PDFfit2 and PDFgui: Computer programs for studying nanostructure in crystals

C. L. Farrow[1], P. Juhas[1], J. W. Liu[1], D. Bryndin[1], J. Bloch[2], Th. Proffen[3] and S. J. L. Billinge[1]

[1]Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824-2320, USA

[2]Institute for Theoretical Physics, University of Regensburg, 93040 Regensburg, Germany

[3]Lujan Neutron Scattering Center, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

E-mail: billinge@pa.msu.edu

J. Phys.: Condens. Matter, to be published

Abstract. PDFfit2 is a program as well as library for real-space refinement of crystal structures. It is capable of fitting a theoretical three dimensional structure to atomic pair distribution function data and is ideal for nanoscale investigations. The fit system accounts for lattice constants, atomic positions and anisotropic thermal factors, correlated atomic motion, and experimental factors that may affect the data. The atomic positions and thermal coefficients can be constrained to follow symmetry requirements of arbitrary space group. The PDFfit2 engine is written in C++ and accessible via Python, allowing it to inter-operate with other Python programs. PDFgui is a graphical interface built on the PDFfit2 engine. PDFgui organises fits and simplifies many data analysis tasks, such as configuring and plotting multiple fits. PDFfit2 and PDFgui are freely available via the Internet.

1. Introduction

Increasingly materials scientists and chemists are interested in complex materials which have structure on the nanometer length-scale. Examples are discrete nanoparticles [1], crystals with embedded nanoscale structures [2, 3, 4], and nanoporous materials with molecules or nanoparticles intercalated inside [5, 6]. It is difficult to study structure on the nanoscale quantitatively using diffraction methods because the diffraction gives no sharp Bragg-peaks but only broad diffuse features in the scattering. Total scattering, in which both Bragg and diffuse components of the scattering are analysed together, is growing in popularity for the study of this kind of problem [2, 7]. One approach to analyse the data is to use reverse Monte Carlo [8], the subject of this journal supplement. An alternative approach is to Fourier transform the data to real-space to obtain the atomic pair distribution function (PDF) and analyse the data directly in real-space [9].

A popular approach for this is to use the profile fitting refinement program PDFfit [10, 11]. This approach assumes that the structure can be described by a relatively small number of atoms in a "unit cell", which may be the crystallographic unit cell or it may be a supercell to accommodate symmetry lowering local distortions. The refined parameters are in direct analogy with those determined crystallographically, lattice parameters and unit cell angles, atomic fractional coordinates and anisotropic thermal factors; however, they correspond to the relevant values in the *local* structure. This allows direct comparison between the local structure in a distorted crystal and the average structure determined from the Bragg peaks alone. In the absence of nanoscale disorder, the values determined crystallographically and those determined from total scattering studies will be the same. This is no longer the case in the interesting materials where the local structure disagrees with the average structure [2, 12, 7]. Here we report a significant redesign and modification of the successful software program PDFfit [13], including a new user-friendly graphical user interface (GUI) with labour-saving data and structure plotting capabilities.

1.1. PDF Technique

The pair distribution function method was originally developed to study structure of materials with no long-range order such as liquids and glasses [14]. The technique is now increasingly applied for investigations of disorder in crystalline materials and for studies of nanomaterials [2]. It is complementary to traditional crystallographic analysis, which assumes perfect periodic material and considers only intensities from Bragg reflections. The PDF method does not require periodic order in the sample - instead it uses the whole measured spectrum, including Bragg and diffuse components, to extract the total scattering structure function, $S(Q)$, which contains coherent scattering intensities from the material [9].

The experimental atomic pair distribution function, $G(r)$, is obtained by Fourier

transformation of $S(Q)$,

$$G(r) = \frac{2}{\pi} \int_0^\infty Q[S(Q) - 1] \sin(Qr)\, \mathrm{d}Q, \tag{1}$$

where Q is the magnitude of the scattering vector. For elastic scattering $Q = 4\pi \sin(\theta)/\lambda$ with 2θ being the scattering angle and λ the wavelength of the measured radiation. The $G(r)$ function is related to the atomic structure through the relation,

$$G(r) = 4\pi r \left[\rho(r) - \rho_0\right], \tag{2}$$

where ρ_0 is the atomic number density of the material and $\rho(r)$ is the atomic pair density which gives the mean weighted density of neighbour atoms at radial distance r from an atom at the origin. Thus,

$$\rho(r) = \frac{1}{4\pi r^2 N} \sum_i \sum_{j \neq i} \frac{b_i b_j}{\langle b \rangle^2} \delta(r - r_{ij}), \tag{3}$$

where the indices i and j go over all N atoms in the sample, b_i is the scattering factor of atom i, $\langle b \rangle$ is the average scattering factor and r_{ij} is the distance between atoms i and j. Full details on the determination of an experimental PDF can be found elsewhere [9].

