



# Representational analysis of extended disorder in atomistic ensembles derived from total scattering data

James R. Neilson<sup>a\*</sup> and Tyrel M. McQueen<sup>b</sup>

Received 20 January 2015  
Accepted 2 September 2015

Edited by T. Proffen, Oak Ridge National Laboratory, USA

**Keywords:** extended disorder; atomistic ensembles; modeling; pair distribution function analysis.

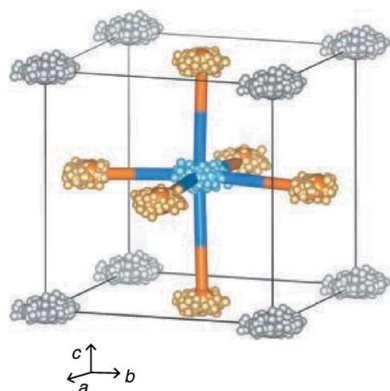
<sup>a</sup>Department of Chemistry, Colorado State University, CO 80523-1872, USA, and <sup>b</sup>Department of Chemistry, Department of Materials Science and Engineering, and Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA. \*Correspondence e-mail: james.neilson@colostate.edu

With the increased availability of high-intensity time-of-flight neutron and synchrotron X-ray scattering sources that can access wide ranges of momentum transfer, the pair distribution function method has become a standard analysis technique for studying disorder of local coordination spheres and at intermediate atomic separations. In some cases, rational modeling of the total scattering data (Bragg and diffuse) becomes intractable with least-squares approaches, necessitating reverse Monte Carlo simulations using large atomistic ensembles. However, the extraction of meaningful information from the resulting atomistic ensembles is challenging, especially at intermediate length scales. Representational analysis is used here to describe the displacements of atoms in reverse Monte Carlo ensembles from an ideal crystallographic structure in an approach analogous to tight-binding methods. Rewriting the displacements in terms of a local basis that is descriptive of the ideal crystallographic symmetry provides a robust approach to characterizing medium-range order (and disorder) and symmetry breaking in complex and disordered crystalline materials. This method enables the extraction of statistically relevant displacement modes (orientation, amplitude and distribution) of the crystalline disorder and provides directly meaningful information in a locally symmetry-adapted basis set that is most descriptive of the crystal chemistry and physics.

## 1. Introduction

Achieving an atomistic description of solids continues to provide a challenge to the study of materials, especially as we learn that imperfections and disorder of crystals can give rise to the emergence of unexpected materials properties. For example, the multifunctional properties of the perovskite manganites can only be explained by understanding the relationships between the local and average structures (Božin *et al.*, 2007; Wu *et al.*, 2007). Therefore, we strive to further classify and quantify the nature of any local ordering (short-range order) that is patterned in a disordered fashion. Pair distribution function (PDF) analysis of total scattering data has become a common technique for the characterization of local distortions and disorder in crystals, as well as of nanoparticle structures (Egami & Billinge, 2012; Billinge & Levin, 2007; Young & Goodwin, 2011; Keen & Goodwin, 2015).

Modeling of atomistic structures – with an emphasis on capturing the correct local structure – from experimentally derived atom–atom histograms poses a great challenge, especially when the best description of the PDF has a short finite correlation length (a domain) that becomes averaged into a higher symmetry in the crystallographic structure. To



OPEN  ACCESS

obtain an atomistic description of such a model with these domains (where each domain consists of a few unit cells), simulations containing thousands of atoms can be used to model the total scattering data. By employing a large-scale simulation, the limitations from periodic boundary conditions are lifted, thus allowing disordered aspects of the structure either to average out into the Debye–Waller factor in the case of crystalline disorder, or to lack any attributes of long-range order over the range of data provided in reciprocal space (after convolution with the finite size of the simulation) in order to describe amorphous solids (Renninger *et al.*, 1974; McGreevy & Pusztai, 1988; Elliott, 1984). However, analysis of these large-scale atomistic ensembles containing thousands of atoms has been nontrivial, both in the challenge of extracting information relative to the average crystallographic structure and also in providing statistically meaningful information; there are typically many more free parameters in these simulations than there are independent observations (*i.e.* data).

Herein, we develop a systematic approach for analyzing the disorder in large atomistic simulations of complex crystal structures using representational analysis. The determination of crystallographic superstructures resulting from displacive distortions *via* symmetry-mode analysis of a statistical distribution of ensembles has proven to be very powerful (*cf.*  $\text{WO}_3$  and  $\text{LaMnO}_3$ ) (Kerman *et al.*, 2012). Another similar approach, but coupled to a different analysis, has also made it possible to extract phonon dispersions from powder diffraction data (Dimitrov *et al.*, 1999; Goodwin *et al.*, 2004, 2005). Here, we use a variation of this technique adapted to the understanding of local structural variations by projecting displacements of atoms from their average crystallographic sites in atomistic ensembles onto a tight-binding-like basis formed from the symmetry-adapted<sup>1</sup> modes of a single unit cell, as depicted in Fig. 1; we define these modes as ‘tight-binding modes’. When displacements from an ideal crystallographic site are projected onto this locally symmetry-adapted basis, the disorder can be quantified and statistically analyzed to determine the frequency of specific displacement magnitudes and orientations. This manuscript outlines the analytical method and presents two illustrative applications of the method: the observation of a trigonal distortion in  $\text{BaTiO}_3$  at room temperature and the identification of the local displacement modes in the charge-ice pyrochlore  $\text{Bi}_2\text{Ti}_2\text{O}_7$ . More broadly, our approach is equally important for the analysis of experimental diffraction and scattering data (Shoemaker *et al.*, 2010; Shoemaker & Seshadri, 2010; King *et al.*, 2011), *ab initio* and force-field-based simulations (Dixon & Elliott, 2014; Palin *et al.*, 2014), and combinations of the two (White *et al.*, 2010*a,b*). Furthermore, this approach provides a common language and representation for bridging experiment- and theory-derived models.

<sup>1</sup> We define the modes as being *locally* symmetry-adapted, since the symmetry relationship that we use is only strictly defined for  $k = (0, 0, 0)$ . In a single unit cell, these modes are fully symmetry adapted.

## 2. Method

### 2.1. Introduction to total scattering methods

The analytical method described here operates on an ensemble of atoms that can be described as a enlarged ‘big box’ generated from small crystallographic unit cells. The atom positions need not sit on precisely ordered lattice sites; however, upon back-folding the big-box ensemble onto the parent unit cell, the average atom positions should project close to particular lattice sites, each with a position distribution resembling something like a Debye–Waller factor (*i.e.* the model may be paracrystalline). This method is agnostic to how the models are generated; the authors refer the reader to Egami & Billinge (2012), Young & Goodwin (2011), Keen & Goodwin (2015) and Tucker *et al.* (2007, 2001) for descriptions of modeling total scattering data.

Here, we use ‘total scattering’ to refer to the scattering of X-rays or neutrons that describes the structure factor of the crystallographic symmetry (diffraction from periodically ordered components) and the diffuse scattering that can arise from displacements of atoms from their ideal lattice points, including displacements from thermal motion and static disorder in the crystal (Egami & Billinge, 2012). If the total scattering structure factor,  $S(Q)$ , is measured to a sufficiently high momentum transfer [ $Q_{\text{max}} \gtrsim 15 \text{ \AA}^{-1}$ ;  $Q = (4\pi \sin \theta)/\lambda$ , where  $\theta$  is half the scattering angle and  $\lambda$  is the wavelength of the incident radiation, one can numerically take a sine Fourier transform to convert  $S(Q)$  into the reduced PDF,  $G(r)$ :

$$G(r) = 4\pi r \rho_0 [g(r) - 1] \\ = (2/\pi) \int_0^\infty Q[S(Q) - 1] \sin(Qr) dQ, \quad (1)$$

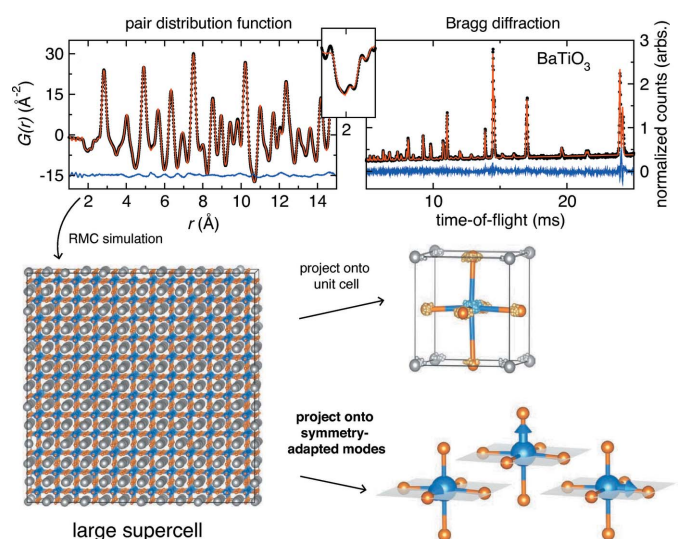


Figure 1

A schematic illustration of the workflow for generating a large ensemble simulation using reverse Monte Carlo simulations. Either this can be folded back onto a small unit cell, or the atomic displacements can be projected onto the tight-binding modes of a small unit cell. Inset:  $G(r)$  centered at  $r = 2 \text{ \AA}$  highlights the Ti–O bond lengths.

where  $\rho_0$  is the average number density of the material and  $g(r)$  is the atomic PDF. The atomic PDF,  $g(r)$ , is a direct measure of the relative positions of the atoms in a solid, *i.e.* an experimentally accessible real-space histogram of all atom–atom separations in the solid (of both periodically ordered and disordered atoms). Because of the crystallographic phase problem, without the use of isotopic labeling or anomalous scattering it is not possible to assign peaks directly in the PDF to specific atoms, so atomic scale modeling must be used to make assignments to individual peaks.

‘Small-box’ models, which allow the extraction of bond lengths and a description of the thermal motion (*i.e.* Debye–Waller factors), can be obtained from least-squares (LS) optimization of a crystallographic unit cell, or some small variant thereof, to the experimental PDF using the software *PDFgui* (Egami & Billinge, 2012; Proffen & Billinge, 1999; Farrow *et al.*, 2007). LS optimization is susceptible to finding local minima in the goodness-of-fit and is numerically cumbersome when the model contains many degrees of freedom, as applicable here. Additionally, these short-range-ordered models often fail to provide an accurate description of the crystallographic observations (Neilson *et al.*, 2012, 2013; King *et al.*, 2013).

