in Fig. 2(e). The location of the yellow cluster identifies the ordering
 160 wavevector to be $q_{CDW} = (0.5, 0.5, 0)$ and symmetry equivalents with respect to the cubic Bragg peaks, without
 161 any prior knowledge. Label smoothing keeps the clustering output to be smoothly connected in the vicinity of each
 162 peak, simplifying interpretation. Plotting the CDW order parameters extracted at each doping, we can track the
 163 evolution of the critical temperature T_c as a function of chemical pressure (Fig. 2(f)) and obtain the full quantum
 164 phase diagram. The doping-dependent T_c obtained using *X-TEC* allows us to map out the quantum phase diagram
 165 associated with the CDW ordering, much as neutron scattering has been used to obtain quantum phase diagrams
 166 associated with spin order. Earlier studies of this family of compounds identified quantum critical behaviour using
 167 thermodynamic, transport, and phonon measurements [29, 30], but this is the first to be determined directly from the
 168 CDW order parameter and shows the efficiency with which X-TEC analysis can extract structural phase diagrams.
 169 The suppression of the CDW order upon doping and the emergence of superconductivity from a strange metal with

170 linear T resistivity is reminiscent of the cuprates [31].

171 [8] We now employ X -TECs and X -TECd in tandem to study hidden IUC order and order parameter fluctuations
 172 in the pyrochlore metal $\text{Cd}_2\text{Re}_2\text{O}_7$ [32–34] (see Fig. 3(a)), where the nature of its two E_u -symmetry structural
 173 transitions have recently regained interest [35] after the discovery of a purported T_{2u} electronic order from second
 174 harmonic generation (SHG) [9]. $\text{Cd}_2\text{Re}_2\text{O}_7$ goes through a second-order transition at $T_{s1} = 200$ K with a large
 175 thermodynamic signature in the specific heat (Fig. 3(b)) from the cubic pyrochlore $Fd\bar{3}m$ structure (phase I) to a
 176 structure that breaks inversion symmetry (phase II). Most studies conclude that the space group of the phase II is
 177 $I\bar{4}m2$ [35], but this is now questioned in light of the SHG data [9, 36, 37], which also reveal the surprising fact that
 178 the E_u structural order (which it also sees) does not have the expected temperature dependence of a primary order
 179 parameter (unlike the T_{2u} signal, which does). At lower temperature, a first-order transition at $T_{s2} = 113$ K (phase
 180 III) has been observed, and is proposed to arise from the other component of E_u which is the $I4_122$ space group
 181 [35]. Again, this is controversial, in that earlier SHG data [38] do not see the expected rotation of the signal that
 182 should accompany such a phase transition. Moreover, recent Raman data [39] see line splittings consistent with a
 183 lowering to orthorhombic symmetry below about 80 K [39] which was speculated to be due to an $F222$ space group. A
 184 combination of small atomic displacements with crystallographic twinning [40] has made it challenging to determine
 185 the true structure of these low symmetry states using traditional crystallographic approaches [41, 42]. Still, previous
 186 results for phase II are consistent with the above picture, where $I\bar{4}m2$ and $I4_122$ are the two components of the E_u
 187 order parameter, a rank-2 tensor. The degeneracy between these two states is lifted at sixth order in Landau theory
 188 [43], resulting in a quasi-Goldstone mode encoding fluctuations between the two phases [44, 45] (see Fig. 1(f)). Raman
 189 scattering [10] sees a strong central peak that appears to be the Goldstone mode, along with a higher frequency mode
 190 which appears to be the Higgs mode, though this has been recently questioned based on pump-probe measurements
 191 [46]. The uniqueness of this situation is that although pseudo-Goldstone modes have been seen in other materials,
 192 notably ferroelectrics, they typically exist at much higher frequencies [45]. The fact that this is not the case for
 193 $\text{Cd}_2\text{Re}_2\text{O}_7$ indicates that the anisotropy in the Landau free energy is anomalously small. Confirmation of such low
 194 frequency fluctuations has been beyond the reach of XRD, as has been the relation of the E_u structural order to the
 195 proposed T_{2u} ‘hidden order’ indicated by the SHG data.

196 [9] We performed x-ray scattering measurements over a wide temperature range ($30 \text{ K} < T < 300 \text{ K}$) on a single
 197 crystal of $\text{Cd}_2\text{Re}_2\text{O}_7$, which our measurements show is untwinned, at least in phase II. This may be due to the

198 small volume ($400 \times 200 \times 50 \mu\text{m}^3$) required for our synchrotron measurements. We first performed scans using an x-ray
 199 energy of 87 keV, which contained scattering spanning nearly 15,000 Brillouin zones, in order to search for previously
 200 undetected peaks and determine the systematic (HKL) dependence of the Bragg peak intensities at each temperature
 201 (see SM section III-A). To better understand the order parameter fluctuations, we then reduced the energy to 60 keV
 202 to improve the \vec{Q} -resolution and increased the number of temperatures, particularly near the phase transitions. We
 203 comprehensively analyzed the resulting data sets with a combined volume of nearly 8 TB using $X\text{-TECs}$ and $X\text{-TECd}$.

204 [10] We illustrate the sharp characteristics of the order parameter and its fluctuations by focusing on the cubic-
 205 forbidden peaks in Figures 3 and 4 (see SM III-B for the clustering results that selects cubic-forbidden peaks as the
 206 order parameter of phase II). Fig. 3(c) shows the $K = 2$ clustering means of $X\text{-TECs}$ and $K = 3$ clustering means of
 207 $X\text{-TECd}$ on the cubic-forbidden peaks over the temperature range of [30 K, 150 K] [47]. Both outcomes presented big
 208 surprises. First, the $X\text{-TECs}$ outcome separated the cubic forbidden peaks that behave like the order parameter of
 209 phase II into two subgroups: one that quickly flattens in phase II to abruptly rise in phase III (yellow) and the other
 210 that continues to rise in phase II to abruptly drop in phase III (green). Second, $X\text{-TECd}$ clustering separates out
 211 the diffuse regions associated with each of the subgroups of cubic-forbidden peaks to define their own clusters with
 212 temperature dependencies that are qualitatively different (red and blue in Fig. 3(c)) and distinct from the temperature
 213 dependencies of the peak centers. The reciprocal space distribution of the clusters reveals precise selection rules and
 214 tight correlation between the order parameter tracked in $X\text{-TECs}$ and the fluctuations revealed in $X\text{-TECd}$. Due to
 215 the orders of magnitude differences in intensity scales, $X\text{-TECs}$ is dominated by the peak centers. $X\text{-TECd}$ separated
 216 out the peak centers from the halos of diffuse regions. Combining the two results, we present the $X\text{-TECs}$ outcome
 217 through the color of the peak centers detected in $X\text{-TECd}$. The (HKL) assignments of the two subgroups in $X\text{-TECs}$,
 218 and their associated diffuse halos in $X\text{-TECd}$ (Fig. 3(d)) reveal strict selection rules. Yellow peaks (with red halos)
 219 are of the form $(4n_1, 4n_2, 4n_3 + 2)$, while green peaks (with blue halos) have $(4n_1 + 2, 4n_2, 4n_3)$ or $(4n_1, 4n_2 + 2, 4n_3)$,
 220 in the cubic indices of phase I. The mean intensity trajectories of red and blue clusters in Fig. 3(c) indicate that
 221 the red halo sustains intensity throughout phase II to only dive down at $T_{s2} = 113$ K while the blue halo picks up
 222 intensity at around T_{s2} to abruptly die out at around 90 K. The temperature evolution of representative line cuts
 223 shown in Fig. 3(e-f) confirms these observations in the raw data.

224 [11] The systematics in the temperature dependencies of different cubic-forbidden peaks and their diffuse halos
 225 revealed using the two modes of $X\text{-TEC}$ on the entire 8 TB of data present an unprecedented opportunity to extract