For a known structure model the experimental PDF can be calculated using

$$G_{calc}(r) = \frac{1}{Nr} \sum_i \sum_{j \neq i} \left[\frac{b_i b_j}{\langle b \rangle^2} \delta(r - r_{ij}) \right] - 4\pi r \rho_0. \tag{4}$$

There are many methods of extracting structural information from measured PDFs. These range from simple analysis of peak positions, widths and intensities to extract bonding information directly [9, 15, 16] reverse Monte Carlo analysis of large structure models [17, 18], emerging *ab-initio* structure solution methods [19], to the common down-hill least-square refinement as implemented in the PDFfit program [11, 10].

2. PDFfit2

PDFfit2 is a major upgrade to PDFfit [10], and inherits many of its features. PDFfit is capable of fitting a theoretical three-dimensional structure to an experimentally determined PDF. It can simultaneously fit multiple structures, accounting for different structural phases in a material. PDFfit has a constraint system that allows expressing structure variables as simple functions of fitted parameters. PDFfit structure variables include lattice constants, data and phase scale factors, atomic site occupation, anisotropic thermal factors, and atomic vibrational correlations. PDFfit has a built-in FORTRAN-style command language that understands simple `for` loops and some built in arithmetic functions.

PDFfit is written in FORTRAN-77, which imposes some limitations on the program. For example, it uses fixed-size arrays for internal storage. This precludes the analysis of structures with large cells without modifying the code. Though the constraint system is powerful, it requires that a constraint equation be accompanied by its first derivative. This places the burden of determining the derivatives on the user,

which can introduce errors that lead to instability in the convergence. Furthermore, the code is monolithic, not easily extensible and hard to integrate with external programs.

The primary focus of PDFfit2 development was to remedy the limitations of PDFfit while extending its functionality. The old PDFfit engine has been completely rewritten in C++, and many bugs have been fixed. The new engine uses dynamic memory allocation so that the size of the structure or extent of the fit-range of the PDF is limited only by the physical memory available. The constraint system has also been upgraded. The program *automatically* computes the analytical derivatives of the constraints that are required by the minimisation routine. This simplifies user input and reduces the possibility of errors. In addition, new fitting parameters for handling dynamic atomic correlations and experimental resolution have been introduced as well.

Instead of rewriting the PDFfit command interpreter, which is used to define the fitting problem and to control and run the refinement, its functions are carried out using the Python language. Python is a powerful, cross-platform, open-source interpreted programming language (i.e., it does not need to be compiled to run, similar to scripting) that emphasises object-oriented and modular design. PDFfit2 scripts written in Python syntax take the place of PDFfit macros and the Python interpreter can handle everything that the old interpreter could, and more. Using Python as an interpreter allows PDFfit2 to be combined with and enhanced by other Python libraries. We make use of this capability with PDFgui as described below.

3. PDFgui

The PDFfit2 engine can be used either directly from Python command line, or as part of larger and more complex software applications. The first application built on PDFfit2 is PDFgui, a graphical environment for PDF fitting.

3.1. Design Principles

PDFgui has been designed to provide users with an easy-to-use yet powerful interface for fitting structure models to PDF data. It makes use of an object oriented, component based architecture, which makes it highly extensible and maintainable. This allows for powerful usability features such as real-time plotting and remote execution of the fitting program whilst visualising the results locally. Highlights in the design include:

- **Python** PDFgui is written in the Python programming language. Python features a relaxed and friendly syntax, supports "write once, run anywhere" portability, and has extensive libraries and modules for virtually every task. Software codes written in a variety of programming languages can be bound into Python, which allows them to be used together. Python is becoming a popular choice in the scientific computation community.
- **GUI** PDFgui's interface is built using wxPython, the Python package for wxWidgets, a mature cross-platform GUI library. Graphical applications written

in wxPython provide a look and feel native to the platform on which they are run.