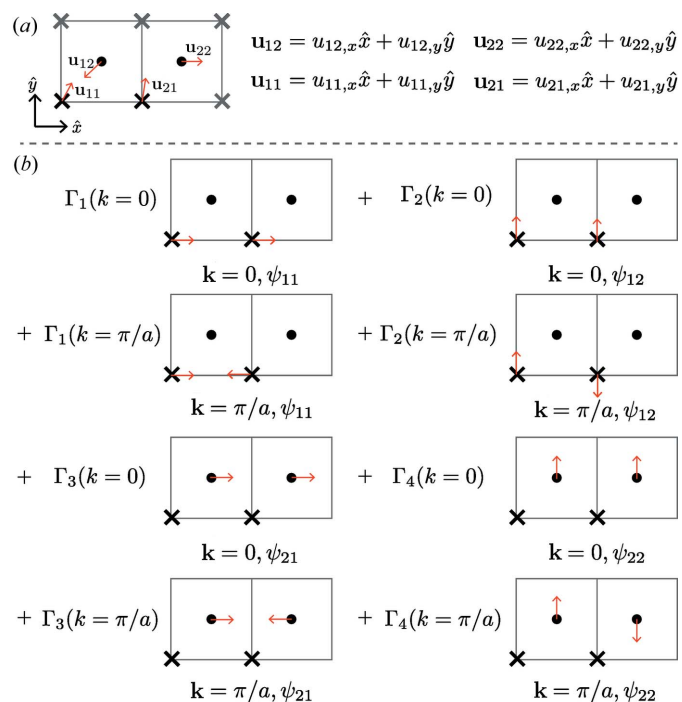
A complimentary approach to extract atomistic configurations from the PDF is to model simultaneously both the crystallographic structure factor and the PDF by employing a ‘large-box’ simulation of the total scattering data. A reverse Monte Carlo (RMC) algorithm can be used to find atomistic configurations of the ensemble consistent with both the experimentally determined  $G(r)$  and  $S(Q)$  (Tucker *et al.*, 2007, 2001).

### 2.2. Coordinate transform and decomposition

The goal of this method is to define the atomic configurations of an ensemble as displacements from the ideal crystallographic positions. Here, we define a ‘big’ or ‘large box’ as an  $M_x \times M_y \times M_z$  enlargement of the crystallographic unit cell to form an atomistic ensemble, but no attempt is made to constrain the symmetry between atoms, within either the subcells or the ‘large box’. The simplest such basis is simply to write down displacement vectors, in Cartesian or lattice coordinates, for each atom in the ensemble. Each atom within the crystallographic unit cell  $i$  has a unique position defined by a vector  $\mathbf{x}_{i,n}$ . The vector  $\mathbf{R}_n$  describes the spatial vector between each unit cell  $n$  within the ensemble. Each atom can be mapped as a displacement from its ideal position in the crystallographic unit cell,  $\mathbf{x}_{i,n}^0$ , by  $\mathbf{u}_{i,n} = \mathbf{x}_{i,n} - \mathbf{x}_{i,n}^0$ , where the values  $\mathbf{x}_{i,n}^0$  are often determined from a traditional crystallographic analysis (Rietveld analysis or single-crystal structural refinement). Such a representation is shown schematically in Fig. 2(a) for a simple two-dimensional ‘toy’ model, a  $2 \times 1$  ‘big box’ built from a crystallographic unit cell with two atoms and  $C_4$  symmetry. While straightforward to compute, this basis (the displacement vectors) lacks any connection to the symmetries that are present, locally or on average, and is thus difficult to interpret. A more refined

approach is to rewrite the displacements in terms of the normal modes of the crystallographic structure, with amplitudes and phases for every mode at every wavevector in the Brillouin zone (as determined by the point symmetry of each wavevector). This normal-mode basis provides physical insight because the atomic displacements are mapped onto symmetry-defined motions away from their ideal positions, and correlations between unit cells are captured. There is, however, an even better choice of basis that keeps many of the advantages of the classic normal-mode approach but retains physical insight into the *local* symmetry changes.

First, the local tight-binding (*i.e.* locally symmetry-adapted) modes are identified. This is accomplished by rewriting all possible atomic motions within a single unit cell into motions consistent with the point symmetry of the crystal at the Brillouin zone center,  $k = (0, 0, 0)$ . Each motion (or mode) can be labeled according to the irreducible representation (irrep) that it transforms under in the point symmetry group and is described by a set of basis vectors describing the actual atomic motions. Identification of these tight-binding modes is straightforward: basis vectors spanning each irreducible representation for each space group have been tabulated by Kovalev (1993), or can be computed by various crystallographic tools, including *KAREP* (Hovestreydt *et al.*, 1992), *SARAh* (Wills, 2002), *BASIREPS* (Rodriguez-Carvajal, 2001), the Bilbao Crystallographic Server (symmetry-adapted modes) (Aroyo *et al.*, 2011; Aroyo, Perez-Mato *et al.*, 2006; Aroyo, Kirov *et al.*, 2006; Kroumova *et al.*, 2003) and the *ISOTROPY* software suite (Stokes *et al.*, 2013). The inputs for these tools are the crystallographic space group and the atom



**Figure 2** Schematic illustrations of the coordinate transformations from atomic displacements into the tight-binding basis, based on locally symmetry-adapted modes. See text for discussion of parts (a) and (b).

positions of the small (crystallographic average) unit cell, as one would derive from Rietveld (or other suitable crystallographic) analysis.

These tight-binding modes provide an orthonormal and *local* basis for describing all possible motional degrees of freedom within a single unit cell, and are analogous to the normal vibrational modes of a molecular system. To retain this physical intuition but capture the degrees of freedom of a ‘large-box’ atomistic ensemble, we adopt a technique analogous to tight-binding methods in electronic structure calculations (Slater & Koster, 1954) and write down modes at a non-zero wavevector in terms of these local basis functions that we define at the Brillouin zone center (as for atomic orbitals), with appropriate phase factors to describe correlations between crystallographic unit cells in an ensemble. Specifically, we define the spatial correlations between unit cells within the ensemble with a quantized reciprocal wavevector,  $\mathbf{k} = 2\pi/\mathbf{R}$ . The vector spans the indices  $\mathbf{k}_x = (2\pi n_x/M_x, 2\pi n_y/M_y, 2\pi n_z/M_z)$  for all  $n_x = 0, 1, \dots, (M_x - 1)$ ,  $n_y = 0, 1, \dots, (M_y - 1)$  and  $n_z = 0, 1, \dots, (M_z - 1)$ ; in other words, the wavevectors are in steps of  $2\pi/M$  along each direction. For mathematical convenience, we define all values of  $\mathbf{k}$  as positive. The amplitude of a tight-binding mode,  $\Gamma_{j,\lambda}(\mathbf{k})$ , with the associated phase factor described by the reciprocal-space wavevector  $\mathbf{k}$ , is defined by

$$\Gamma_{j,\lambda}(\mathbf{k}) = \sum_n \left[ \sum_i \mathbf{u}_{i,n} \cdot \boldsymbol{\psi}_{i,j,\lambda} \exp(-i\mathbf{k} \cdot \mathbf{R}_n) \right], \quad (2)$$

where  $i$  runs over all atoms in the crystallographic unit cell and  $n$  runs over all unit cells contained within the ensemble. The vector  $\mathbf{R}$  points to the  $n$ th crystallographic unit cell in the ensemble. The values  $\boldsymbol{\psi}_{i,j,\lambda}$  are the vectorial contribution of atom  $i$  to the mode described by the  $(j, \lambda)$  pair. The vectorial contributions can span multiple atoms, as pertaining to the crystallographic multiplicity of the particular site in the original crystallographic unit cell. The index  $j$  specifies each set of modes that together transform as an irreducible representation of the point group;  $\lambda$  is equal to the dimensionality of the corresponding irreducible representation and runs over all modes in the set. Together, there are  $3N$  distinct  $(j, \lambda)$  pairs, or tight-binding modes, where  $N$  is the number of atoms in the small crystallographic cell.

There is no index  $k$  on  $\boldsymbol{\psi}$ , just like there is no wavevector dependence on atomic orbitals in the classic tight-binding electronic structure approach, because all wavevector dependences are explicitly included in the phase factors. Further, note that to retain all degrees of freedom we allow the amplitudes of each tight-binding mode to be independent of all others, even if symmetry would constrain them (*i.e.* because one irreducible representation may be spanned by multiple modes). This allows us to consider, but not enforce, symmetry in describing the ‘large-box’ atomistic ensembles. Stated differently, the projection is only a change of basis; all  $3N - 6$  degrees of freedom (for an ensemble of  $N$  atoms) are retained (omitting the three translational and three rotational degrees of freedom) and the exact atomistic ensemble can be reconstructed by the inverse of

$$\mathbf{u}_{i,n} = \sum_{\mathbf{k}} \sum_j \sum_{\lambda} \Gamma_{j,\lambda}(\mathbf{k}) \cdot \boldsymbol{\psi}_{i,j,\lambda} \exp(i\mathbf{k} \cdot \mathbf{R}_n). \quad (3)$$

This method, as applied to the toy model, is shown in Fig. 2(b). We note that this is distinct from typical crystallographic order parameter analysis (Kerman *et al.*, 2012; Dimitrov *et al.*, 1999; Goodwin *et al.*, 2004, 2005; Stokes *et al.*, 2013; Campbell *et al.*, 2006), in which the constraints of the parent crystallographic symmetry are preserved and the primary interoperable variables are the order parameter amplitudes, thus providing one number for a pair of basis vectors that describe a displacement transforming as a two-dimensional irreducible representation, *versus* two numbers in our approach. We retain all possible degrees of freedom.

### 2.3. Continuous symmetry measures

When using our tight-binding modes, we can determine the activity of the mode and the deviation of the ensemble from the crystallographic symmetry, not just from the mode amplitude but also from its mean-squared deviation (MSD) from an ensemble operated on by a symmetry operation of the parent crystallographic space group. There are at least two distinct types of continuous symmetry measures [as developed by Avnir and coworkers (Zabrodsky *et al.*, 1992; Alvarez *et al.*, 2005)] that we characterize here. First, the global activity of a single tight-binding mode ( $j$ , consisting of one or more individual  $\lambda$  modes depending on the dimensionality of the corresponding irreducible representation) can be quantified as the MSD between the  $|\Gamma_{j,\lambda}(\mathbf{k})|$  amplitudes and the new amplitude coefficients,  $|\Gamma_{G,j,\lambda}(\mathbf{k})|$ , following application of a symmetry operation  $G$  of the crystallographic space group:

$$s_{G,j} = \left\{ \sum_{\mathbf{k}} \left[ \sum_{\lambda} |\Gamma_{G,j,\lambda}(\mathbf{k})'|^2 - \sum_{\lambda} |\Gamma_{j,\lambda}(\mathbf{k})|^2 \right]^2 \right\}^{1/2}. \quad (4)$$

For purely symmetry-conserving displacements, the MSD should be zero. Here, it is critical to combine the squared amplitudes of all individual modes that together transform as a single multidimensional irreducible representation (the innermost sums) because the amplitudes of individual modes can be varied simply by changing the choice of basis vectors within that mode set. The sum over all wavevectors is justified to identify local symmetry changes because it corresponds to summing the contributions derived from a single local tight-binding mode (as for atomic orbital) and is exact in the molecular limit. The final square root is provided for convenience to make the magnitude of  $s_{G,j}$  more physically interpretable.