226 atomic scale clues regarding the hidden order. First, we can extract an order parameter critical exponent associated
 227 with the structural transition that is reflecting the entire data set from the X - TEC s mean trajectories. Fig. 4(a)
 228 shows the temperature dependence of the two peak averaged clusters (yellow and green) of cubic-forbidden peaks and
 229 their fits, in which we treat the displacements as order parameters with a common exponent β (see SM III-D). Both
 230 clusters fit to the common exponent of $\beta \approx 0.25$ close to T_{s1} . This is close to the value expected for a 2D-XY system
 231 [48]. This is a surprise in that the E_u signal observed by SHG scales linearly in $T_{s1} - T$ which is 4β instead of the
 232 expected 2β indicated by theory [36], whereas it is the T_{2u} signal that scales like 2β . Second, we can convert the
 233 selection rule revealed by X - TEC into atomic distortions. The selection rule shows that the two clusters correspond
 234 to two distinct classes of structure factor, whose values only depend on the distortions of the Cd and Re sublattices:
 235 the yellow cluster consists of peaks that are dominated by z -axis displacements $(\delta z_{Cd}, \delta z_{Re})$, and those in the green
 236 cluster by in-plane displacements, along x or y depending on the Wyckoff position, $(\delta x_{Cd}, \delta x_{Re})$ (SM III-C) (see
 237 Fig. 4(b)). The flat temperature dependence of the yellow cluster below 180 K results from out-of-phase distortions of
 238 the Cd and Re sublattices. The refined values of $(\delta z_{Cd}$ and $\delta z_{Re})$ are approximately equal and opposite (see Fig. 4(b)).
 239 This is another surprising result. Previous refinements [49] indicate that the Re displacements are small, and this
 240 is consistent with a density functional theory study [39]. Small Re displacements are expected if the $5d$ electrons in
 241 Re play a passive role in the structural transition as the Re are in an almost ideally bonded octahedral environment,
 242 compared to Cd which is underbonded because of its two short Cd-O and six long Cd-O bonds. Therefore, a large
 243 displacement of Re implies that this is a consequence of the $5d^2$ configuration of Re being unstable to spin nematic
 244 order that should lead to valence bond ordering (different Re-Re bonds, as illustrated in Fig. 1(f)) in a given Re
 245 tetrahedron as proposed in other pyrochlores [50]. Third, the connection between the two diffuse halo clusters (red
 246 and blue) and the selection rule for the peak centers draws us to the unusual and distinct temperature dependence of
 247 the diffuse regions (see Fig. 4(c)). Strong critical scattering at T_{s1} is clear in both clusters, but the diffuse contribution
 248 is much stronger in the red halo throughout phase II. The role between the two halos reverses at T_{s2} . We attribute
 249 the fluctuations reflected in the sustained intensity of the red halo to the Goldstone mode manifest through strong
 250 z -axis fluctuations.

251 [12] To investigate this further, we turn to a description of the various modes (see SM III-F for more details of the
 252 calculations). Above T_{s1} one has a soft mode whose energy should go to zero at T_{s1} . Below this, the soft mode splits
 253 into a Higgs mode (fluctuations in the amplitude of the E_u order) and a Goldstone mode (fluctuations in the phase,

254 that is fluctuations between $I\bar{4}m2$ and $I4_122$). The latter would be at zero energy if there were no anisotropy. In
 255 Landau theory, the first anisotropy term appears at sixth order and the next one at eighth order in the free energy.
 256 These two must be of opposite sign in order to have a second transition at T_{s2} [43]. Their difference changes sign at
 257 T_{s2} . The net result is that one has a Goldstone mode that starts at zero energy at T_{s1} , rises slightly with lowering T ,
 258 then dips down again at T_{s2} , and then rises again below this. This can be appreciated by the intensities associated
 259 with the various modes (see Fig. 4(d)), noting that the Goldstone mode's coupling to the x-rays is quadratic in the
 260 E_u order parameter [51] reflecting the fact that it does not exist above T_{s1} (the analog of the soft mode below T_{s1}
 261 is the Higgs mode). From the calculated intensities, one sees that the Goldstone mode completely dominates outside
 262 of the critical region near T_{s1} . The calculated behavior is remarkably similar to the XRD data (Fig. 4(c)), with a
 263 pronounced cusp at T_{s2} . This is strong indication that the diffuse scattering is indeed due to structural fluctuations
 264 associated with the Goldstone mode.

265 [13] In summary, we developed *X-TEC*, an unsupervised and interpretable ML algorithm for voluminous XRD
 266 data that is guided by the fundamental role temperature plays in emergent phenomena. By analyzing the entire
 267 data set over many BZs and making use of temperature evolutions, *X-TEC* can pick up subtle features representing
 268 both order parameters and fluctuations from higher intensity backgrounds. The two modes, *X-TEC*s and *X-TEC*d,
 269 allow for discovery of systematics in order parameters and its fluctuations despite orders of magnitude differences in
 270 intensities. The algorithm is fast with $O(10)$ minutes of run time for the tasks presented here. Using *X-TEC*, we
 271 discovered that the superconductor family $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$ exhibits CDW order and we mapped out its phase
 272 diagram. In $\text{Cd}_2\text{Re}_2\text{O}_7$, we conclusively identified the primary order parameter of the $T_{s1} = 200$ K transition. We
 273 further revealed the nature of the intra-unit-cell atomic distortions in a way that has eluded crystallographic analysis
 274 until now. Finally, we revealed XRD evidence of a structural Goldstone mode for the first time. The unprecedented
 275 degree of microscopic information we have been able to unearth from the XRD is fitting for such comprehensive data
 276 but would have been impossible by manual inspection. Instead of determining critical exponents by fitting a handful
 277 of peaks, *X-TEC* provides a means of including the entire data volume by clustering peak intensities from thousands
 278 of Brillouin zones to produce an analysis that is both robust and rapid in future studies of such phase diagrams.
 279 Given the general structure of *X-TEC*, we anticipate it to be broadly applicable to other fields beyond XRD.