- **Portability** Portability is a crucial requirement for PDFgui. PDFgui is designed to run on Windows, Mac OS, Linux, and all major Unix systems. Portability of PDFgui is assured by Python and the chosen dependencies.
- **Threading** PDFgui has been designed with multitasking in mind. It is multi-threaded so that the work being done by the PDFfit2 engine does not interfere with the tasks of the user interface. This allows for dynamic features such as real-time plotting.

3.2. Capabilities

PDFgui contains all of the functionality of PDFfit2 along with additional enhancements for usability. Mundane tasks are handled by the program and difficult tasks are made simple. Selected features of PDFgui include:

- **Project management** A PDFgui project can manage multiple fits. Each fit can have multiple experimental data sets and structure models. All the initial, final, and intermediate data are stored in a platform independent project file that can be loaded on any computer. PDFgui supports all common management tasks, such as adding, removing, modifying, copying, and pasting of fits, phases, and data sets.
- **Fully graphical** PDFgui is a fully graphical application. All tasks can be done through graphical interface. This includes fit creation, import and export of data and results, fit configuration and editing, structure creation and visualisation, and data plotting.
- **Crystal symmetry** PDFgui supports space group operations. Users can define an asymmetric unit and let PDFgui expand it to a full cell with all symmetry related positions. Another feature is automatic generation of symmetry constrains for atom positions and anisotropic thermal displacements. Users just need to specify the space group, and the program will identify equivalent sites and generate constraint equations for their coordinates and temperature factors to keep the structure consistent with symmetry requirements. This can be done either for all atoms in the structure or for an arbitrary subset - for example when it is known that only certain species show local distortions. The code for space group definitions was provided by the Python Macromolecular Library (mmLib) [22]. PDFgui also supports supercell expansion of a normal unit cell.
- **Plotting** PDFgui is using matplotlib Python package for 2D plotting of data and results. It has a friendly interface so the user can quickly and easily view the results of a fitting. PDFgui lets users
  - Plot data from a series of fits and plot it against selected metadata (temperature, doping, etc.).
  - Plot the results of several fits in the same window.
  - Plot the PDF in real time as the fitting is running.

  - Plot the parameters or variables in real time as the refinement evolves.
  - Save plots in common image formats or export the data to a text file.
- **3D structure visualisation** PDFgui allows viewing of 3D model of the refined structure. Currently this is done by using AtomEye, a 3D visualisation tool [20]. The modular nature of the program allows for other alternatives in the future, such as PyMOL [21].
- **Distributed computation** PDFgui can be configured to run fits on a remote computer. This uses the secure-shell protocol (ssh2) so that the connections are stable and secure. In this distributed mode the core fitting computation does not compete for system resources with the graphical functions, which speeds up the fit and gives faster user response at the console.
- **Linked fitting** PDFgui supports sequential fitting. Fits in a sequence can call upon other fits for their starting parameters. Configured fits can be queued to run while the user is away.
- **Macros** PDFgui supports built-in macros for advanced fits. For example for a set of experimental data from one system at different temperatures or doping levels, PDFgui can expand template fit to a series of related fits. Another PDFgui macro makes it easy to setup boxcar fits, where the same model is fit over different r-ranges of the PDF data.

3.3. Example

The capabilities of PDFgui are demonstrated here on an example fitting of temperature series of neutron PDF data from $LaMnO_3$. The specifics of using PDFgui and PDFfit2 are detailed in the respective user manuals, available with the code.

$LaMnO_3$ has a perovskite structure consisting of corner-shared MnO_6 octahedral units [**?**, 24]. At room temperature this material takes on orthorhombic symmetry (space-group *Pbnm*) where the Jahn-Teller (JT) distorted MnO_6 octahedra contain four short and two long Mn-O bonds and the JT long-bonds are ordered in space in a checkerboard fashion [13]. As temperature is increased through 750 K, a structural phase transition occurs to a pseudo-cubic phase, without a change in space group [25]. Using neutron powder diffraction data collected at the NPDF diffractometer [26] at the Lujan Center at Los Alamos National Laboratory, PDFgui was used to create a temperature series fit to investigate the orthorhombic to pseudo-cubic phase transition. More detail about this analysis can be found in references [27] and [28]. The average crystallographic structure for $LaMnO_3$ was used as a starting point for the fits. In practise this can be done by typing in the asymmetric unit, or loading it from a CIF-format file and letting the program expand it.