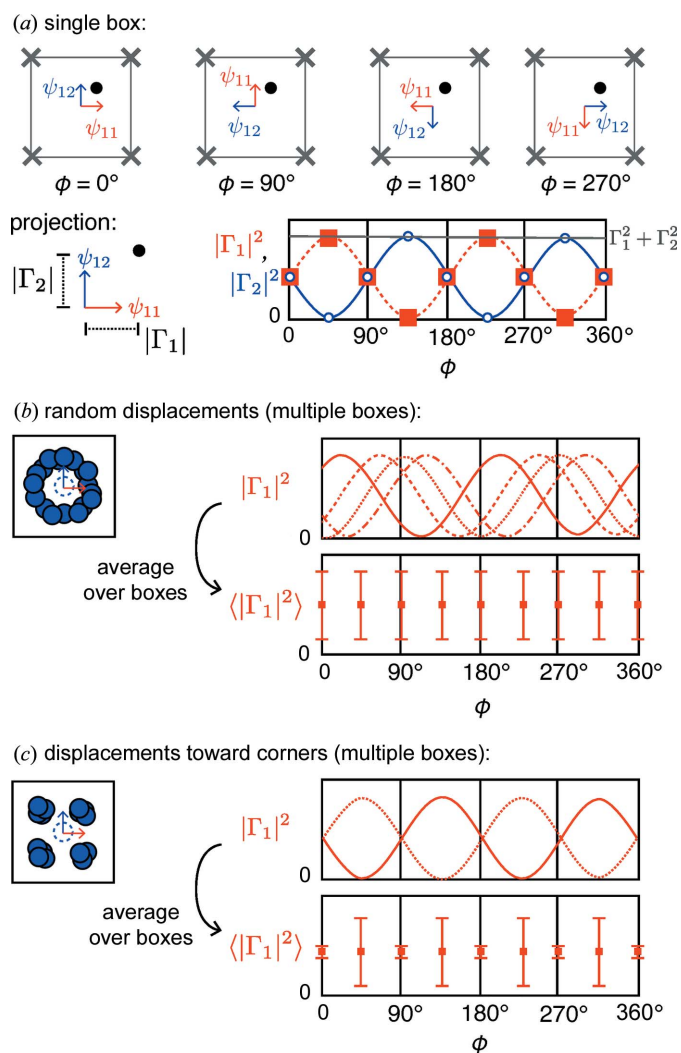
The related MSD, not broken down by individual mode sets, is similarly simple to calculate:

$$s_G = \left\{ \sum_j \sum_{\mathbf{k}} \left[ \sum_{\lambda} |\Gamma_{G,j,\lambda}(\mathbf{k})'|^2 - \sum_{\lambda} |\Gamma_{j,\lambda}(\mathbf{k})|^2 \right]^2 \right\}^{1/2}, \quad (5)$$

where again the final square root is provided for convenience.

The second type of deviation from the parent space group that can be identified is distortions that do not retain an equivalence of mode amplitudes within a single mode set that

transforms as a multidimensional irreducible representation. To illustrate this, consider a single box with  $C_4$  symmetry and an atom in the center displaced along the diagonal direction (Fig. 3). Projected onto the two basis vectors  $\Gamma_1$  and  $\Gamma_2$ , which together span the two-dimensional irreducible representation  $E$  in the corresponding point group, the amplitudes along each basis are initially equal. As the Euler angle that defines the absolute orientation of the basis vectors is varied, the intensity of  $\Gamma_2$  reduces while  $\Gamma_1$  increases until  $\Gamma_1$  is collinear with the atom displacement; this oscillatory pattern continues the rest of the way. Note that the sum of the square amplitudes from the two contributions ( $\Gamma_1$  and  $\Gamma_2$ ) is a constant (this is required, as the magnitude of the displacement is not changing). However, across multiple subcells or across multiple simulation runs, one can differentiate between random and ordered displacements. Let  $\phi_0$  be the initial angle of the



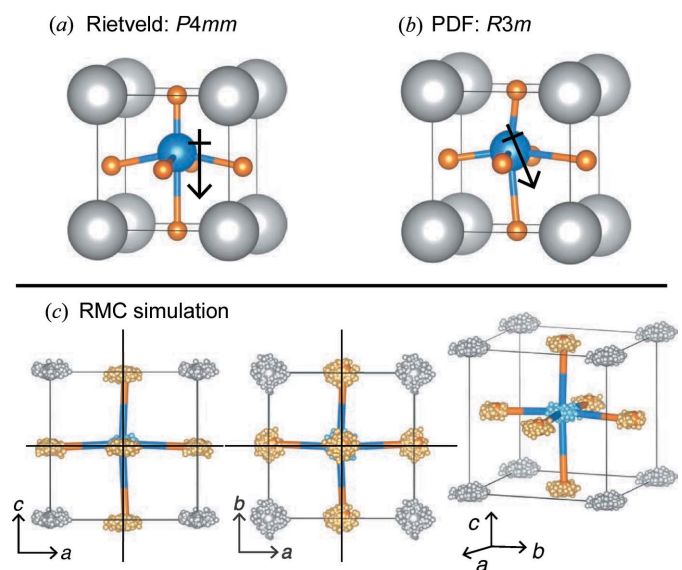
**Figure 3** Schematic illustrations of the rotation of the Euler angle  $\phi$  about the rotation axis of a particular symmetry element ( $C_4$ ). (a) In a single box, the amplitudes of the displacements oscillate as a function of  $\phi$ . (b) For random displacements across multiple subcells or boxes, the amplitude averages to a constant value as a function of  $\phi$  with a high and constant variance (denoted by error bars). (c) For displacements towards the four corners, the amplitude averages to a constant value as a function of  $\phi$ , but the variance oscillates as a function of  $\phi$  angle.

displacement of the central atom. Different values of  $\phi_0$  correspond to phase shifts of the values of  $\Gamma_1^2$  (and  $\Gamma_2^2$ ). If the  $\phi_0$  values are completely random, then their average is a flat line as a function of Euler angle, with variances that are also flat (Fig. 3b). On the other hand, if the  $\phi_0$  values are pinned to specific directions, then only a subset of the phase shifts is present. This will often result in an average that is still flat as a function of Euler angle, e.g. if they are pinned every  $90^\circ$ , but the variances will no longer be uniform (Fig. 3c). This can be exploited to determine whether the displacements are approximately random or fixed in some subset of orientations relative to the parent unit-cell coordinate system.

### 3. Case studies

#### 3.1. Trigonal displacements in tetragonal BaTiO<sub>3</sub>

The ferroelectric ceramic BaTiO<sub>3</sub> at  $T = 298$  K provides an excellent example of local distortion that averages out to a higher crystallographic symmetry in the unit cell. The average crystallographic symmetry determined from Rietveld analysis is tetragonal,  $P4mm$ , which was used to define the tight-binding modes. However, the local bonding environment is significantly distorted and better described by the symmetry of the low-temperature  $R3m$  configuration (Kwei *et al.*, 1995; Ravel *et al.*, 1998; Page *et al.*, 2010), as illustrated in Figs. 4(a) and 4(b). While the  $R3m$  model provides a quantitative fit to the PDF within one unit cell, it does not provide quantitative



**Figure 4** (a) Room-temperature  $P4mm$  crystal symmetry of BaTiO<sub>3</sub>, with exaggerated displacements of the Ti and O positions to illustrate the ferroelectric dipole. (b) The PDF analysis reveals a local distortion present at room temperature that resembles the low-temperature  $R3m$  crystal structure (Kwei *et al.*, 1995; Ravel *et al.*, 1998; Page *et al.*, 2010), illustrated here with exaggerated Ti and O displacements. (c) The folded atomistic big-box ensemble generated from an RMC simulation, overlain with the anisotropic displacement ellipsoids determined from small-box modeling of the PDF ( $R3m$  structure). The (200), (020) and (002) planes are shown to illustrate the net displacement of the O-atom positions rather than the Ti- and Ba-atom positions.

**Table 1**

Basis vector components along each crystallographic direction for BaTiO<sub>3</sub>, described in the *P4mm* space group setting, using the fractional atom coordinates Ba (0, 0, 0), Ti ( $\frac{1}{2}, \frac{1}{2}, 0.516$ ), O1a ( $\frac{1}{2}, 0, 0.487$ ), O1b (0,  $\frac{1}{2}, 0.487$ ) and O2 ( $\frac{1}{2}, \frac{1}{2}, 0.978$ ).

Irrep	Atom	Basis vector components		
		$\mathbf{u} \parallel \mathbf{a}$	$\mathbf{u} \parallel \mathbf{b}$	$\mathbf{u} \parallel \mathbf{c}$
Ba <i>A</i> <sub>1</sub>	Ba	0	0	1
Ba <i>E</i> (1)	Ba	1	0	0
Ba <i>E</i> (2)	Ba	0	1	0
Ti <i>A</i> <sub>1</sub>	Ti	0	0	1
Ti <i>E</i> (1)	Ti	1	0	0
Ti <i>E</i> (2)	Ti	0	1	0
O1 <i>A</i> <sub>1</sub>	O1a	0	0	1
O1 <i>B</i> <sub>1</sub>	O1a	0	0	−1
O1 <i>E</i> (1)	O1a	1	0	0
O1 <i>E</i> (2)	O1a	0	0	0
O1 <i>E</i> (3)	O1a	0	1	0
O1 <i>E</i> (4)	O1a	0	0	0
O2 <i>A</i> <sub>1</sub>	O2	0	0	1
O2 <i>E</i> (1)	O2	1	0	0
O2 <i>E</i> (2)	O2	0	1	0

information on the medium-range order, such as information on the correlations between unit cells or the coherence length scale, even though such information can (and should) exist within the PDF.