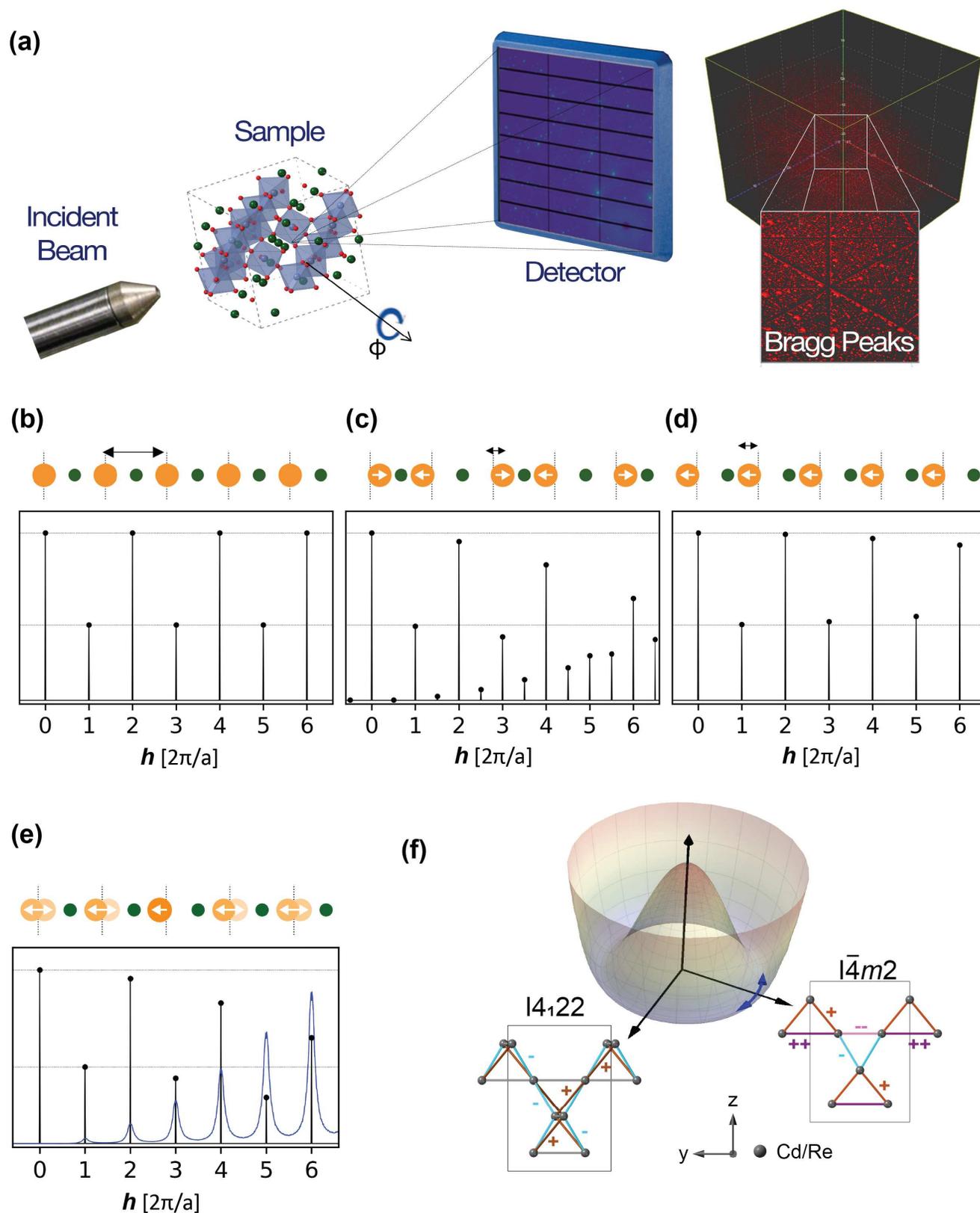
- [1] William Henry Bragg and William Lawrence Bragg, “The reflection of x-rays by crystals,” *Proc. Roy. Soc. London. A* **88**, 428–438 (1913), <https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1913.0040>.
- [2] Takeshi Egami and Simon JL Billinge, *Underneath the Bragg peaks: structural analysis of complex materials* (Newnes, 2012).
- [3] P. Abbamonte, A. Rusydi, S. Smadici, G. D. Gu, G. A. Sawatzky, and D. L. Feng, “Spatially modulated ‘Mottness’ in $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$,” *Nature Physics* **1**, 155–158 (2005).
- [4] E. M. Forgan, E. Blackburn, A. T. Holmes, A. K. R. Briffa, J. Chang, L. Bouchenoire, S. D. Brown, Ruixing Liang, D. Bonn, W. N. Hardy, N. B. Christensen, M. V. Zimmermann, M. Hücker, and S. M. Hayden, “The microscopic structure of charge density waves in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6.54}$ revealed by x-ray diffraction,” *Nature Communications* **6**, 10064 (2015).
- [5] M. J. Lawler, K. Fujita, Jinhwan Lee, A. R. Schmidt, Y. Kohsaka, Chung Koo Kim, H. Eisaki, S. Uchida, J. C. Davis, J. P. Sethna, and Eun-Ah Kim, “Intra-unit-cell electronic nematicity of the high- T_c copper-oxide pseudogap states,” *Nature* **466**, 347–351 (2010).
- [6] Eduardo Fradkin, Steven A. Kivelson, Michael J. Lawler, James P. Eisenstein, and Andrew P. Mackenzie, “Nematic fermi fluids in condensed matter physics,” *Ann. Rev. Cond. Matt.* **1**, 153–178 (2010), <https://doi.org/10.1146/annurev-conmatphys-070909-103925>.
- [7] Rafael M. Fernandes, Peter P. Orth, and Jörg Schmalian, “Intertwined vestigial order in quantum materials: Nematicity and beyond,” *Ann. Rev. Cond. Matt.* **10**, 133–154 (2019), <https://doi.org/10.1146/annurev-conmatphys-031218-013200>.
- [8] Liang Fu, “Parity-Breaking Phases of Spin-Orbit-Coupled Metals with Gyrotropic, Ferroelectric, and Multipolar Orders,” *Phys. Rev. Lett.* **115**, 026401 (2015).
- [9] J. W. Harter, Z. Y. Zhao, J.-Q. Yan, D. G. Mandrus, and D. Hsieh, “A parity-breaking electronic nematic phase transition in the spin-orbit coupled metal $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Science* **356**, 295–299 (2017), <https://science.sciencemag.org/content/356/6335/295.full.pdf>.
- [10] C. A. Kendziora, I. A. Sergienko, R. Jin, J. He, V. Keppens, B. C. Sales, and D. Mandrus, “Goldstone-mode phonon dynamics in the pyrochlore $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Phys. Rev. Lett.* **95**, 125503 (2005).
- [11] Yi Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch’ng, H. Eisaki, S. Uchida, J. C. Séamus Davis, Ehsan Khatami, and Eun-Ah Kim, “Machine learning in electronic-quantum-matter imaging experiments,” *Nature* **570**, 484–490 (2019).
- [12] Annabelle Bohrdt, Christie S. Chiu, Geoffrey Ji, Muqing Xu, Daniel Greif, Markus Greiner, Eugene Demler, Fabian Grusdt, and Michael Knap, “Classifying snapshots of the doped hubbard model with machine learning,” *Nature Physics* **15**, 921–924 (2019).
- [13] Sayak Ghosh, Michael Matty, Ryan Baumbach, Eric D. Bauer, K. A. Modic, Arkady Shekhter, J. A. Mydosh, Eun-Ah Kim, and B. J. Ramshaw, “One-component order parameter in URu_2Si_2 uncovered by resonant ultrasound spectroscopy and machine learning,” *Science Advances* **6** (2020), 10.1126/sciadv.aaz4074, <https://advances.sciencemag.org/content/6/10/eaaz4074.full.pdf>.
- [14] Giacomo Torlai, Brian Timar, Evert P. L. van Nieuwenburg, Harry Levine, Ahmed Omran, Alexander Keesling, Hannes Bernien, Markus Greiner, Vladan Vuletić, Mikhail D. Lukin, Roger G. Melko, and Manuel Endres, “Integrating neural networks with a quantum simulator for state reconstruction,” *Phys. Rev. Lett.* **123**, 230504 (2019).
- [15] P. Ronhovde, S. Chakrabarty, D. Hu, M. Sahu, K. K. Sahu, K. F. Kelton, N. A. Mauro, and Z. Nussinov, “Detection of hidden structures for arbitrary scales in complex physical systems,” *Scientific Reports* **2**, 329 (2012).
- [16] Maxim Ziatdinov, Dohyung Kim, Sabine Neumayer, Rama K. Vasudevan, Liam Collins, Stephen Jesse, Mahshid Ahmadi, and Sergei V. Kalinin, “Imaging mechanism for hyperspectral scanning probe microscopy via gaussian process modelling,” *npj Computational Materials* **6**, 21 (2020).
- [17] Harry S. Geddes, Helen Blade, James F. McCabe, Leslie P. Hughes, and Andrew L. Goodwin, “Structural characterisation of amorphous solid dispersions via metropolis matrix factorisation of pair distribution function data,” *Chem. Commun.* **55**, 13346–13349 (2019).
- [18] Christopher J. Wright and Xiao-Dong Zhou, “Computer-assisted area detector masking,” *Journal of Synchrotron Radiation* **24**, 506–508 (2017).
- [19] C. J. Long, D. Bunker, X. Li, V. L. Karen, and I. Takeuchi, “Rapid identification of structural phases in combinatorial thin-film libraries using x-ray diffraction and non-negative matrix factorization,” *Review of Scientific Instruments* **80**, 103902 (2009).
- [20] Valentin Stanev, Velimir V Vesselinov, A Gilad Kusne, Graham Antoszewski, Ichiro Takeuchi, and Boian S Alexandrov, “Unsupervised phase mapping of x-ray diffraction data by nonnegative matrix factorization integrated with custom clustering,” *npj Computational Materials* **4**, 1–10 (2018).
- [21] Zhantao Chen, Nina Andrejevic, Nathan Drucker, Thanh Nguyen, R Patrick Xian, Tess Smidt, Yao Wang, Ralph Ernstorfer, Alan Tennant, Maria Chan, *et al.*, “Machine learning on neutron and x-ray scattering,” arXiv preprint arXiv:2102.03024 (2021).
- [22] R.K. Pathria and P.D. Beale, *Statistical Mechanics* (Elsevier Science, 2011).
- [23] Douglas A Reynolds, Thomas F Quatieri, and Robert B Dunn, “Speaker verification using adapted gaussian mixture models,” *Digital signal processing* **10**, 19–41 (2000).
- [24] Kevin P. Murphy, *Machine learning : a probabilistic perspective* (MIT Press, Cambridge, Mass. [u.a.], 2013).
- [25] Yurong You, Yan Wang, Wei-Lun Chao, Divyansh Garg, Geoff Pleiss, Bharath Hariharan, Mark E. Campbell, and Kil-