Each fit in the series was part of the same project and configured identically. The La and Mn atoms were constrained to have isotropic thermal factors as were the axial and polar oxygen atoms. Using the PDFfit2 constraint mechanism allows one to explicitly include known correlations between the physical parameters of the fit and leads to less

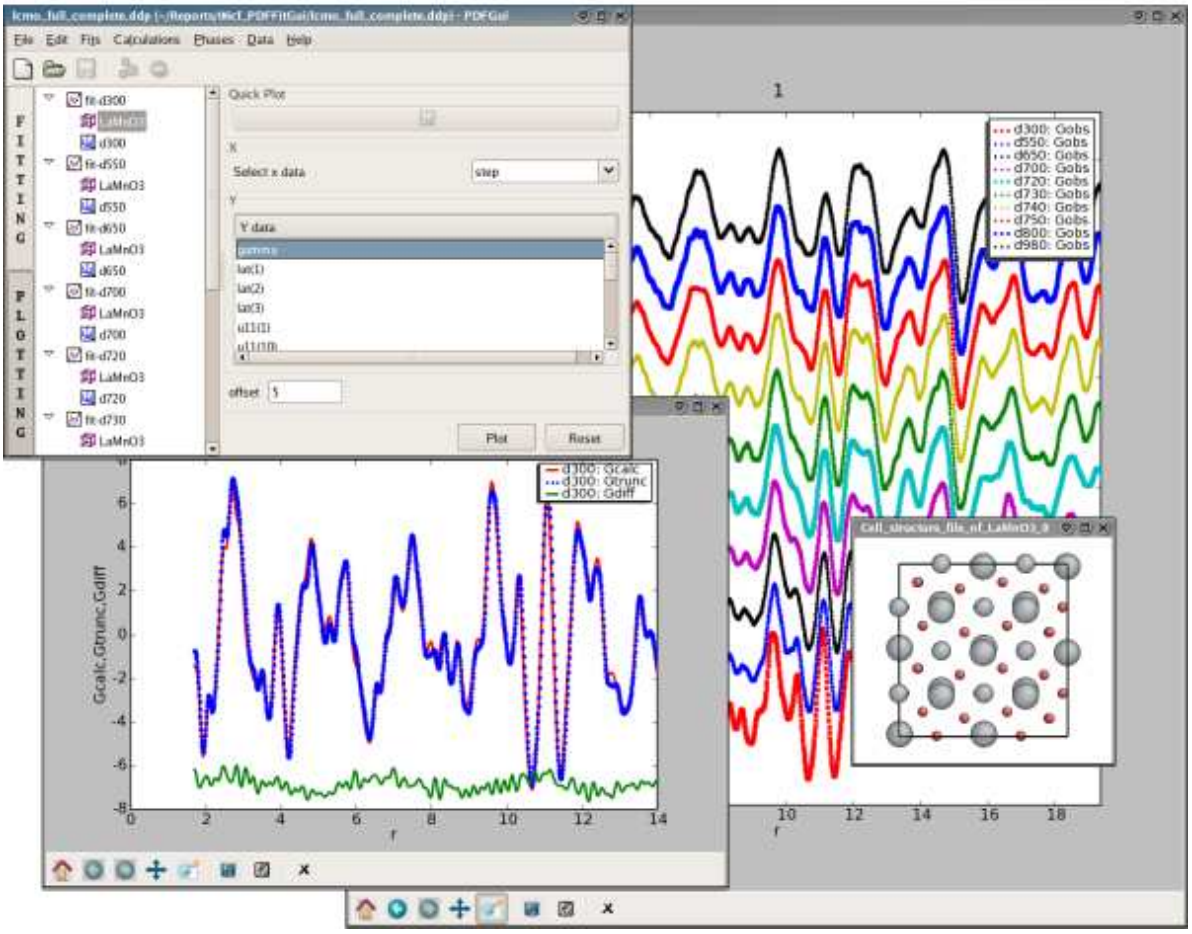

Figure 1. The PDFgui plot window, plots of the $LaMnO_3$ PDF fit at 300 K and the data at all temperatures, and the refined structure at 300 K.

uncertainty in the resulting refinement. To speed up the convergence of the fits, the starting values used for a given temperature were taken from the converged values of the previous temperature in the series. Chaining fits together in such a way places the fitting parameters in the basin of attraction for the minimum in the parameter space. Figure 1 shows the fit organisation and plot interface of PDFgui, a stacked plot of all of the project data, a plot of the 300 K fit results, and the refined structure from the 300 K fit.