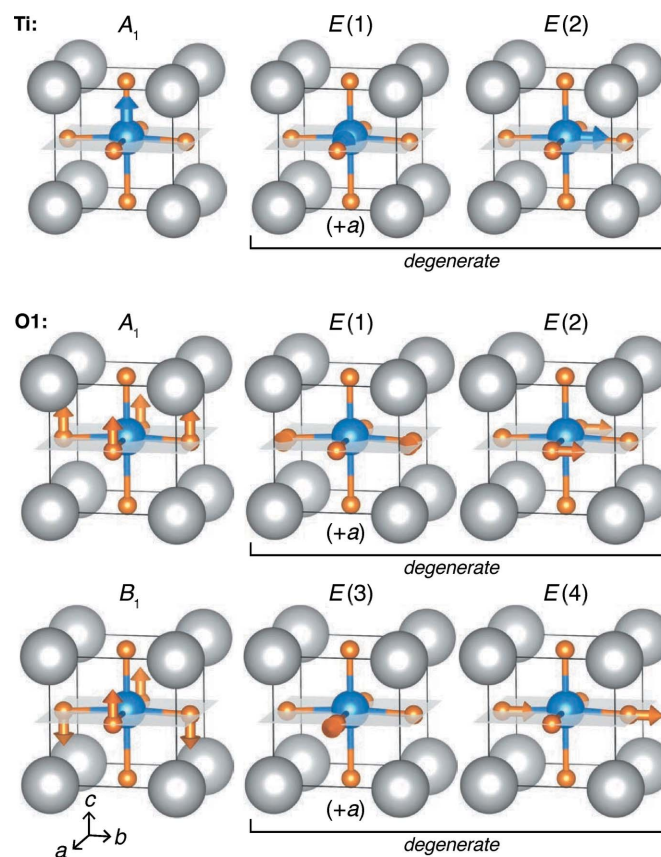
The experimental data used for this analysis were collected using the NPDF instrument (Lujan Neutron Scattering Center, Los Alamos National Laboratory, New Mexico, USA) and were re-analyzed with adjusted relative absorption corrections [such that a scale factor was not needed to fit the intensity  $G(r)$ ]; the experimental details and original report of the experimental data are given by Page *et al.* (2010). The Bragg profile and PDF were used to constrain RMC simulations using the *RMCprofile* code (Tucker *et al.*, 2007), as illustrated in Figs. 1(a) and 1(b). The simulation ensemble is a  $12 \times 12 \times 12$  enlarged big-box ensemble of the tetragonal *P4mm* unit cell (8640 atoms) that was determined from Rietveld analysis. The ensembles were constrained by  $G(r)$  (in the range  $1 < r < 24 \text{ \AA}$ ) in addition to the Bragg profile from the 90° detector bank of the NPDF ( $1.7 < Q < 15.7 \text{ \AA}^{-1}$ ,  $3.7 > d > 0.4 \text{ \AA}$ ). In addition to hard-sphere cutoffs, a small penalty was applied to the simulations for breaking [TiO<sub>6</sub>] coordination in order to accelerate the simulations. Two hundred different simulations were performed from the same starting configuration in order to build statistics in the analysis. Each simulation ensemble can be back-folded into the unit cell; the atom positions fall within a cloud-like distribution centered around the average crystallographic site (Fig. 4c).

Using the analysis method presented here, the atom positions were then decomposed into the tight-binding basis of the *P4mm* space group with a  $k$ -mesh divided into 12 discrete

steps along each  $x$ ,  $y$  and  $z$  direction with  $M_x = M_y = M_z = 12$ . The irreducible representations and corresponding basis vectors for the tight-binding (locally symmetry-adapted) modes were identified using the Bilbao Crystallographic Server (symmetry-adapted modes) (Kroumova *et al.*, 2003) and are listed in Table 1; some basis functions are represented graphically in Fig. 5.

For the analysis of a single ensemble of BaTiO<sub>3</sub>, the tight-binding mode amplitudes that describe displacements along the ferroelectric polarization are not very large (Ti *A*<sub>1</sub>, O1 *A*<sub>1</sub>, O2 *A*<sub>1</sub>, O2 *B*<sub>1</sub>, Table 2). This makes sense, since the average positions of the Ti and O2 atoms are off-center along the elongated  $c$ -axis direction (Table 1) (Megaw, 1945, 1973). However, the displacements in the  $ab$  plane are significantly enlarged. This is represented graphically by the ‘point-cloud’ distributions of the atom positions in Fig. 4(c) that are overlain on top of the *R3m* unit cell used to describe the PDF by Page *et al.* (2010).

One problem with RMC simulations is that, if the data are insufficiently resolved such that some atoms are poorly


**Figure 5**

Visualization of selected tight-binding modes of BaTiO<sub>3</sub> in the *P4mm* space group. For Ti (Wyckoff position 1b), all three basis vectors are shown, and they demonstrate the retained  $3N$  degrees of freedom, as the pair that together transform as *E* are allowed to have independent amplitudes. For the O1 site (Wyckoff position 2c), the *A*<sub>1</sub> and *B*<sub>1</sub> modes each join two O-atom positions, but  $3N$  degrees of freedom are retained for the two atoms generated from that Wyckoff position, noted by the six independent modes. While the *E* pairs will transform together if the local symmetry is also *P4mm*, all amplitudes are allowed to vary independently in this analysis.

**Table 2**

Amplitudes of the tight-binding modes (summed over all wavevectors) for the numerical average of 12 BaTiO<sub>3</sub> ensembles, and control simulations assuming *P4mm* or *R3m* local symmetry.

Values in parentheses are the variance of the coefficient. Values highlighted in bold are enlarged from the expected, isotropic values.

Irrep	Tight-binding mode coefficients (Å <sup>2</sup> )		
	Data	<i>P4mm</i> control	<i>R3m</i> control
Ba A <sub>1</sub>	0.0034 (1)	0.0059 (4)	0.0026 (3)
Ba E (1)	<b>0.0127 (4)</b>	0.0066 (5)	0.007 (1)
Ba E (2)	<b>0.0130 (3)</b>	0.0065 (5)	0.007 (1)
Ti A <sub>1</sub>	0.0089 (3)	0.0082 (6)	0.0024 (3)
Ti E (1)	<b>0.0143 (6)</b>	0.0077 (6)	0.009 (1)
Ti E (2)	<b>0.0128 (5)</b>	0.0076 (6)	0.009 (1)
O1 A <sub>1</sub>	0.0044 (1)	0.0043 (3)	0.0031 (4)
O1 E (1)	<b>0.0103 (3)</b>	0.0068 (5)	0.008 (1)
O1 E (2)	<b>0.0101 (3)</b>	0.0068 (5)	0.008 (1)
O2 A <sub>1</sub>	0.0040 (1)	0.0045 (3)	0.0025 (3)
O2 B <sub>1</sub>	0.0039 (1)	0.0044 (3)	0.0024 (3)
O2 E (1)	<b>0.0116 (4)</b>	0.0076 (6)	0.009 (1)
O2 E (2)	<b>0.0114 (4)</b>	0.0076 (6)	0.009 (1)
O2 E (3)	<b>0.0090 (3)</b>	0.0066 (5)	0.007 (1)
O2 E (4)	<b>0.0099 (2)</b>	0.0066 (5)	0.007 (1)

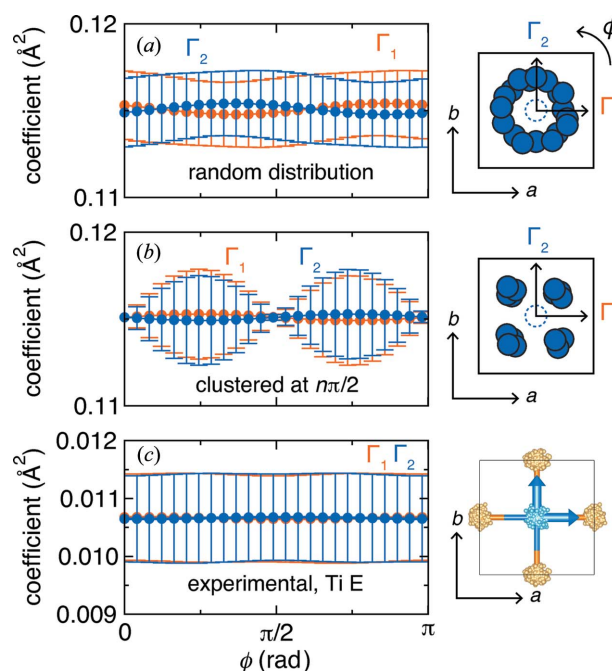
constrained, then the simulation atoms can wander away from their ideal positions. This would give the same graphical appearance as in Fig. 4(c). However, the quantitative data presented in Table 2 show that these displacements are significant on average within an ensemble and that their variance is tightly defined, even across 200 simulations.

As a control, we performed RMC simulations constrained by simulated PDFs. In one case (*P4mm* control), we computed  $G(r)$  from the *P4mm* crystal structure obtained by Rietveld analysis (convoluted with the appropriate instrumental resolution parameters,  $Q_{\text{damp}}$  and  $Q_{\text{broad}}$ ); the Bragg profile was the experimental Bragg profile. The simulated  $G(r)$  and Bragg profile were used to constrain 200 RMC simulations for analysis. For another control (*R3m* control), we took the reported *R3m* model determined from small-box modeling for the PDF [as reported by Page *et al.* (2010)] and simulated  $G(r)$  from that structural model; the Bragg profile was the experimental Bragg profile. These then constrained 90 independent RMC simulations for analysis. The *P4mm* control is a negative control that does not have additional displacements within the *ab* plane (beyond thermal disorder modeled by a Debye–Waller factor); the *R3m* control is a positive control for a known displacement in the *ab* plane coincident with thermal disorder. The tight-binding mode coefficients resulting from analysis of the *P4mm* control simulations do not have a substantial anisotropy (Table 2); while there is a statistically significant increase in the coefficients of displacements in the *ab* plane, this may be biased from using the experimental Bragg peaks in conjunction with the simulated  $G(r)$ . For the *R3m* control, there is a significant and expected increase in displacements within the *ab* plane. This analysis informs us that the tight-binding mode amplitudes are capable of identifying aperiodic displacements when expected; however, the

values of the coefficients alone do not inform us as to whether particular symmetry operations are broken.

With a local trigonal distortion, the *R3m*-based model implies that there are specific vectors along which the Ti displacements are oriented; these are the vectors that point directly at the faces of the [TiO<sub>6</sub>] octahedra (*i.e.* the {111} directions, as referenced to the *P4mm* or *Pm3m* unit cells of BaTiO<sub>3</sub>). However, looking at the graphical representation in Fig. 4(c), it is impossible to tell if particular directions are preferred. Because the tight-binding modes within a set are mutually orthogonal and therefore yield locally orthogonal displacements, it is trivial to rotate the reference frame of the basis vectors and recompute their coefficient as a function of the Euler angle along the rotation axis of the multidimensional irreducible representation. In the *P4mm* description, this angle ( $\phi$ ) rotates around the fourfold axis of the unit cell.