- ian Q. Weinberger, “Pseudo-lidar++: Accurate depth for 3d object detection in autonomous driving,” in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020* (OpenReview.net, 2020).
- [26] Percy Liang and Dan Klein, “Online em for unsupervised models,” in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL ’09* (Association for Computational Linguistics, USA, 2009) pp. 611–619.
- [27] Lina E. Klintberg, Swee K. Goh, Patricia L. Alireza, Paul J. Saines, David A. Tompsett, Peter W. Logg, Jinhu Yang, Bin Chen, Kazuyoshi Yoshimura, and F. Malte Grosche, “Pressure- and composition-induced structural quantum phase transition in the cubic superconductor $(\text{Sr}, \text{Ca})_3\text{Ir}_4\text{Sn}_{13}$,” *Phys. Rev. Lett.* **109**, 237008 (2012).
- [28] Matthew J Krogstad, Stephan Rosenkranz, Justin M Wozniak, Guy Jennings, Jacob P C Ruff, John T Vaughey, and Raymond Osborn, “Reciprocal Space Imaging of Ionic Correlations in Intercalation Compounds,” *Nature Materials* **19**, 63–68 (2020), 1902.03318.
- [29] S K Goh, D A Tompsett, P J Saines, H C Chang, T Matsumoto, M Imai, K Yoshimura, and F M Grosche, “Ambient Pressure Structural Quantum Critical Point in the Phase Diagram of $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$,” *Phys. Rev. Lett.* **114**, 097002 (2015).
- [30] Y. W. Cheung, Y. J. Hu, M. Imai, Y. Tanioku, H. Kanagawa, J. Murakawa, K. Moriyama, W. Zhang, K. T. Lai, K. Yoshimura, F. M. Grosche, K. Kaneko, S. Tsutsui, and Swee K. Goh, “Evidence of a structural quantum critical point in $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$ from a lattice dynamics study,” *Phys. Rev. B* **98**, 161103 (2018).
- [31] Bernhard Keimer, Steven A Kivelson, Michael R Norman, Shinichi Uchida, and J Zaanen, “From quantum matter to high-temperature superconductivity in copper oxides,” *Nature* **518**, 179–186 (2015).
- [32] R Jin, J He, S McCall, C S Alexander, F Drymiotis, and D Mandrus, “Superconductivity in the correlated pyrochlore $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Phys. Rev. B* **64**, 180503 (2001), cond-mat/0107309.
- [33] M Hanawa, Y Muraoka, T Tayama, T Sakakibara, J Yamaura, and Z Hiroi, “Superconductivity at 1 K in $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Phys. Rev. Lett.* **87**, 187001 (2001), cond-mat/0106421.
- [34] Hironori Sakai, Kazuyoshi Yoshimura, Hiroyuki Ohno, Harukazu Kato, Shinsaku Kambe, Russell E Walstedt, Tatsuma D Matsuda, Yoshinori Haga, and Yoshichika Onuki, “Superconductivity in a pyrochlore oxide, $\text{Cd}_2\text{Re}_2\text{O}_7$,” *J. Phys. Cond. Matt.* **13**, L785–L790 (2001).
- [35] Zenji Hiroi, Jun-ichi Yamaura, Tatsuo C. Kobayashi, Yasuhito Matsubayashi, and Daigorou Hirai, “Pyrochlore oxide superconductor $\text{Cd}_2\text{Re}_2\text{O}_7$ revisited,” *J. Phys. Soc. Jpn* **87**, 024702 (2018), <https://doi.org/10.7566/JPSJ.87.024702>.
- [36] S. Di Matteo and M. R. Norman, “Nature of the tensor order in $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Phys. Rev. B* **96**, 115156 (2017).
- [37] M R Norman, “Crystal structure of the inversion-breaking metal $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Phys. Rev. B* **101**, 045117 (2020).
- [38] Jesse C. Petersen, Michael D. Caswell, J. Steven Dodge, Ivan A. Sergienko, Jian He, Rongying Jin, and David Mandrus, “Nonlinear optical signatures of the tensor order in $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Nature Physics* **2**, 605–608 (2006).
- [39] Konrad J. Kapcia, Maureen Reedyk, Mojtaba Hajialamdari, Andrzej Ptok, Przemysław Piekarczyk, Armin Schulz, Fereidoon S. Razavi, Reinhard K. Kremer, and Andrzej M. Oleś, “Discovery of a low-temperature orthorhombic phase of the $\text{Cd}_2\text{Re}_2\text{O}_7$ superconductor,” *Phys. Rev. Research* **2**, 033108 (2020).
- [40] John-Paul Castellan, B D Gaulin, J van Duijn, M J Lewis, Mark D Lumsden, R Jin, J He, Stephen E Nagler, and David Mandrus, “Structural ordering and symmetry breaking in $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Phys. Rev. B* **66**, 134528 (2002).
- [41] Jun-Ichi Yamaura and Zenji Hiroi, “Low Temperature Symmetry of Pyrochlore Oxide $\text{Cd}_2\text{Re}_2\text{O}_7$,” *J. Phys. Soc. Jpn* **71**, 2598 – 2600 (2002).
- [42] Jun-ichi Yamaura, Keiki Takeda, Yoichi Ikeda, Naohisa Hirao, Yasuo Ohishi, Tatsuo C. Kobayashi, and Zenji Hiroi, “Successive spatial symmetry breaking under high pressure in the spin-orbit-coupled metal $\text{Cd}_2\text{Re}_2\text{O}_7$,” *Phys. Rev. B* **95**, 020102 (2017).
- [43] Ivan A. Sergienko and Stephanie H. Curnoe, “Structural Order Parameter in the Pyrochlore Superconductor $\text{Cd}_2\text{Re}_2\text{O}_7$,” *J. Phys. Soc. Jpn* **72**, 1607 – 1610 (2003).
- [44] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg, “Broken symmetries,” *Phys. Rev.* **127**, 965–970 (1962).
- [45] Quintin N. Meier, Adrien Stucky, Jeremie Teyssier, Sinéad M. Griffin, Dirk van der Marel, and Nicola A. Spaldin, “Manifestation of structural higgs and goldstone modes in the hexagonal manganites,” *Phys. Rev. B* **102**, 014102 (2020).
- [46] 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, “Evidence of an improper displacive phase transition in $\text{Cd}_2\text{Re}_2\text{O}_7$ via time-resolved coherent phonon spectroscopy,” *Phys. Rev. Lett.* **120**, 047601 (2018).
- [47] For each X -TEC clustering we increase K until there is no gain in information.
- [48] Steven Bramwell and P C W Holdsworth, “Magnetization and universal sub-critical behaviour in two-dimensional XY magnets,” *Journal of Physics: Condensed Matter* **5**, L53 – L59 (1993).
- [49] Mark T Weller, Robert W Hughes, Joanna Rooke, Christopher S Knee, and Jon Reading, “The pyrochlore family—a potential panacea for the frustrated perovskite chemist,” *Dalton transactions*, 3032–3041 (2004).
- [50] Oleg Tchernyshyov, R. Moessner, and S. L. Sondhi, “Spin-peierls phases in pyrochlore antiferromagnets,” *Phys. Rev. B* **66**, 064403 (2002).
- [51] P A Fleury, “The effects of soft modes on the structure and properties of materials,” *Annual Review of Materials Science* **6**, 157–180 (1976).
- [52] Andrew Ng, “Cs229 lecture notes,” CS229 Lecture notes (2017).

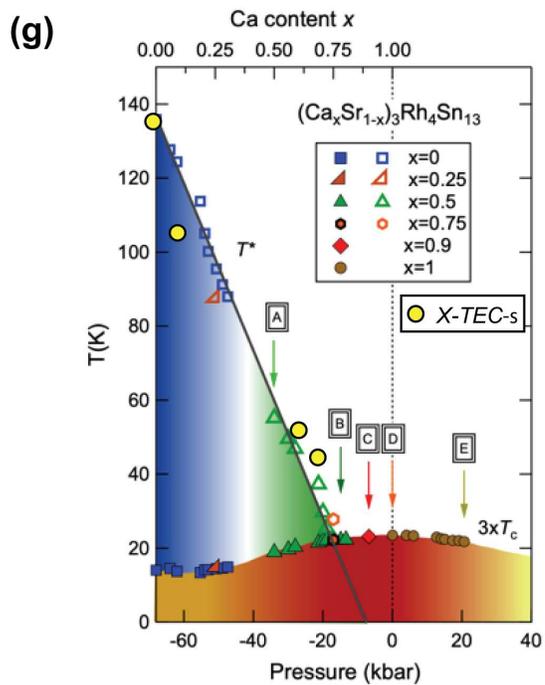
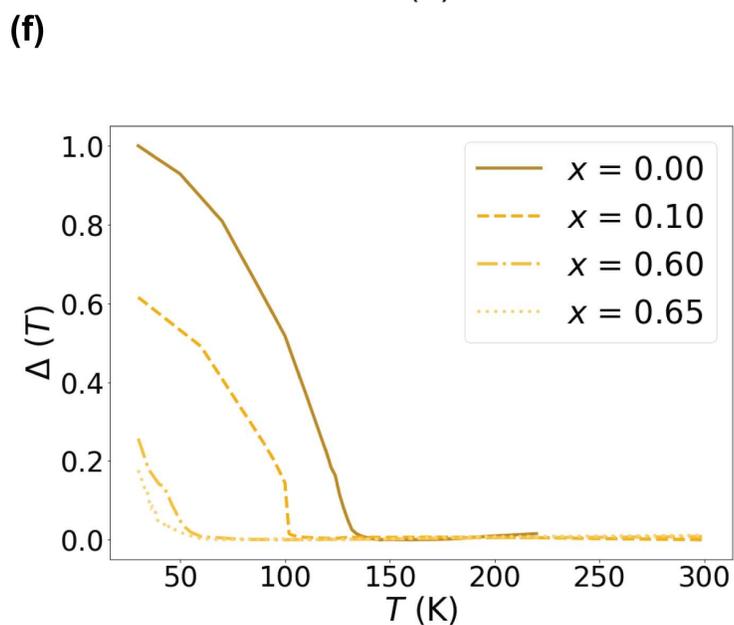
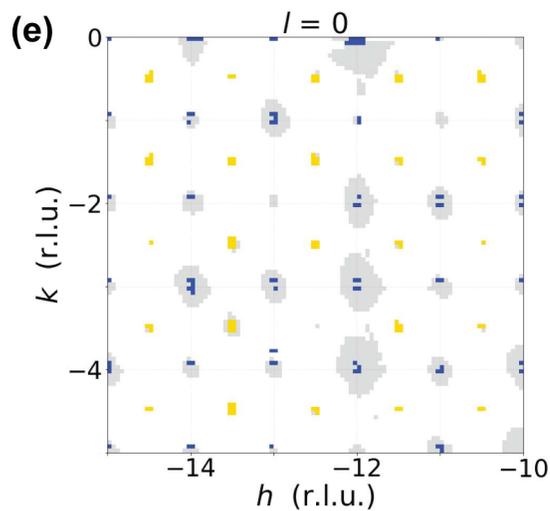
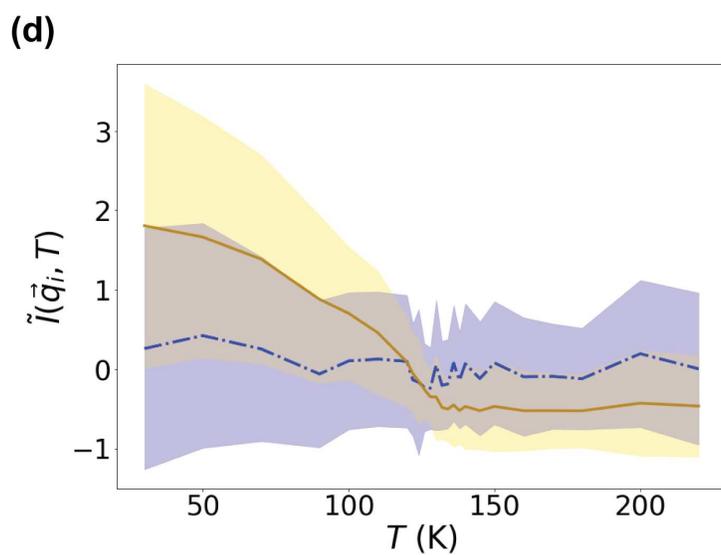
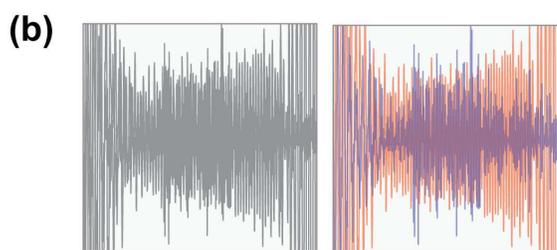
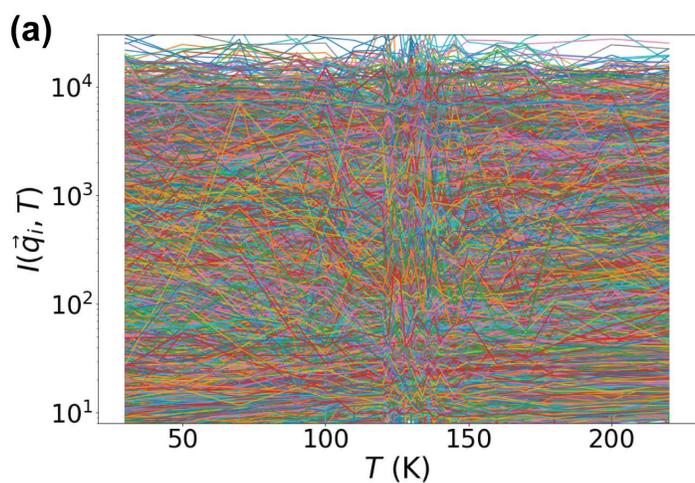
ACKNOWLEDGMENTS