One of the goals of this series of fits was to track the orientation of the MnO_6 octahedra. A major strength of PDFgui is the ability to plot any structure variable or a parameter used in a fit. This allows users to easily and creatively investigate the complex correlations among the fit parameters, without having to manually mine data out of output files. An example of this can be seen in Fig. 2. The figure shows a screen shot a PDFgui plot window displaying the refined fractional x-coordinate of one of the planar oxygens. The expected structural phase transition is clearly indicated by a spike in the value of this parameter at the transition temperature.

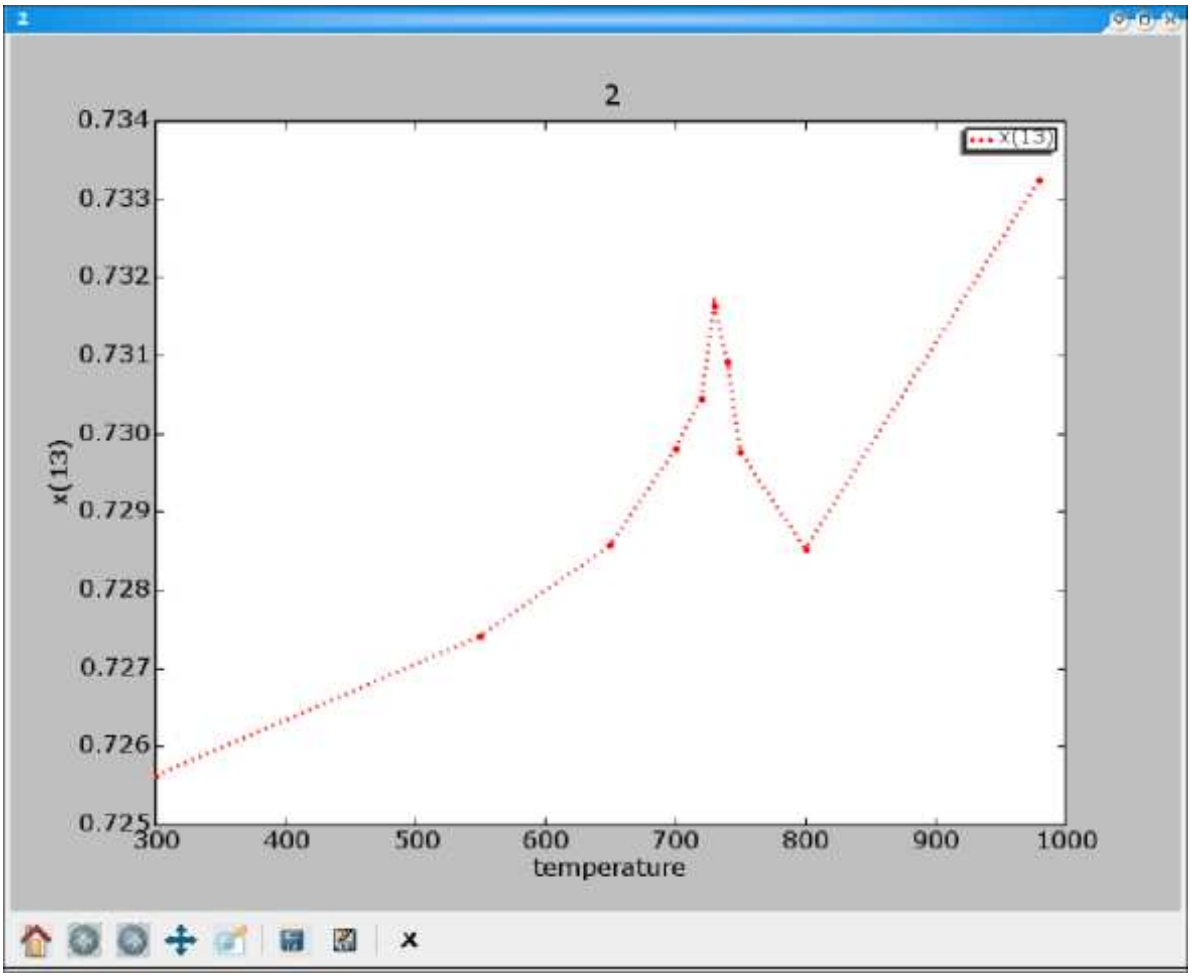

Figure 2. The full PDFgui plot window showing the fractional x-coordinate of an axial oxygen versus temperature.