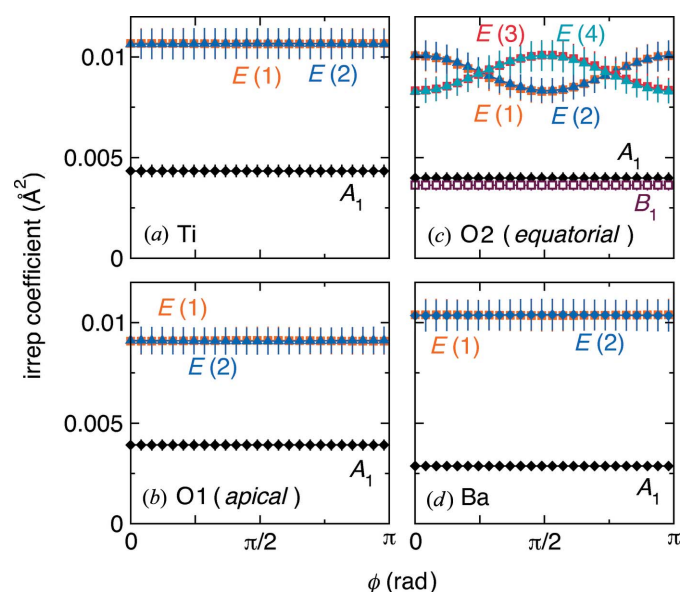
In our analysis, we decomposed the atomic displacements into amplitudes of specific tight-binding modes as a function of rotation about the Euler angle,  $\phi$  (Fig. 6). To illustrate this analysis, we employ two control simulations. Shown in Fig. 6(a) is a simulation of the displacements of Ti atoms around an approximately random distribution of  $\phi$  angles. In Fig. 6(b), we show a simulation with Ti atoms displaced at the same



**Figure 6** Euler angle analysis of a multidimensional tight-binding mode set that transforms as a multidimensional irreducible representation. (a) A random distribution of atom positions from the crystallographic location (dashed circle in cartoon) produces an equivalent variance of the mixing coefficients (denoted with vertical bars) between the two basis vectors ( $\psi_1$  and  $\psi_2$ ) when the basis vectors are rotated about the Euler angle,  $\phi$ , parallel to the  $C_4$  axis of the crystal structure. (b) A clustering of positions at regular intervals, such as  $\pi/2$ , will produce the same mixing coefficients as in part (a) for each tight-binding mode when averaged over all  $\mathbf{k}$  and over all ensembles. However, a clustering of positions will yield a significant variance in the coefficients, denoted by the vertical bars. (c) The coefficients provided from the experimentally derived BaTiO<sub>3</sub> ensembles do not display significant differences when the basis vectors are rotated about the Euler angle.

magnitude as in Fig. 6(a), but the angles are constrained to be a random integer multiple of  $\pi/2$  rad. Therefore, the Ti atoms are clustered into four groups (akin to the  $\langle 111 \rangle$  displacements). In both cases, the average coefficient of the tight-binding modes that together form a set and span a multi-dimensional irreducible representation will not change as a function of  $\phi$ , since the Ti atoms are displaced from the center by the same distance. However, the variance between tight-binding mode amplitudes [ $E(1)$  versus  $E(2)$ ] will be distinct for each Euler angle (*cf.* Fig. 3). For the completely random distribution in Fig. 6(a), the coefficient multiplying  $\Gamma_1$  of the irreducible representation  $E$  will vary continuously between 0 and the maximum value, as the basis vector is orthogonal and collinear with the atomic displacement; the second basis vector ( $\Gamma_2$ ) will also vary by the same amount, but its amplitude will be  $\pi/2$  out of phase with  $\Gamma_1$ . Therefore, each basis vector will have the same variance with  $\phi$ , denoted by the error bars in Fig. 6.

As in Fig. 6(b), if the atom displacements are clustered into groups, then the variation of basis vector coefficients will not be constant with  $\phi$ . When  $\phi = 0$  rad, such that  $\Gamma_1$  is oriented along the  $a$  axis, then its mixing coefficient will be  $2^{1/2}$  times the average value; the coefficient of  $\Gamma_2$  will be identical. Therefore, the difference in coefficients is zero. However, when  $\phi$  orients one of the basis vectors directly towards the clustered displacements, one coefficient is maximal and the other is zero; this produces a large variation in the basis vector amplitudes. In the experimental simulations, there does not appear to be explicit clustering of the Ti atoms along particular displacement vectors (Fig. 6c). Looking at the variation in coefficients for all atoms in the unit cell, depicted by the error bars in Fig. 7, there does not appear to be any clustering of displacements as a function of Euler angle. While the two-dimensional irreducible representations  $E$  for atom O2 appear to exhibit a

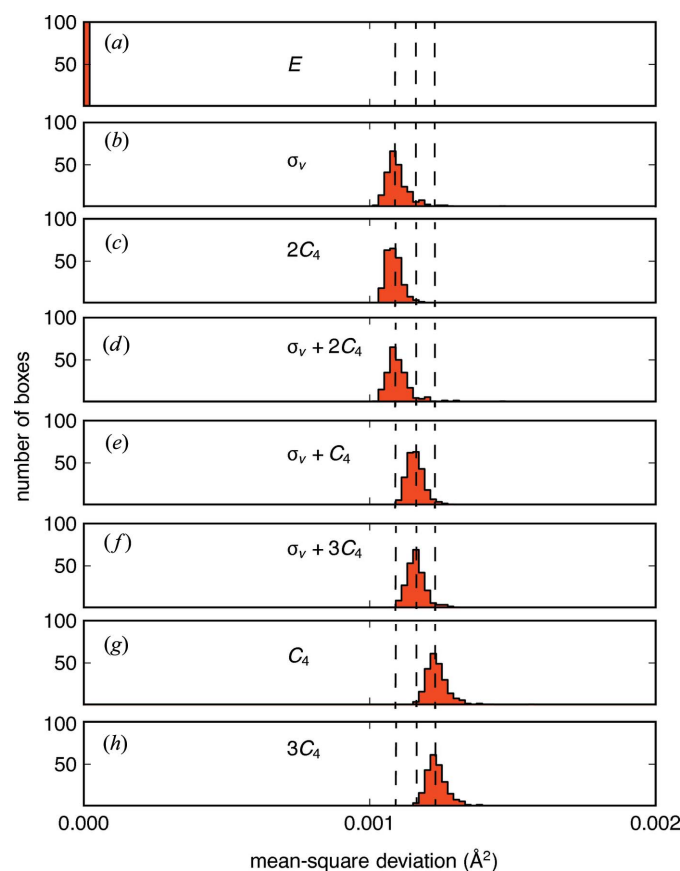


**Figure 7**  
Tight-binding mode coefficients as a function of Euler angle,  $\phi$ , parallel to the  $C_4$  axis of  $P4mm$  for the (a) Ti, (b) O1, (c) O2 and (d) Ba crystallographic sites, illustrating a constant variance as a function of  $\phi$ .

trend with  $\phi$ , the change in the average value of the coefficient reflects the definition of the basis vectors; the variations of the coefficients, as indicated by the error bars, do not change with  $\phi$ . This result is consistent for RMC simulations run for different times (as disorder tends to be artificially maximized for longer simulation runs).

While a variation in coefficients with Euler angle can indicate clustering of displacements described by a multi-dimensional irreducible representation, it does not provide any indication of whether the degeneracy-inducing symmetry operation is broken. To find broken degeneracies, we turn to continuous symmetry measures as defined in the *Method* section. For  $\text{BaTiO}_3$ , we compute the MSD for each generated symmetry operation of the crystallographic space group ( $P4mm$ ). With four symmetry operations ( $E$ ,  $\sigma_v$ ,  $C_2$  and  $C_4$ ), there are a total of eight symmetry-related atoms that are generated from a general position; therefore, we test all unique combinations of these operations (each combination that generates one of the general positions).

The MSDs illustrate that the atomic displacements in the ensembles show the highest deviation away from the fourfold rotation symmetry element. Histograms of all MSDs computed for  $\text{BaTiO}_3$  (summed over all  $\mathbf{k}$  and irreducible



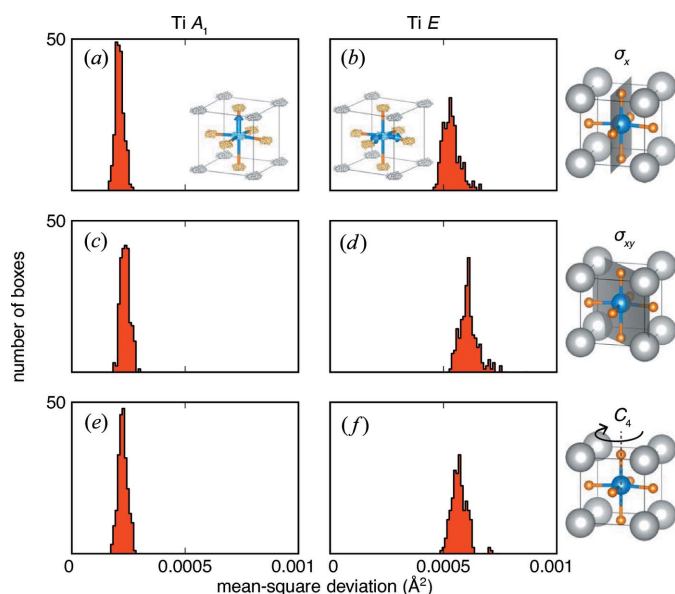
**Figure 8**  
Histograms of the mean-square displacements, summed over all  $\mathbf{k}$  and mode sets ( $j$ ) for 240 ensembles for each symmetry operation of  $P4mm$ . (a) The identity  $E$ , (b) a vertical mirror along  $x$ ,  $\sigma_v$ , (c) a twofold axis,  $2C_4 = C_2$ , (d) a vertical mirror and twofold axis  $\sigma_v + 2C_4$ , (e) a vertical mirror plane along  $xy$ ,  $\sigma_v + C_4 = \sigma_{xy}$ , (f) the fourfold rotation axis  $C_4$  and (g)  $3C_4$ .



representations) are illustrated in Fig. 8 for each symmetry operation. The histograms for related symmetry elements are clustered together; those combinations that equate to a four-fold rotation have the most significant MSD (Figs. 8g and 8h), followed by mirror planes parallel to the {110} planes, then mirror planes parallel to the {100} planes.

To probe which irreducible representations are most symmetry conserving, histograms of MSDs summed over all  $\mathbf{k}$  for each irreducible representation are shown in Fig. 9; the histograms bin together the MSDs computed for the equivalent symmetry operations shown on the right. The histograms for the Ti  $A_1$  irreducible representation (Figs. 9a, 9c and 9e) show tightly grouped and low-value MSDs, indicating that the vertical Ti displacements tend to preserve the  $P4mm$  symmetry operations. However, the displacements that project onto the Ti irreducible representation  $E$  tend to break the symmetry operations, as inferred previously. The fourfold rotation axis appears to be the symmetry operation most frequently broken, as expected naively from the small-box trigonal model illustrated in Fig. 4(b), which does have a vertical mirror plane parallel to the (110) plane.

The analyses presented here for  $\text{BaTiO}_3$  provide results that are sufficiently simple for easy comparison with small-box models of  $\text{BaTiO}_3$ . The use of RMC simulations allows one to extract a single statistically relevant model of the atom positions that describes both the data regarding local atom separations (the PDF) and the average crystallographic symmetry (Bragg profile). For  $\text{BaTiO}_3$ , the coefficients of the tight-binding modes and their spatial dependence reveal the presence of a significant distortion from the  $P4mm$  crystallographic symmetry. The resulting ensemble reveals that the atom positions are mostly displaced in the  $ab$  plane, which closely resembles the low-temperature  $R3m$  crystal structure.



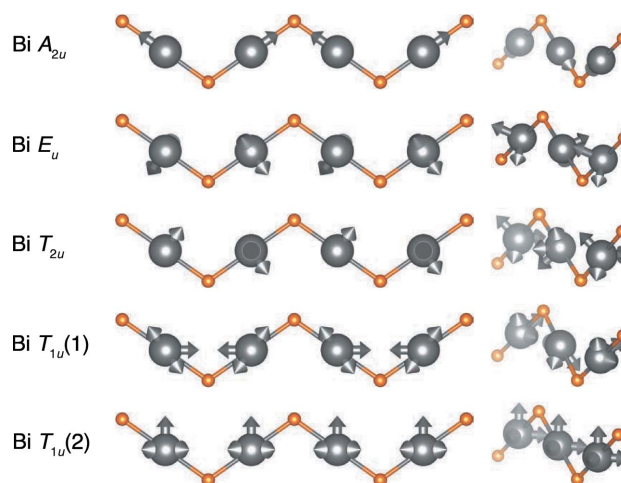
**Figure 9**  
Histograms of the mean-square displacements of the (a), (c), (e) Ti  $A_1$  and (b), (d), (f) Ti  $E$  mode sets, summed over all  $\mathbf{k}$  for each unique symmetry orientation: (a), (b) the  $\sigma_x$  mirror plane, (c), (d) the  $\sigma_{xy}$  mirror plane and (e), (f) the  $C_4$  rotation axis.