401 We acknowledge the assistance of Anshul Kogar in the TiSe_2 measurements. We thank Jeffrey Lynn and Johnpierre
402 Paglione for assistance in preparing the $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$ samples. Initial development of *X-TEC* (EAK, AW, KW,
403 GP) was supported by NSF HDR-DIRSE award number OAC-1934714 and testing on TiSe_2 data was supported by
404 U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering under
405 Award DE-SC0018946 (JV). The experiments on $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$ and $\text{Cd}_2\text{Re}_2\text{O}_7$ (MK, SR, RO, PU, DP), and
406 the subsequent machine learning analysis and theoretical interpretations of the results (EAK, VK, JV, MN, KM),
407 were supported by the US DOE, Office of Science, Office of Basic Energy Sciences, Division of Material Sciences
408 and Engineering. MM acknowledges support by the National Science Foundation (Platform for the Accelerated
409 Realization, Analysis, and Discovery of Interface Materials (PARADIM)) under Cooperative Agreement No. DMR-
410 1539918 and the Cornell Center for Materials Research with funding from the NSF MRSEC program (DMR-1719875).
411 This research used resources of the Advanced Photon Source, a US DOE Office of Science User Facility operated for
412 the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Research
413 conducted at CHESS is supported by the National Science Foundation via Awards DMR-1332208 and DMR-1829070.
414