4. Availability

PDFfit2 and PDFgui are open source and distributed under a BSD license. They run on Windows, Mac OS, Linux, and all major Unix systems. The source code is freely available at http://www.totalscattering.org. Some dependencies must be satisfied to compile the code from its sources, requiring Python 2.3 or later (http://www.python.org), GSL - GNU scientific library (http://www.gnu.org/software/gsl/), and a c++ compiler. PDFgui requires third-party Python packages: wxPython (http://www.wxpython.org), NumPy (http://numpy.scipi.org) and matplotlib (http://matplotlib.sourceforge.net). AtomEye (http://164.107.79.177/Archive/Graphics/A/) is required for structural visualisation.

4.1. Contacts

For more information please contact Professor Simon Billinge (billinge@pa.msu.edu) or consult the web-page http://www.totalscattering.org. News of updates and releases will

be posted at this web-site and on the total-scattering email-list. Instructions for joining the email-list can be found on the web-page.

¡¡¡¡¡¡¡ JPCM$_P DF FitGui.tex$

4.2. Citation to this paper

Since this paper has not been published, the citation to it is
C. L. Farrow, P. Juhas, J. W. Liu, D. Bryndin, J. Bloch, Th. Proffen and S. J. L. Billinge, PDFfit2 and PDFgui: Computer programs for studying nanostructure in crystals, *J. Phys: Condens. Matter*, submitted (2006)

5. Acknowledgments

=======

6. Acknowledgements

¿¿¿¿¿¿¿ 1.21

We would like to acknowledge HyunJeong Kim, He Lin, Emil Božin, and Wenduo Zhou for help and useful discussions during the development and all the Billinge-group members for help in testing. PDFgui was developed as part of the Distributed Analysis of Neutron Scattering Data (DANSE) project which is funded by NSF under DMR-0520547. The PDFfit2 development was funded by NSF under contract DMR-0304391.

References

[1] B. Gilbert, F. Huang, H. Zhang, G. A. Waychunas, and J. F. Banfield. Nanoparticles: Strained and stiff. *Science*, 305:651–654, 2004.

[2] S. J. L. Billinge and M. G. Kanatzidis. Beyond crystallography: the study of disorder, nanocrystallinity and crystallographically challenged materials. *Chem. Commun.*, pages 749–760, 2004.

[3] He Lin, E. S. Božin, S. J. L. Billinge, Eric Quarez, and M. G. Kanatzidis. Nanoscale clusters in the high performance thermoelectric $AgPb_mSbTe_{m+2}$. *Phys. Rev. B*, 72:174113, 2005.

[4] S. Vensky, L. Kienle, R. E. Dinnebier, A. S. Masadeh, S. J. L. Billinge, and M. Jansen. The real structure of Na_3BiO_4 by electron microscopy, HR-XRD and PDF analysis. *Z. Kristallogr.*, 220:231–244, 2005.

[5] V. Petkov, S. J. L. Billinge, T. Vogt, A. S. Ichimura, and J. L. Dye. Structure of intercalated Cs in zeolite ITQ-4: an array of metal ions and electrons confined in a pseudo-1D nanoporous host. *Phys. Rev. Lett.*, 89:075502, 2002. (Highlighted in Phys. Rev. Focus: http://focus.aps.org/story/v10/st4).

[6] Simon J. L. Billinge, Emily J. McKimmey, Mouath Shatnawi, HyunJeong Kim, Valeri Petkov, Didier Wermeille, and Thomas J. Pinnavaia. Mercury binding sites in thiol-functionalized mesostructured silica. *J. Am. Chem. Soc.*, 127:8492–8498, 2005.

[7] M. G. Tucker, M. P. Squires, M. T. Dove, and D. A. Keen. Dynamic structural disorder in cristobalite: neutron total scattering measurement and reverse Monte Carlo modelling. *J. Phys: Condens. Matter*, 13(3):403–423, 2001.

[8] R. L. McGreevy and L. Pusztai. Reverse Monte Carlo simulation: a new technique for the determination of disordered structures. *Mol. Simul.*, 1:359–367, 1988.

[9] T. Egami and S. J. L. Billinge. *Underneath the Bragg peaks: structural analysis of complex materials*. Pergamon Press, Elsevier, Oxford, England, 2003.