This example illustrates how such an analysis may be performed on materials with more complexity, in terms of both their crystal structure and their crystalline disorder, as described in the next section.

### 3.2. Correlated O and Bi displacements in $\text{Bi}_2\text{Ti}_2\text{O}_7$

The analysis methods presented here are generally applicable to materials with more complex structures. The ‘charge-ice’ pyrochlore oxide  $\text{Bi}_2\text{Ti}_2\text{O}_7$  has a large unit cell that contains 88 atoms; direct inspection of ensembles becomes prohibitive with this many degrees of freedom in a single crystallographic unit cell (Hector & Wiggin, 2004). In  $\text{Bi}_2\text{Ti}_2\text{O}_7$  there is extensive disorder of the Bi sublattice, attributed to stereochemical activity of the lone pair – derived from the  $[\text{Xe}]5d^{10}6s^2$  electron configuration of  $\text{Bi}^{\text{III}}$  – on a geometrically frustrated lattice. The geometry of the diamond lattice prevents long-range ordering of the dipoles, in a manner related to Pauling’s ice rules (Seshadri, 2006). Previously, RMC simulations of total neutron scattering have been used to gain an atomistic representation of the static Bi displacements, which form a toroidal distribution of Bi atoms that encircle the ideal crystallographic site. Furthermore, the  $\text{O}'$  atoms (Wyckoff site 8a) are connected to the non-spherically distributed Bi atoms and therefore become displaced from their ideal crystallographic sites into tetrahedral volumes. The original report, experimental data and experimental details are given by Shoemaker *et al.* (2010). The crystallographic  $\text{Bi}_2\text{Ti}_2\text{O}_7$  unit cell is described by the  $Fd\bar{3}m$  space group, which defines the irreducible representations and tight-binding modes used here.

By rewriting the atomic displacements in terms of the tight-binding modes, a straightforward examination of their amplitudes reveals several characteristics that lead to many of the same conclusions as presented by Shoemaker *et al.* (2010);



**Figure 10**  
Visualization of the tight-binding modes for the different Bi irreducible representations. The  $E_u$ ,  $T_{2u}$  and  $T_{1u}(1)$  mode sets have the largest amplitude. They describe the displacements that generate a toroidal distribution of Bi positions and agree with the predicted displacement modes of  $E_u$  and  $T_{1u}$  symmetry from *ab initio* density functional theory calculations (Shoemaker *et al.*, 2010).

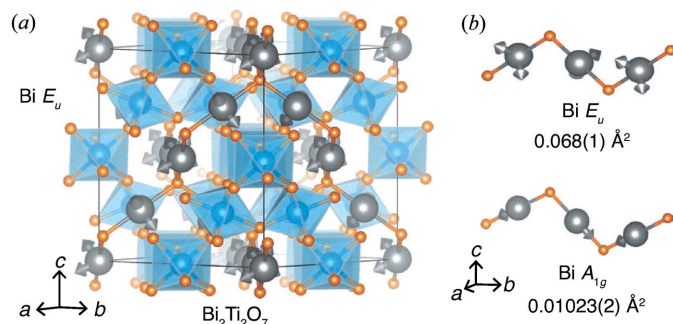
**Table 3**

 Tight-binding mode amplitudes obtained from averaging 320 RMC simulations of  $\text{Bi}_2\text{Ti}_2\text{O}_7$ .

 Atom positions are defined as Bi (0, 0, 0), Ti ( $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$ ), O1 (0.43114, 0.125, 0.125) and O2 (0.125, 0.125, 0.125). Amplitudes are also averaged across atoms that are related through face centering.

Irrep	Coefficient ( $\text{\AA}^2$ )	Irrep	Coefficient ( $\text{\AA}^2$ )	Irrep	Coefficient ( $\text{\AA}^2$ )
Bi $A_{2u}$	0.010 (1)	Ti $A_{2u}$	0.003 (4)	O1 $A_{1g}$	0.021 (2)
Bi $E_u$ (1)	0.067 (8)	Ti $E_u$ (1)	0.008 (1)	O1 $A_{2u}$	0.006 (1)
Bi $E_u$ (2)	0.067 (7)	Ti $E_u$ (2)	0.008 (1)	O1 $E_u1$	0.006 (1)
Bi $T_{2u}$ (1)	0.068 (8)	Ti $T_{2u}$ (1)	0.008 (1)	O1 $E_u2$	0.006 (1)
Bi $T_{2u}$ (2)	0.067 (8)	Ti $T_{2u}$ (2)	0.008 (1)	O1 $E_g$	0.006 (1)
Bi $T_{2u}$ (3)	0.067 (8)	Ti $T_{2u}$ (3)	0.008 (1)	O1 $E_g$	0.006 (1)
Bi $T_{1u1}$ (1)	0.030 (4)	Ti $T_{1u1}$ (1)	0.005 (1)	O1 $T_{2u1}$ (1)	0.007 (1)
Bi $T_{1u1}$ (2)	0.030 (4)	Ti $T_{1u1}$ (2)	0.005 (1)	O1 $T_{2u1}$ (2)	0.007 (1)
Bi $T_{1u1}$ (3)	0.029 (3)	Ti $T_{1u1}$ (3)	0.005 (1)	O1 $T_{2u1}$ (3)	0.008 (1)
Bi $T_{1u2}$ (1)	0.049 (6)	Ti $T_{1u2}$ (1)	0.006 (1)	O1 $T_{2u2}$ (1)	0.007 (1)
Bi $T_{1u2}$ (2)	0.049 (6)	Ti $T_{1u2}$ (2)	0.006 (1)	O1 $T_{2u2}$ (2)	0.007 (1)
Bi $T_{1u2}$ (3)	0.048 (6)	Ti $T_{1u2}$ (3)	0.006 (1)	O1 $T_{2u2}$ (3)	0.007 (1)
O1 $T_{2g1}$ (1)	0.007 (1)	O1 $T_{1u2}$ (1)	0.006 (1)	O2 $T_{2g}$ (1)	0.029 (3)
O1 $T_{2g1}$ (2)	0.007 (1)	O1 $T_{1u2}$ (2)	0.006 (1)	O2 $T_{2g}$ (2)	0.029 (3)
O1 $T_{2g1}$ (3)	0.007 (1)	O1 $T_{1u2}$ (3)	0.006 (1)	O2 $T_{2g}$ (3)	0.029 (3)
O1 $T_{2g2}$ (1)	0.006 (1)	O1 $T_{1u3}$ (1)	0.007 (1)	O2 $T_{1u}$ (1)	0.030 (3)
O1 $T_{2g2}$ (2)	0.006 (1)	O1 $T_{1u3}$ (2)	0.007 (1)	O2 $T_{1u}$ (2)	0.030 (3)
O1 $T_{2g2}$ (3)	0.006 (1)	O1 $T_{1u3}$ (3)	0.007 (1)	O2 $T_{1u}$ (3)	0.030 (4)
O1 $T_{2g3}$ (1)	0.007 (1)	O1 $T_{1g1}$ (1)	0.007 (1)		
O1 $T_{2g3}$ (2)	0.007 (1)	O1 $T_{1g1}$ (2)	0.007 (1)		
O1 $T_{2g3}$ (3)	0.007 (1)	O1 $T_{1g1}$ (3)	0.007 (1)		
O1 $T_{1u1}$ (1)	0.008 (1)	O1 $T_{1g2}$ (1)	0.007 (1)		
O1 $T_{1u1}$ (2)	0.008 (1)	O1 $T_{1g2}$ (2)	0.007 (1)		
O1 $T_{1u1}$ (3)	0.008 (1)	O1 $T_{1g2}$ (3)	0.007 (1)		

these coefficients are tabulated in Table 3 (the coefficients are averaged across modes related by face centering, over all  $\mathbf{k}$  and across 320 distinct ensembles). Of the Bi modes (depicted in Fig. 10), those spanning the  $E_u$  and  $T_{2u}$  irreducible representations generate displacements that reproduce the toroidal distribution of Bi positions observed by Shoemaker *et al.* (2010) and have the most significant amplitude. The mode spanning the  $A_{2u}$  representation is orthogonal to the  $C_\infty$  rotational axis of the torus and has a small amplitude. The modes spanning the  $T_{1u}$  (1) and  $T_{1u}$  (2) representations have intermediate orientations and amplitudes. The decomposition

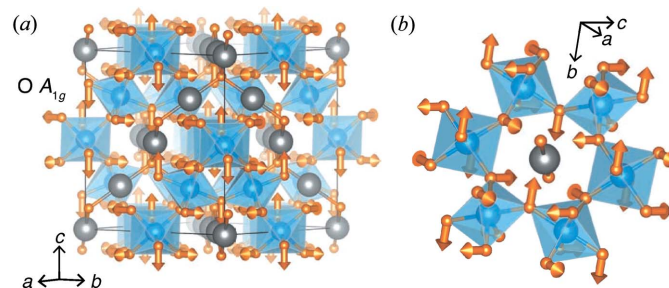

**Figure 11**

(a) Visualization of the tight-binding modes corresponding to the Bi  $T_{2u}$  modes of  $\text{Bi}_2\text{Ti}_2\text{O}_7$ . (b) The Bi  $T_{2u}$  mode set, showing the Bi atom displaced from its nominal  $180^\circ$   $\text{O}'\text{-Bi-O}'$  angle, orthogonal to the linear axis; this is reflected in the small magnitude of the Bi  $A_{2u}$ -derived modes.

of atomic displacements into tight-binding modes reproduces the physically meaningful and intuitive results presented by Shoemaker *et al.* (2010); here, the averaging across many wavevectors and simulations identifies the robustness of these displacements.

Furthermore, identification of these high-amplitude modes allows one to create a ‘small-box’ model for a symmetry-constrained refinement. In work by Shoemaker *et al.* (2010) and Fennie *et al.* (2007), imaginary phonon modes were discovered at the Brillouin zone center; the symmetries of the polarization eigenvectors belong to the  $T_{1u}$  and  $E_u$  irreducible representations (Fig. 11). The tight-binding modes spanning these irreducible representations have high-amplitude coefficients in the analysis performed here. Distortion of the  $Fd\bar{3}m$  lattice along these polarization modes yields a small unit cell of  $Cm$  symmetry that provides an excellent description of  $G(r)$  for  $r < 3.5 \text{ \AA}$  (Shoemaker *et al.*, 2010). The agreement of the high-amplitude tight-binding modes with the theory-predicted distortion modes and small-box refinement illustrates another utility of this approach for unknown systems.