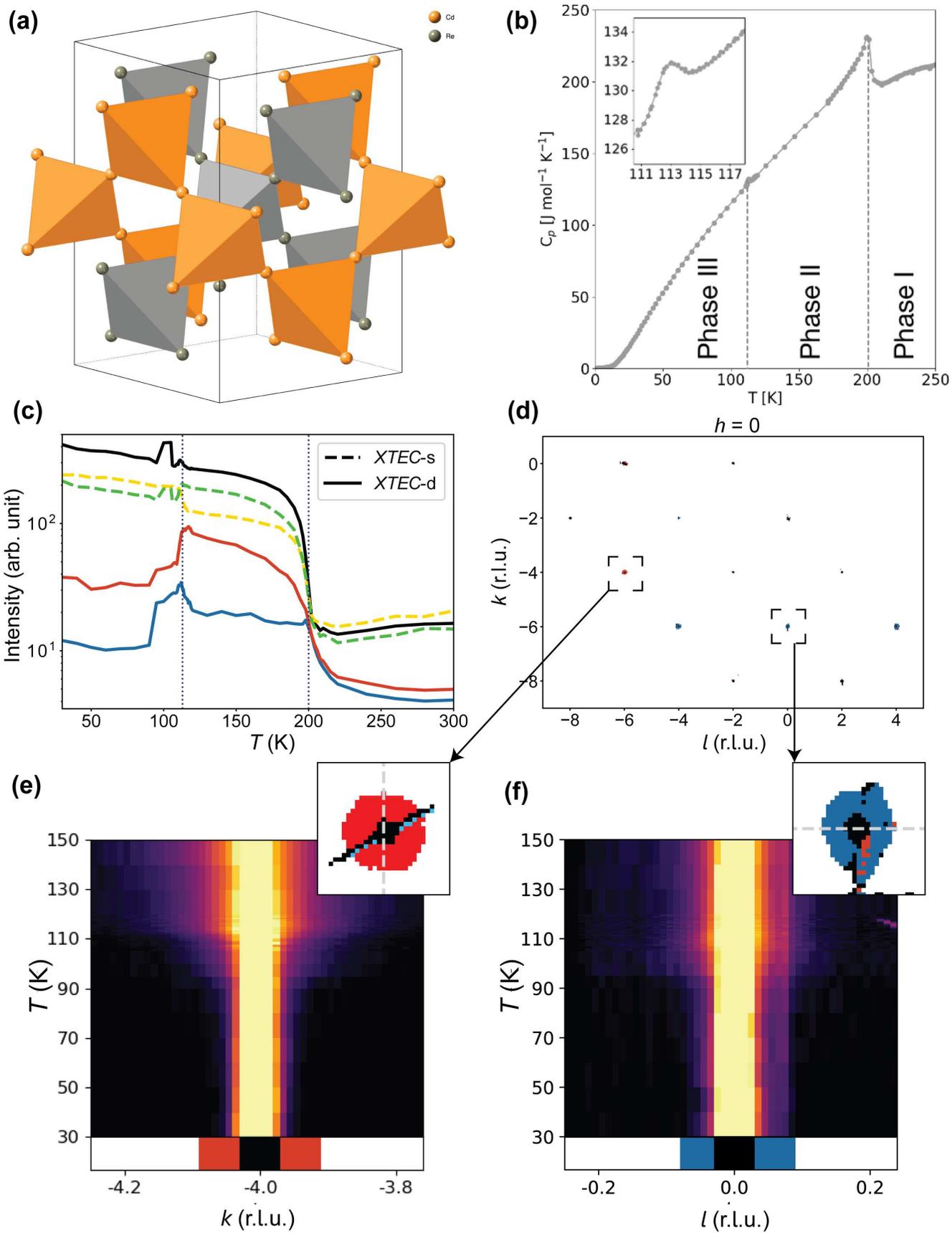
FIGURE CAPTIONS



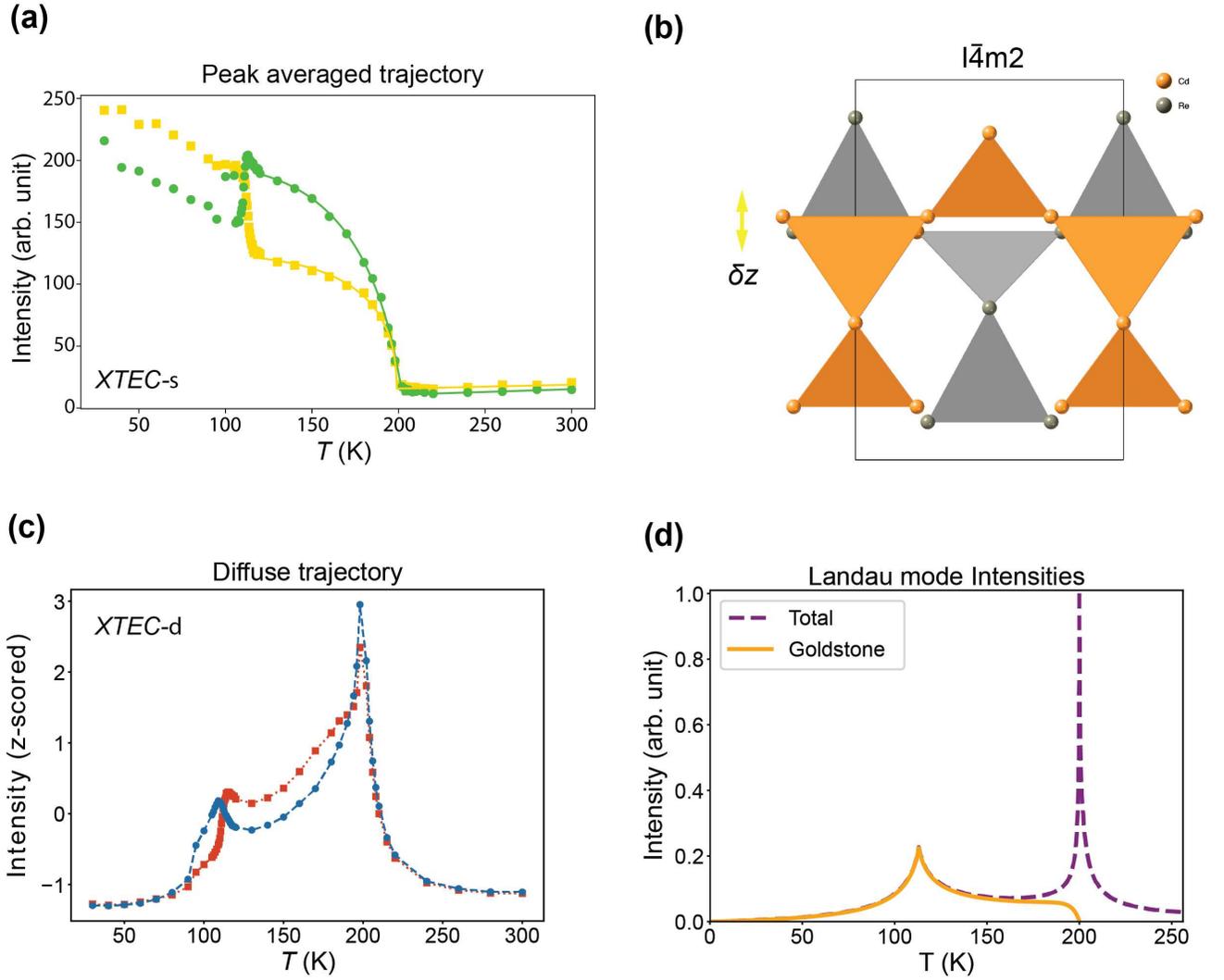
418 on the sample, which rotates about the orthogonal ϕ axis while images are captured on a fast area detector. The
 419 reciprocal space map shows the \vec{q} -coverage of a single plane in the 3D volume after capturing images over a full 360°
 420 sample rotation. A three-dimensional volume of reciprocal space covered by the x-ray scattering is shown on the
 421 right. Each red dot is a single Bragg peak. With an x-ray energy of 87 keV, a volume of over $10,000 \text{ \AA}^{-3}$ is measured,
 422 containing over ten thousand Brillouin zones if the unit cell dimension is 10 \AA . **(b-e)** Real space positions of atoms
 423 (top) and the corresponding scattering intensities (bottom) calculated from simulated one-dimensional crystals with
 424 a unit cell containing two atoms, illustrating (b) a high symmetry phase, with (c) distortions due to CDW order,
 425 (d) IUC order and (e) short-range IUC order. In (b), the high symmetry phase produces peaks at integer \vec{q} . In (c),
 426 displacements of the orange atoms by $\pm\delta$ double the size of the unit-cell producing additional super-lattice peaks at
 427 half-integer \vec{q} as well as changes in the other peak intensities. In (d), IUC distortions of the orange atoms by $-\delta$
 428 change the peak intensities without producing additional super-lattice peaks. In (e), every orange atom is displaced
 429 by $\pm\delta$, with a 70% probability of nearest neighbors having the same displacement. This finite correlation length has
 430 a small impact on the total scattering (black), but produces broad diffuse scattering (blue, x70000 scale compared
 431 to total scattering). **(f)** Bond patterns on the pyrochlore lattice associated with an E_u distortion as inferred in
 432 $\text{Cd}_2\text{Re}_2\text{O}_7$. The two space groups refer to the two different components of E_u with each bond color denoting a
 433 different bond length. The amount of distortion of each bond from the average bond (grey) is indicated by ++, -,
 434 etc. along with the respective bond color.



436 **Fig 2:** *X-TEC* with label smoothing (*X-TECs*) on $(\text{Ca}_x\text{Sr}_{1-x})_3\text{Rh}_4\text{Sn}_{13}$. **(a):** Example of raw intensity trajectories
437 for $\text{Sr}_3\text{Rh}_4\text{Sn}_{13}$. The plot shows the collection of individual raw temperature series $I(\vec{q}_i, T)$ for each point \vec{q}_i in the
438 data set spanning the reciprocal space $(h, k, l = 0)$ where $h, k \in [-15, 15]$ reciprocal lattice units (r.l.u.). **(b)** Sound
439 waveform of two people simultaneously talking (left) can be separated through clustering represented by different
440 colors (right). **(c)** Performing depth estimation for self driving cars, aggregating multiple sensor information with
441 label smoothing. Depth estimation from LIDAR (yellow) are highly accurate but sparse, while depth estimation
442 from cameras (blue) are dense but noisy. Label smoothing synthesizes the two sources, aligning the noisy camera
443 observations to match LIDAR observations [25]. **(d)** Two-cluster results of XRD data from $\text{Sr}_3\text{Rh}_4\text{Sn}_{13}$ with the
444 clustering assignments color-coded as yellow and blue. Each raw intensity trajectories of (a) are re-scaled $[\tilde{I}(\vec{q}_i, T)]$
445 by dividing their individual mean over temperature and subtracting one, before clustering. The lines represent
446 cluster means and the shaded region shows one standard deviation, interpolated between 24 temperature points of
447 measurement. **(e)** The corresponding yellow/blue cluster assignments of the \vec{q}_i pixels that passed the thresholding.
448 The image is zoomed to a section of the $(h, k, 0)$ plane. The low intensity background (white) and the $\tilde{I}(\vec{q}_i, T)$ with
449 low temperature variance (grey) are eliminated by thresholding (see SM. II-B). **(f)** The cluster means of the CDW
450 clusters are interpolated and plotted to reveal order parameter $\Delta(T)$ like behavior for four samples at different values
451 of Ca doping x . $\Delta(T)$ is estimated from the cluster means by subtracting the minimum from each cluster mean
452 and appropriate normalization. **(g)** The critical temperatures estimated from $\Delta(T)$ (yellow filled circles) overlaid
453 onto the known phase diagram from [29] based on phase boundaries from thermodynamic measurements and transport.



456 **Fig. 3** *X-TEC* analysis of $\text{Cd}_2\text{Re}_2\text{O}_7$. **(a)** Crystal structure of $\text{Cd}_2\text{Re}_2\text{O}_7$ showing only Cd and Re, in the high
 457 temperature cubic phase. **(b)** Temperature dependence of the specific heat of $\text{Cd}_2\text{Re}_2\text{O}_7$, showing the second-order
 458 phase transition at $T_{s1}=200$ K and the first-order phase transition at $T_{s2}=113$ K (see SM III-A). Three temperature
 459 ranges are marked as phase I ($T > T_{s1} = 200\text{K}$), phase II ($T_{s2} = 113\text{K} < T < T_{s1}$), and phase III ($T < T_{s2}$). **(c)**
 460 *X-TEC* results on the cubic forbidden Bragg peaks from high resolution XRD data, showing temperature dependence
 461 of the mean intensity of each cluster (the cluster assignments are obtained from $30 \text{ K} \leq T \leq 150 \text{ K}$ data, see SM
 462 III-C for details). The solid lines show three-cluster ($K = 3$) *X-TEC*d trajectories, color coded as black, red and
 463 blue. The dashed line shows two-cluster ($K = 2$) *X-TEC*s (peak averaged) trajectories, colored yellow and green.
 464 The temperatures of the two structural phase transitions are shown as dotted lines. **(d)** The *X-TEC*d cluster-color
 465 assignments of the thresholded pixels in a section of the $h = 0$ plane, where k and l are in reciprocal lattice units
 466 (r.l.u.). The pixels are assigned black, red and blue colors as in (c). The regions in the vicinity of two Bragg peaks at
 467 $0\bar{4}\bar{6}$ (left) and $0\bar{6}0$ (right) are magnified to show that the peak centers in both belong to the black cluster while halos
 468 form two distinct clusters (red and blue respectively) separated from their peak centers. **(e-f)** The raw intensity
 469 plotted for $0\bar{4}\bar{6}$ (left) and $0\bar{6}0$ (right) along a line cut (the grey dashed line shown in the respective zoom-ins) confirm
 470 the temperature dependence of the red and blue halo intensities represented by the cluster means in (c). Specifically,
 471 the $0\bar{4}\bar{6}$ peak has enhanced diffuse scattering above $T_{s2} \approx 113$ K, consistent with the temperature dependence of the
 472 red cluster mean. The $0\bar{6}0$ peak shows an anomaly near T_{s2} and a suppressed diffuse scattering above, consistent
 473 with the temperature dependence of the blue cluster mean.