[10] Th. Proffen and S. J. L. Billinge. PDFFIT, a program for full profile structural refinement of the atomic pair distribution function. *J. Appl. Crystallogr.*, 32:572–575, 1999.

[11] S. J. L. Billinge. Real-space Rietveld: full profile structure refinement of the atomic pair distribution function. In S. J. L. Billinge and M. F. Thorpe, editors, *Local Structure from Diffraction*, page 137, New York, 1998. Plenum.

[12] M. G. Tucker, A. L. Goodwin, M. T. Dove, D. A. Keen, S. A. Wells, and J. S. O. Evans. Negative thermal expansion in ZrW_2O_8: Mechanisms, rigid unit modes, and neutron total scattering. *Phys. Rev. Lett.*, 95:255501, 2005.

[13] Th. Proffen, R. G. DiFrancesco, S. J. L. Billinge, E. L. Brosha, and G. H. Kwei. Measurement of the local Jahn-Teller distortion in $LaMnO_{3.006}$. *Phys. Rev. B*, 60:9973, 1999.

[14] B. E. Warren. *X-ray diffraction*. Dover, New York, 1990.

[15] A. C. Wright. Diffraction studies of glass structure: the first 70 years. *Glass. Phys. Chem.*, 24:148–179, 1998.

[16] I-K. Jeong, Th. Proffen, F. Mohiuddin-Jacobs, and S. J. L. Billinge. Measuring correlated atomic motion using x-ray diffraction. *J. Phys. Chem. A*, 103:921–924, 1999.

[17] R. L. McGreevy. Reverse Monte Carlo modelling. *J. Phys: Condens. Matter*, 13(46):R877–R913, 2001.

[18] B. H. Toby and T. Egami. Accuracy of pair distribution function analysis applied to crystalline and noncrystalline materials. *Acta Cryst. A*, A48(3):336–46, 1992.

[19] P. Juhas, D. M. Cherba, P. M. Duxbury, W. F. Punch, and S. J. L. Billinge. Ab initio determination of solid-state nanostructure. *Nature*, 440(7084):655–658, 2006.

[20] J. Painter and E. A. Merritt. mmLib Python toolkit for manipulating annotated structural models of biological macromolecules. *J. Appl. Crystallogr.*, 37:174–178, 2004.

[21] J. Li. AtomEye: an efficient atomistic configuration viewer. *Model. Simul. Mater. Sc.*, 11(2):173–177, 2003.

[22] DeLano W. L. *The PyMOL User's Manual*. DeLano Scientific, San Carlos, CA, USA., 2002.

[23] Jacqueline B. A. A. Elemans, B. Van Laar, K. R. Van der Veen, and B. O. Loopstra. The crystallographic and magnetic structures of $La_{1-x}Ba_xMn_{1-x}Me_xO_3$ (Me = Mn or Ti). *J. Solid State Chem.*, 3(2):238–42, 1971.

[24] J. Rodriguez-Carvajal, M. Hennion, F. Moussa, A. H. Moudden, L. Pinsard, and A. Revcolevschi. Neutron-diffraction study of the Jahn-Teller transition in stoichiometric $LaMnO_3$. *Phys. Rev. B*, 57(6):R3189–R3192, 1998.

[25] Tapan Chatterji, Francois Fauth, Bachir Ouladdiaf, P. Mandal, and B. Ghosh. Volume collapse in $LaMnO_3$ caused by an orbital order-disorder transition. *Phys. Rev. B*, 68:052406, 2003.

[26] Th. Proffen, T. Egami, S. J. L. Billinge, A. K. Cheetham, D. Louca, and J. B. Parise. Building a high resolution total scattering powder diffractometer - upgrade of NPD at MLNSC. *Appl. Phys. A*, 74:s163–s165, 2002.

[27] Xiangyun Qiu, Th. Proffen, J. F. Mitchell, and S. J. L. Billinge. Orbital correlations in the pseudocubic O and rhombohedral R-phases of $LaMnO_3$. *Phys. Rev. Lett.*, 94:177203, 2005.

[28] E. S. Božin, X. Qiu, M. Schmidt, G. Paglia, J. F. Mitchell, P. G. Radaelli, Th. Proffen, and S. J. L. Billinge. Local structural aspects of the orthorhombic to pseudo-cubic phase transformation in $La_{1-x}Ca_xMnO_3$. *Physica B*, 2006. In press.