Additionally, the representational analysis performed here suggests that there are correlated Bi displacements, as well as correlated  $\text{O-Bi-O}$  displacements that are not immediately observed from direct inspection of the atomic displacements. While a correlation between the Bi and  $\text{O}'$  displacements was made previously (Shoemaker *et al.*, 2010), the tight-binding mode amplitudes show that there are large displacements of Bi and O. The highest modes corresponding to the  $48f$  O atom relate to the O  $A_{1g}$  irreducible representation, which can be described as a subtle elongation and twisting of the  $[\text{TiO}_6]$  octahedron (Fig. 12a). This large displacement is also mirrored in the anisotropic atomic displacement parameter of the  $48f$  O-atom position obtained from Rietveld analysis. A possible origin of the high amplitude of this distortion is illustrated in Fig. 12(b): the  $48f$  O-atom positions form a hexagon encircling the linear  $\text{O}'\text{-Bi-O}'$  linkages in an orthogonal orientation. With significant Bi displacements, as indicated by the large amplitude of the Bi  $T_{2u}$  spanning modes, the O atoms are displaced from their ideal positions around the hexagon in order to accommodate the shifted Bi atoms.


**Figure 12**

(a) The O  $A_{1g}$  representation, carrying the largest magnitude of all degrees of freedom for the  $48f$  O atoms. (b) Vectors depicting the O  $A_{1g}$  mode surrounding a linear  $\text{O}'\text{-Bi-O}'$  group illustrate how the O atoms move out of the plane in which the modes of the Bi  $T_{2u}$  representation act (as illustrated in Fig. 11). While all vectors are illustrated, not all of the atoms shown are spanned by a single irreducible representation (*i.e.* those related by face centering are included).

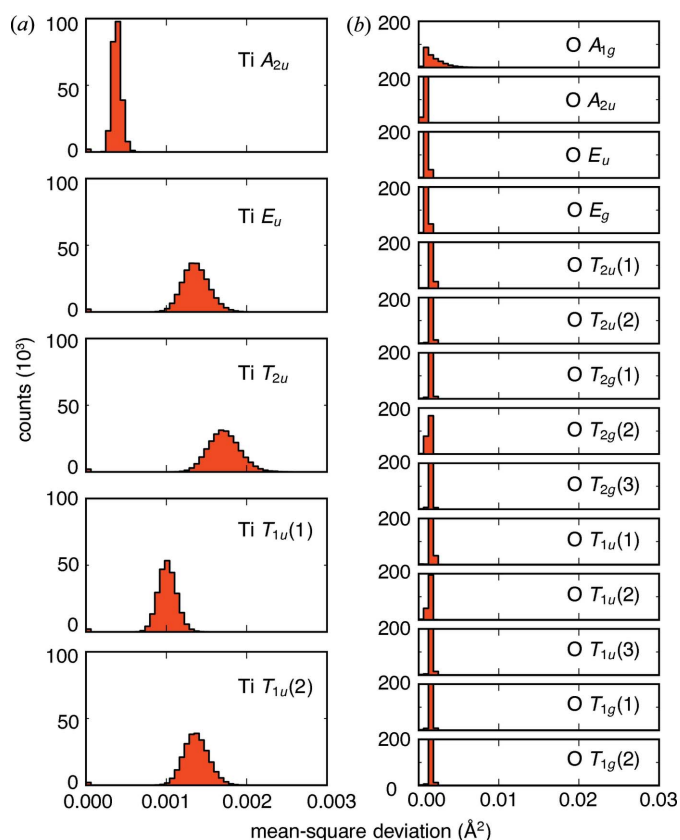
By comparing different simulation runs, it is possible to gauge the uncertainty in how distorted or ideal the connectivity is in different parts of the lattice. For example, the mode amplitudes corresponding to the Ti–O sublattice are shown in Fig. 13. The histograms show that the displacement amplitudes have a narrow distribution across all length scales and between many simulation runs. Furthermore, the displacements appear to be reasonably isotropic, as consistent with thermal disorder of a cubic lattice.

In contrast, the distribution of Bi-atom displacements is varied (Fig. 14). The Bi  $A_{2u}$  mode distribution is comparable to the Ti–O sublattice. However, many of the Bi displacement modes corresponding to multidimensional irreducible representations have high amplitudes and broad distributions, indicative of substantial static disorder in directions orthogonal to the linear O'–Bi–O' bond axis. This strongly suggests that those displacement modes locally break the  $Fd\bar{3}m$  symmetry of the crystal structure.

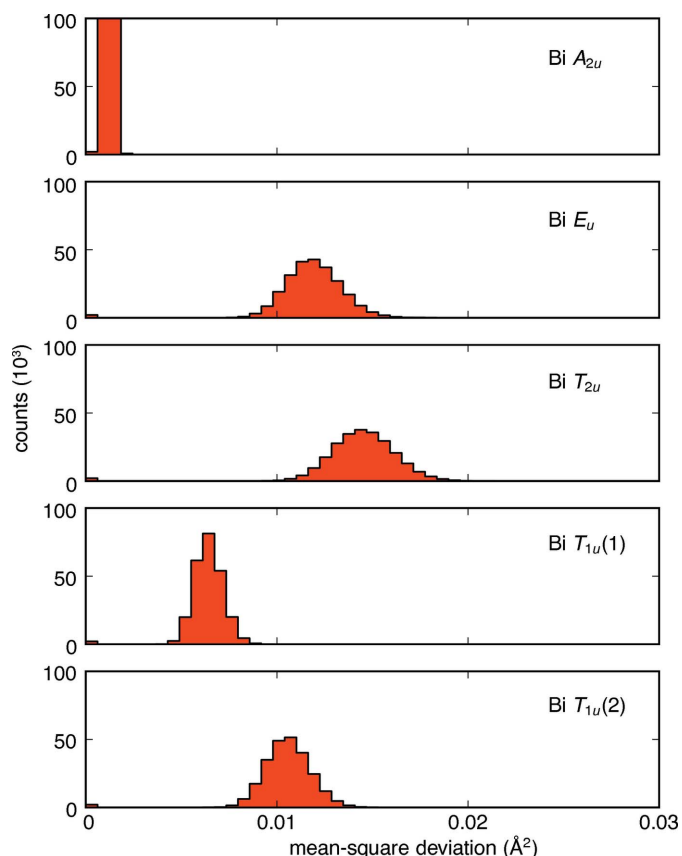
In this analysis, the multidimensional irreducible representations are broken into their individual components (so as to retain the total number of degrees of freedom); however, the values of each tight-binding mode are identical in the case of  $\text{Bi}_2\text{Ti}_2\text{O}_7$ . Then, to identify if and by how much the atomic displacements break the symmetry elements linking together tight-binding modes spanning a multidimensional irreducible

representation, the continuous symmetry measure of each irreducible representation can be calculated. Fig. 15 contains histograms of the MSDs for each irreducible representation after operation on the simulation box by a specific symmetry operation (equation 4). The modes spanning the  $A_{2u}$  representation do not show any dependence on the symmetry operation, while the modes corresponding to the  $E_u$  and  $T_{2u}$  representations do show a dependence on the operations. Specifically, the face-centering  $[(\frac{1}{2}, 0, \frac{1}{2}), (0, \frac{1}{2}, \frac{1}{2})]$  and inversion ( $i$ ) symmetry operations show the highest MSDs as well as the broadest distributions, suggesting that those symmetries deviate by the largest magnitude and in the most ways. In future work, it will be informative to analyze the compatibility relationships as the degeneracy of different modes changes as  $\mathbf{k} \neq (0, 0, 0)$ .

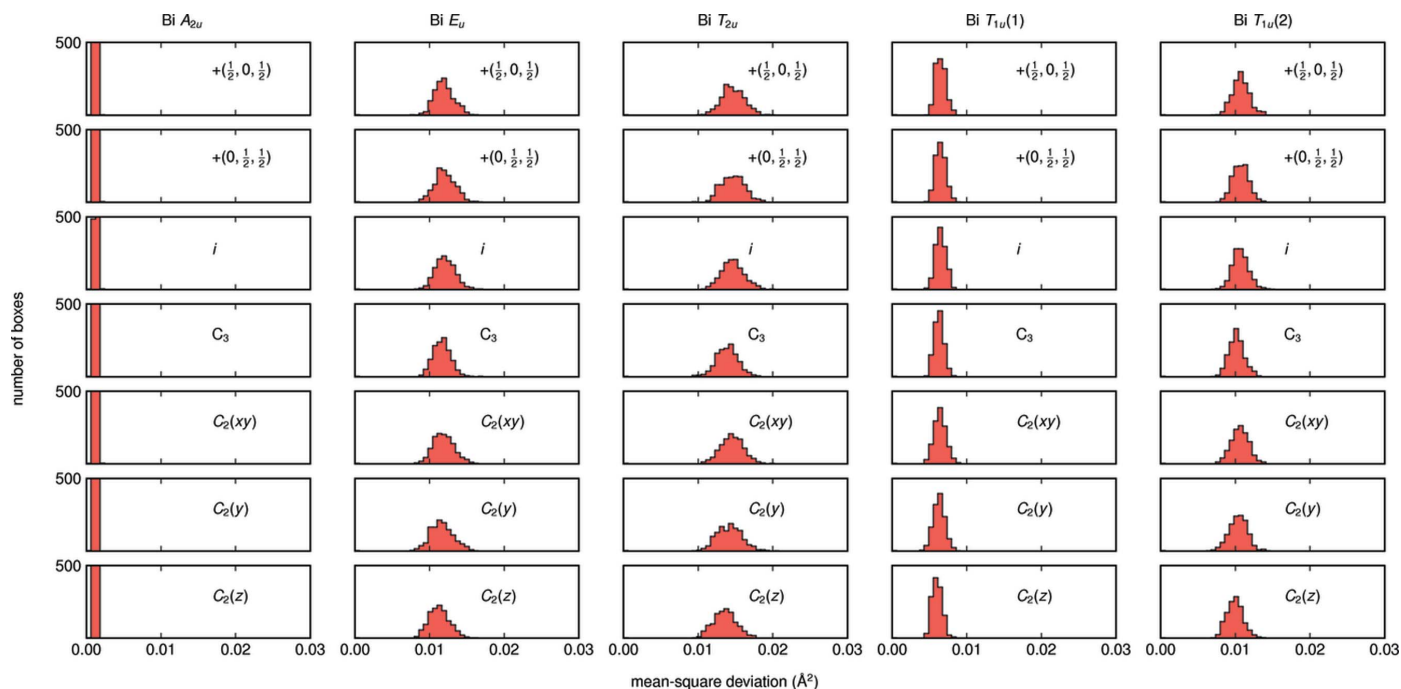
The crystal structure of  $\text{Bi}_2\text{Ti}_2\text{O}_7$  presents a very complex problem as the unit cell contains 88 atoms, resulting in 264 degrees of freedom or 264 distinct tight-binding modes to describe all atom displacements. When trying to analyze a large ensemble simulation of this structure, analysis in Cartesian coordinates becomes unwieldy. Decomposition of the structure into the crystallographically relevant local basis allows one to determine the highest amplitude disorder in the lattice, the distribution of amplitudes, the direction of the



**Figure 13** Histograms of the mean-square displacements of the (a) Ti and (b) O mode sets, summed over all  $\mathbf{k}$ , all symmetry operations and 320 ensembles, illustrating that the Ti–O sublattice is not disrupted. Only the O  $A_{1g}$  mode shows any distribution of the MSD.