474

475 **Fig. 4** Order parameters and their fluctuations inferred from *X-TEC* outcomes. (a) The filled symbols are the
 476 two-cluster mean intensity trajectories of peak averaged data [yellow and green trajectories from Fig. 3(c)], and solid
 477 lines are fits to these cluster means based on the model assuming δx displacements (yellow) and δz displacements
 478 (green) of cations to vary as $(T - T_c)^\beta$, with a common order parameter exponent of $\beta = 0.25$ as discussed in SM
 479 Section III-D. (b) Schematic diagram of the relative z -axis displacements of cation sublattices for the Cd (orange) and
 480 Re (gray) with respect to the cubic phase, inferred from the fit in (a). The *X-TEC*-discovered selection rule and the
 481 fit establish the approximately equal magnitude but out-of-phase displacements δz_{Cd} and δz_{Re} . (c) The characteristic
 482 temperature dependence of the diffuse clusters are revealed by the z-scored intensities (for each intensity, subtract
 483 their mean over T and then divide their standard deviation in T). The red and blue trajectories correspond to the
 484 respective cluster average of the z-scored intensities. Lines are guides for the eyes. The symbols square (circle) in (a)
 485 and (c) indicate that red (blue) diffuse clusters are associated with yellow (green) Bragg peaks. (d) The calculated

486 Landau mode intensities as a function of T (see SM III-F). Outside of the critical region near T_{s1} (200 K), the
487 intensity is dominated by the Goldstone mode intensity. Note the resemblance of the calculated intensity to the
488 diffuse trajectory in (c).

Figures

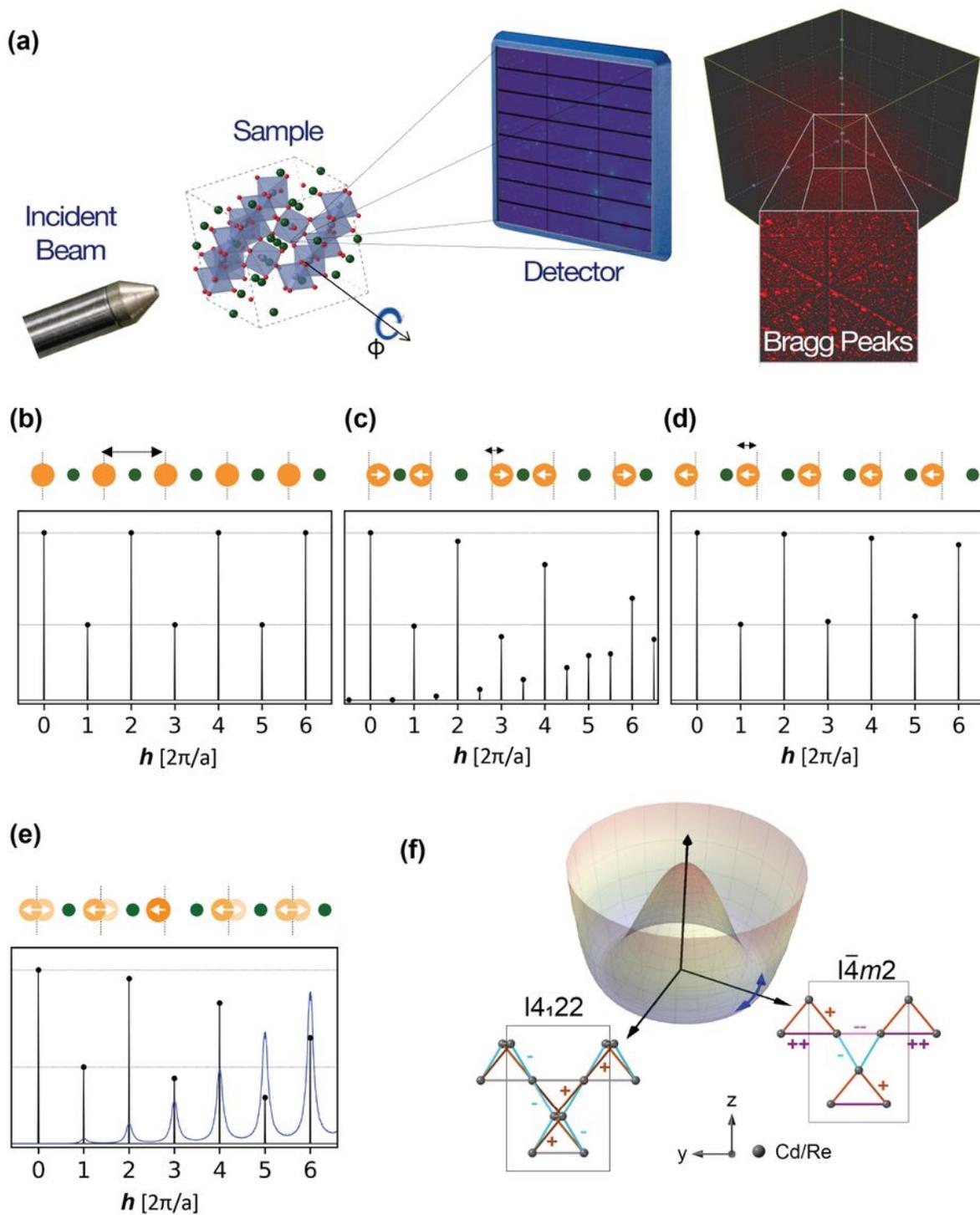


Figure 1

Schematic geometry of the x-ray scattering measurements. (see Manuscript file for full figure legend)

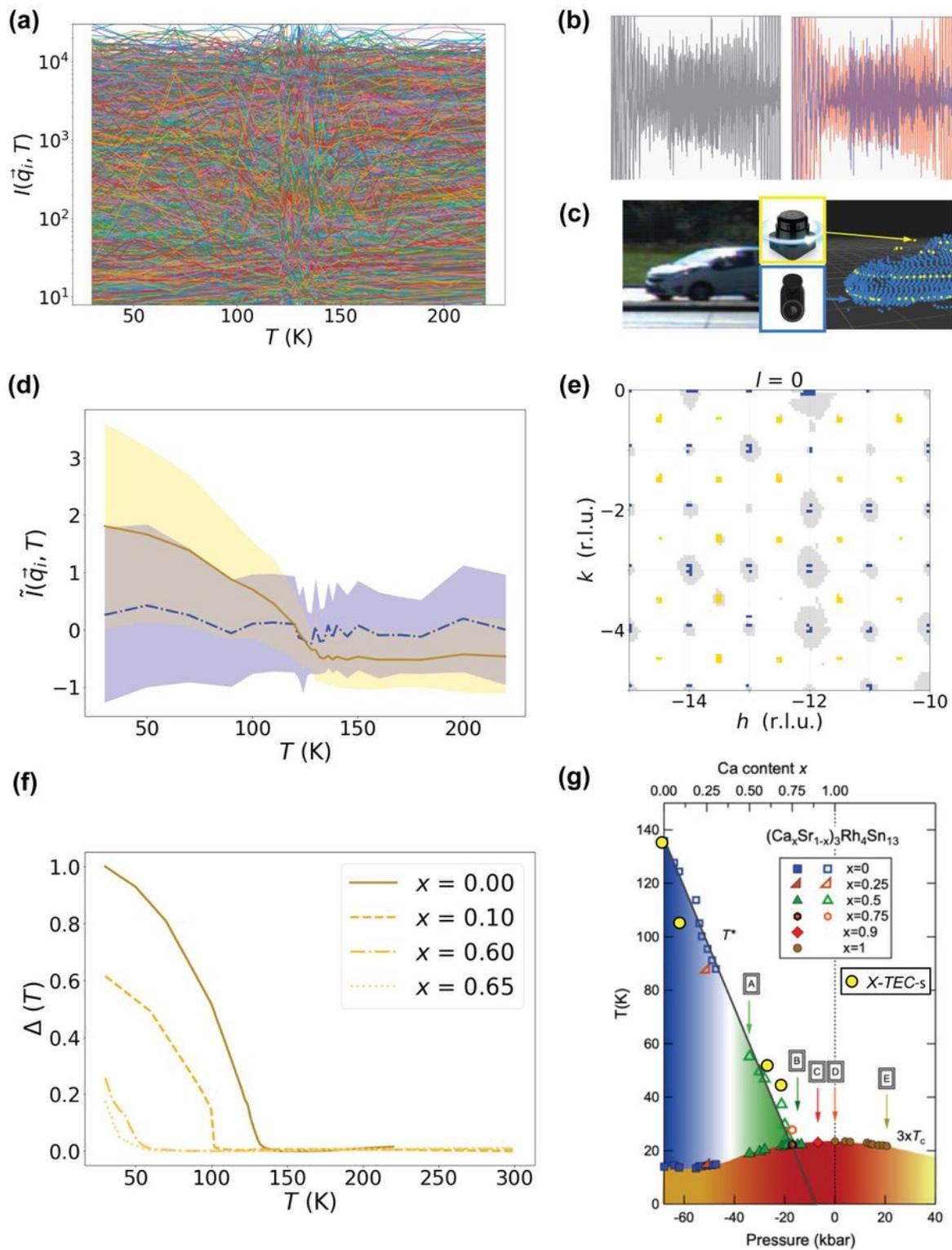


Figure 2

X-TEC with label smoothing (X-TECs) on $(\text{Ca}_x\text{Sr}_{1-x})_{436} 3\text{Rh}_4\text{Sn}_{13}$. (see Manuscript file for full figure legend)