**Figure 14** Histograms of the mean-square displacements of the Bi mode sets, summed over all  $\mathbf{k}$ , all symmetry operations and 320 ensembles, illustrating that the  $E_u$ ,  $T_{2u}$  and  $T_{1u}(2)$  representations break the symmetry operations of  $Fd\bar{3}m$ ; those modes correspond to the toroidal displacements shown in Fig. 10.


**Figure 15**

Histograms of the mean-square displacements of the Bi mode sets for each symmetry operation of  $Fd\bar{3}m$ , summed over all  $\mathbf{k}$  and 320 ensembles. The amplitudes and breadths of the MSDs show that the face-centering  $[\frac{1}{2}, 0, \frac{1}{2}]$ ,  $+(0, \frac{1}{2}, \frac{1}{2})$  and inversion ( $i$ ) symmetry operations are not well followed by the  $E_u$ ,  $T_{2u}$  and  $T_{1u}(2)$  representations.

atomic displacements causing the disorder, and how the disorder breaks specific symmetry elements of the crystallographic space group and by how much.

#### 4. Conclusions

The representational analysis of large atomistic ensembles generated from simulations (such as from reverse Monte Carlo simulations of total scattering data) using a tight-binding basis derived from locally symmetry-adapted modes is a robust method that allows one to quantify disorder in the lattice. In many RMC simulations, the goal is often to characterize subtle deviations from the lattice. These types of displacement are subtle perturbations from a lattice that possesses a modicum of moderately isotropic thermal disorder. Therefore, isolation and quantification of the disorder (*i.e.* of infrequent events) requires statistical analysis. By representing the disorder with respect to a local basis of the background signal (*i.e.* symmetry-adapted modes of the crystallographic space group), displacements appear as a positive signal, are amplified and can be analyzed statistically. Additionally, the approach presented here permits a framework for analyzing other types of degrees of freedom, such as occupational/compositional disorder (*e.g.* solid solutions) or magnetism. Such a rigorous group-theoretical treatment is currently implemented in *ISODISPLACE* (Campbell *et al.*, 2006).

#### Acknowledgements

The source code used in this analysis is freely available at <https://occamy.chemistry.jhu.edu/references/pubsoft/index>.

php. The authors thank D. P. Shoemaker, K. Page and R. Seshadri for sharing their neutron scattering data for this analysis and for helpful discussions. This research was principally supported by the US DOE, Office of Basic Energy Sciences (BES), Division of Materials Sciences and Engineering, under award No. DE-FG02-08ER46544. TMM acknowledges support from the David and Lucile Packard Foundation. This work benefited from the use of the NPDF at the Lujan Center at Los Alamos Neutron Science Center, funded by DOE BES. Los Alamos National Laboratory is operated by Los Alamos National Security LLC under DOE contract No. DE-AC52-06NA25396. The upgrade of the NPDF was funded by the National Science Foundation through grant No. DMR 00-76488. This research utilized the CSU ISTeC Cray HPC System supported by NSF grant No. CNS-0923386.

#### References

- Alvarez, A., Alemany, P. & Avnir, D. (2005). *Chem. Soc. Rev.* **34**, 313–326.
- Aroyo, M. I., Kirov, A., Capillas, C., Perez-Mato, J. M. & Wondratschek, H. (2006). *Acta Cryst. A* **62**, 115–128.
- Aroyo, M. I., Perez-Mato, J. M., Capillas, C., Kroumova, E., Ivantchev, S., Madariaga, G., Kirov, A. & Wondratschek, H. (2006). *Z. Kristallogr.* **221**, 15–27.
- Aroyo, M. I., Perez-Mato, J. M., Orobengoa, D., Tasci, E., de la Flor, G. & Kirov, A. (2011). *Bulg. Chem. Commun.* **43**, 183–197.
- Billinge, S. J. L. & Levin, I. (2007). *Science*, **316**, 561–565.
- Božin, E., Schmidt, M., DeConinck, A., Paglia, G., Mitchell, J., Chatterji, T., Radaelli, P., Proffen, Th. & Billinge, S. J. L. (2007). *Phys. Rev. Lett.* **98**, 137203.
- Campbell, B. J., Stokes, H. T., Tanner, D. E. & Hatch, D. M. (2006). *J. Appl. Cryst.* **39**, 607–614.

- Dimitrov, D. A., Louca, D. & Röder, H. (1999). *Phys. Rev. B*, **60**, 6204–6207.
- Dixon, J. A. & Elliott, S. R. (2014). *Appl. Phys. Lett.* **104**, 141905.
- Egami, T. & Billinge, S. J. L. (2012). *Underneath The Bragg Peaks: Structural Analysis of Complex Materials*, 2nd ed. Oxford: Pergamon Press, Elsevier.
- Elliott, S. R. (1984). *Physics of Amorphous Materials*. London: Longman.
- Farrow, C. L., Juhas, P., Liu, J. W., Bryndin, D., Božin, E. S., Bloch, J., Proffen, T. & Billinge, S. J. L. (2007). *J. Phys. Condens. Matter*, **19**, 335219.
- Fennie, C. J., Seshadri, R. & Rabe, K. M. (2007). *arXiv*, 0712.1846.
- Goodwin, A. L., Tucker, M. G., Cope, E. R., Dove, M. T. & Keen, D. A. (2005). *Phys. Rev. B*, **72**, 214304.
- Goodwin, A. L., Tucker, M. G., Dove, M. T. & Keen, D. A. (2004). *Phys. Rev. Lett.* **93**, 075502.
- Hector, A. L. & Wiggin, S. B. (2004). *J. Solid State Chem.* **177**, 139–145.
- Hovestreydt, E., Aroyo, M., Sattler, S. & Wondratschek, H. (1992). *J. Appl. Cryst.* **25**, 544.
- Keen, D. A. & Goodwin, A. L. (2015). *Nature*, **521**, 303–309.
- Kerman, S., Campbell, B. J., Satyavarapu, K. K., Stokes, H. T., Perselli, F. & Evans, J. S. O. (2012). *Acta Cryst. A* **68**, 222–234.
- King, G., Abakumov, A. M., Woodward, P. M., Llobet, A., Tsirlin, A. A., Batuk, D. & Antipov, E. V. (2011). *Inorg. Chem.* **50**, 7792–7801.
- King, G., Thompson, C. M., Greedan, J. E. & Llobet, A. (2013). *J. Mater. Chem. A*, **1**, 10487–10494.
- Kovalev, O. V. (1993). *Representations of the Crystallographic Space Groups: Irreducible Representations, Induced Representations and Corepresentations*, 2nd ed, edited by H. T. Stokes & D. M. Hatch. Reading: Gordon and Breach Science Publishers.
- Kroumova, E., Aroyo, M. I., Perez-Mato, J. M., Kirov, A., Capillas, C., Ivantchev, S. & Wondratschek, H. (2003). *Phase Transitions*, **76**, 155–170.
- Kwei, G. H., Billinge, S. J. L., Cheong, S. W. & Saxton, J. G. (1995). *Ferroelectrics*, **164**, 57–73.
- McGreevy, R. L. & Pusztai, L. (1988). *Mol. Simul.* **1**, 359–367.
- Megaw, H. D. (1945). *Nature*, **155**, 484–485.
- Megaw, H. D. (1973). *Crystal Structures: A Working Approach*. Philadelphia: Saunders.
- Neilson, J. R., Llobet, A., Stier, A. V., Wu, L., Wen, J. J., Tao, J., Zhu, Y., Tesanovic, Z. B., Armitage, N. P. & McQueen, T. M. (2012). *Phys. Rev. B*, **86**, 054512.
- Neilson, J. R., McQueen, T. M., Llobet, A., Wen, J. J. & Suchomel, M. R. (2013). *Phys. Rev. B*, **87**, 045124.
- Page, K., Proffen, T., Niederberger, M. & Seshadri, R. (2010). *Chem. Mater.* **22**, 4386–4391.
- Palin, E. J., Dove, M. T., Redfern, S. A. T., Ortega-Castro, J., Sainz-Díaz, C. I. & Hernández-Laguna, A. (2014). *Int. J. Quantum Chem.* **114**, 1257–1286.
- Proffen, Th. & Billinge, S. J. L. (1999). *J. Appl. Cryst.* **32**, 572–575.
- Ravel, B., Stern, E. A., Vedrinskii, R. I. & Kraizman, V. (1998). *Ferroelectrics*, **206**, 407–430.
- Renninger, A. L., Reichtin, M. D. & Averbach, B. L. (1974). *J. Non-Cryst. Solids*, **16**, 1–14.
- Rodriguez-Carvajal, J. (2001). *Mater. Sci. Forum*, **378–381**, 268–273.
- Seshadri, R. (2006). *Solid State Sci.* **8**, 259–266.
- Shoemaker, D. P. & Seshadri, R. (2010). *Phys. Rev. B*, **82**, 214107.
- Shoemaker, D. P., Seshadri, R., Hector, A. L., Llobet, A., Proffen, Th. & Fennie, C. J. (2010). *Phys. Rev. B*, **81**, 144113.
- Slater, J. C. & Koster, G. F. (1954). *Phys. Rev.* **94**, 1498–1524.
- Stokes, H. T., Campbell, B. J. & Hatch, D. M. (2013). *ISOTROPY*, Version 9.3. Brigham Young University, Provo, Utah, USA. <http://iso.byu.edu>.
- Tucker, M. G., Dove, M. T. & Keen, D. A. (2001). *J. Appl. Cryst.* **34**, 630–638.
- Tucker, M. G., Keen, D. A., Dove, M. T., Goodwin, A. L. & Hui, Q. (2007). *J. Phys. Condens. Matter*, **19**, 335218.
- White, C. E., Provis, J. L., Proffen, Th., Riley, D. P. & van Deventer, J. S. J. (2010a). *Phys. Chem. Chem. Phys.* **12**, 3239–3245.
- White, C. E., Provis, J. L., Proffen, Th., Riley, D. P. & van Deventer, J. S. J. (2010b). *J. Phys. Chem. A*, **114**, 4988–4996.
- Wills, A. S. (2002). *Appl. Phys. Mater. Sci. Process.* **74**, s856–s858.
- Wu, L., Klie, R. F., Zhu, Y. & Jooss, Ch. (2007). *Phys. Rev. B*, **76**, 174210.
- Young, C. A. & Goodwin, A. L. (2011). *J. Mater. Chem.* **21**, 6464–6476.
- Zabrodsky, H., Peleg, S. & Avnir, D. (1992). *J. Am. Chem. Soc.* **114**, 7843–7851.