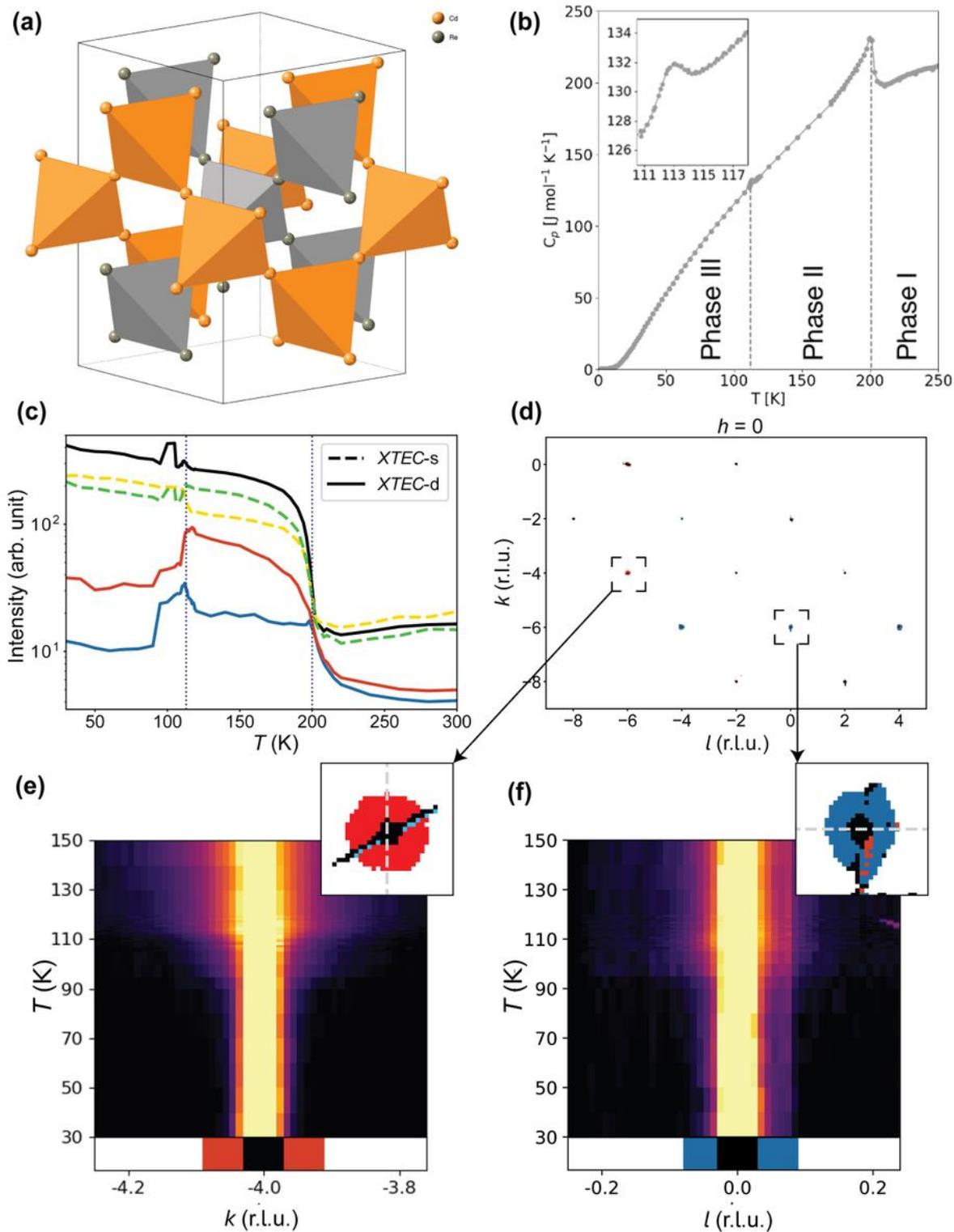


Figure 3

X-TEC analysis of $\text{Cd}_2\text{Re}_2\text{O}_7$. (see Manuscript file for full figure legend)

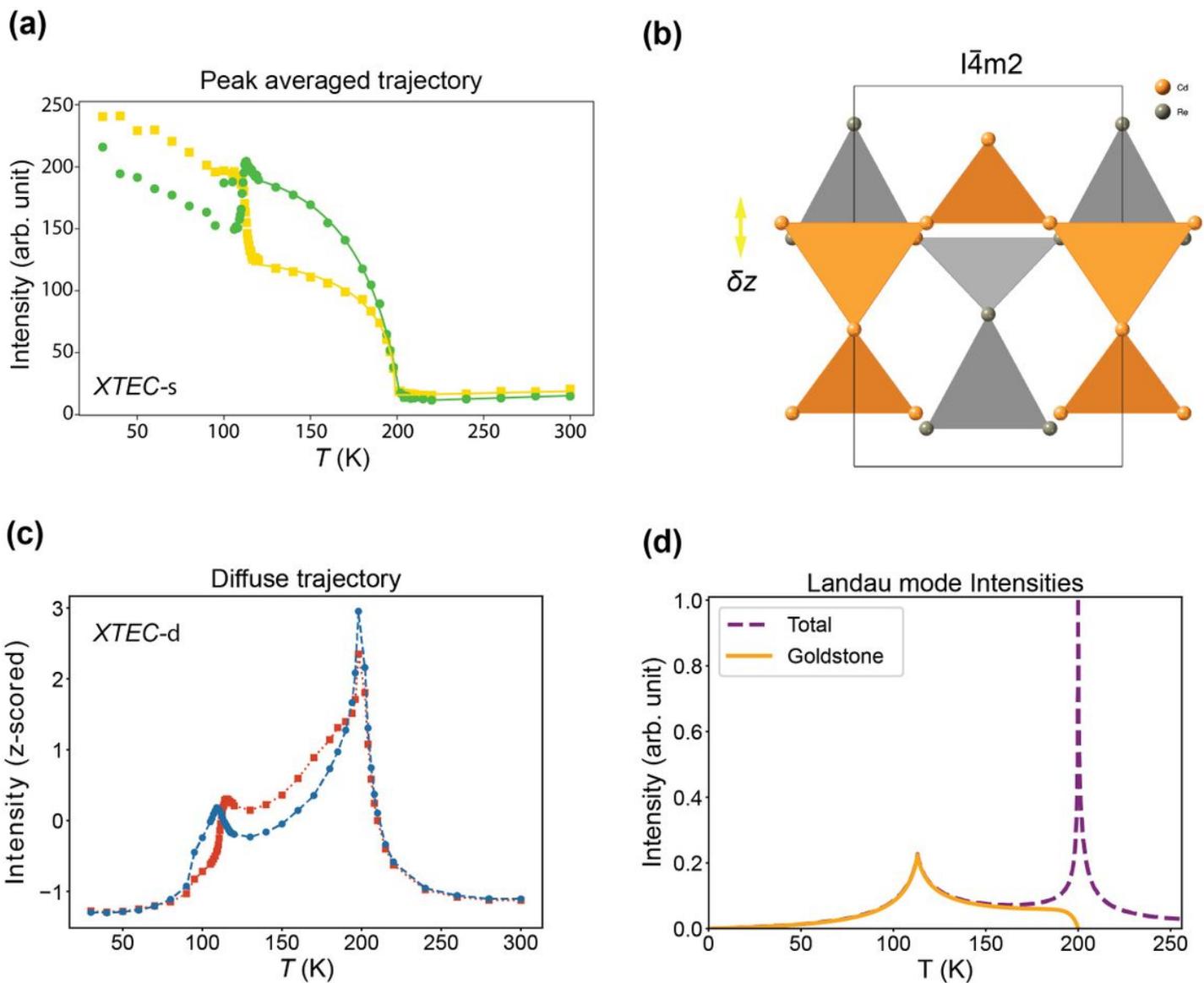


Figure 4

Order parameters and their fluctuations inferred from X-TEC outcomes. (see Manuscript file for full figure legend)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [sm.pdf](#)