



Received 28 July 2017

Accepted 13 September 2017

Edited by K. Chapman, Argonne National
Laboratory, USA

Keywords: local structure; reverse Monte Carlo;
scattering.

Advancing reverse Monte Carlo structure refinements to the nanoscale

M. Eremenko, V. Krayzman, A. Gagin and I. Levin*

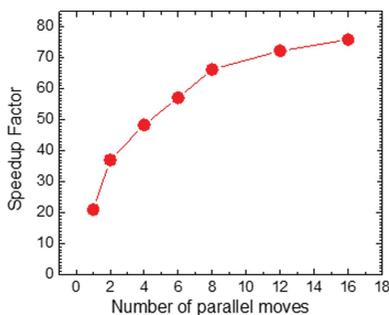
Materials Measurement Science Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. *Correspondence e-mail: igor.levin@nist.gov

Over the past decade, the *RMCPProfile* software package has evolved into a powerful computational framework for atomistic structural refinements using a reverse Monte Carlo (RMC) algorithm and multiple types of experimental data. However, realizing the full potential of this method, which can provide a consistent description of atomic arrangements over several length scales, requires a computational speed much higher than that permitted by the current software. This problem has been addressed *via* substantial optimization and development of *RMCPProfile*, including the introduction of the new parallel-chains RMC algorithm. The computing speed of this software has been increased by nearly two orders of magnitude, as demonstrated using the refinements of a simulated structure with two distinct correlation lengths for the atomic displacements. The new developments provide a path for achieving even faster performance as more advanced computing hardware becomes available. This version of *RMCPProfile* permits refinements of atomic configurations of the order of 500 000 atoms (compared to the current limit of 20 000), which sample interatomic distances up to 10 nm (*versus* 3 nm currently). Accurate, computationally efficient corrections of the calculated X-ray and neutron total scattering data have been developed to account for the effects of instrumental resolution. These corrections are applied in both reciprocal and real spaces, thereby enabling RMC fitting of an atomic pair distribution function, which is obtained as the Fourier transform of the total-scattering intensity, over the entire nanoscale distance range accessible experimentally.

1. Introduction

Exploitable properties of many functional materials (*e.g.* dielectrics, ferroelectrics, thermoelectrics, ion conductors *etc.*) are critically influenced by atomic order on the sub-nanometre and nanometre scales, which frequently differs from that described by average symmetry seen by Bragg diffraction. Correlated deviations from average symmetry are manifested in the diffuse scattering that occurs near and/or between the Bragg peaks. Thus, to determine the local atomic arrangements, one needs to convert such diffuse-intensity distributions into real-space correlations between atomic site occupancies and/or displacements. In polycrystalline materials, this problem is increasingly addressed by fitting a structural model to an atomic pair distribution function (PDF), obtained as the Fourier transform of the total (Bragg plus diffuse) X-ray/neutron diffracted intensity, to determine the relevant local-structure descriptors (Egami & Billinge, 2012; Keen & Goodwin, 2015).

In one common approach (McGreevy, 2001), a crystal in a polycrystalline sample is represented by a large atomic configuration, typically under periodic boundary conditions. During the fit, individual atomic positions and/or their chemical types are adjusted according to a reverse Monte



© 2017 International Union of Crystallography

Carlo (RMC) method until the calculated signal (*e.g.* PDF) matches the experimental data. In a version of the Metropolis algorithm, which is adopted by the existing RMC software, at each step, an atom is selected and moved at random. Subsequently, a signal of interest is calculated and compared with experimental data. If a given atomic move improves the fit, it is accepted unconditionally. Otherwise, the move can still be accepted but with a probability that decays exponentially with an increasing misfit between the calculated and experimental signals. The RMC method enables explicit treatment of local chemical disorder and atomic displacements as encountered, for example, in solid solutions. However, freedom from effective parameters comes at the expense of an under-determined problem and, therefore, the plausibility of structural models obtained using this approach depends critically on the completeness of data included in a fit.

The most developed software package for RMC structural refinements, *RMCPProfile* (Tucker *et al.*, 2007; Krayzman *et al.*, 2009; Playford *et al.*, 2014), permits simultaneous fitting of real- and reciprocal-space representations of neutron and X-ray total scattering data, Bragg intensities, extended X-ray absorption fine structure (EXAFS) datasets, and single-crystal diffuse scattering. Additionally, various constraints, including those defined by empirical interatomic potentials (Duncan *et al.*, 2016), can be imposed on atomic configurations. Modern neutron diffractometers provide PDFs that contain usable signals to interatomic distances greater than 10 nm (~ 6 nm for typical X-ray PDF setups). However, the computing speeds enabled by the existing *RMCPProfile* software limit the practical sizes of atomic configurations, so that the distance ranges probed by these refinements rarely exceed 3 nm.

In principle, an RMC approach can provide a model that consistently describes structural behavior on the local, nanometre and macroscopic scales. However, for such a description to be adequate, the atomic configuration must be significantly larger than the effective length of the relevant interatomic correlations in the system. This problem can be illustrated using materials that exhibit pseudosymmetry (*i.e.* their symmetry varies with the sampled volume). In some of these systems, such as perovskite BaTiO_3 or its solid solutions, the local symmetry averages to the macroscopic symmetry at scales as low as 1–2 nm and, therefore, the currently used configuration sizes can be sufficient for describing the whole structure (Comes *et al.*, 1968; Levin *et al.*, 2014, 2016; Senn *et al.*, 2016). In other cases, however, a crossover to the macroscopic symmetry only occurs on a larger scale. One such system is perovskite-like $\text{Na}_{1/2}\text{Bi}_{1/2}\text{TiO}_3$ (NBT) (Jones & Thomas, 2002; Rödel *et al.*, 2009), which has attracted much interest as a promising source for lead-free piezoelectrics. NBT exhibits a hierarchical structure of coherent nanodomains which arise from coexisting types of atomic displacements with dissimilar correlation lengths (Levin & Reaney, 2012). The smallest domains, 1–3 nm in size, assemble into larger domains having dimensions of 10–20 nm. The symmetry seen by Bragg powder diffraction is that of the larger nanodomain assemblages, not of the nanodomains themselves. As indicated above, the longest distances probed by the atomic

configurations used in current RMC refinements are limited to ~ 3 nm; that is, for NBT, these configurations would sample predominantly the smallest domains, in which the symmetry and the types of atomic displacements differ considerably from those probed by Bragg diffraction. Thus, simultaneous fitting of a PDF (limited to 3 nm) and a Bragg profile, as used previously (Keeble *et al.*, 2012; Aksel *et al.*, 2013), simply cannot be expected to produce a fully correct structural model. A similar challenge exists for structural determination in many relaxor ferroelectrics (Burns & Scott, 1973; Xu *et al.*, 2004), which supposedly contain polar nanoregions, 1–3 nm in size, having symmetry that is lower than the average, as well as in $\text{Pb}(\text{Zr},\text{Ti})\text{O}_3$ piezoelectrics (Zhang *et al.*, 2014).

Addressing the pseudosymmetry problem would require atomic configurations that could sample distances up to at least 10 nm; however, such configuration sizes are currently computationally prohibitive, especially if using multiple experimental datasets. Additionally, recovering complex nanoscale correlations requires a relatively large number of atomic moves, which slows convergence to the correct solution, making it unachievable in a practical time. Clearly, determination of the nanoscale structures using RMC refinements requires much faster computing than that enabled by the current *RMCPProfile* software. Larger sizes of atomic configurations are also required to model diffuse scattering near Bragg reflections in single-crystal X-ray or neutron diffraction, which could be fitted simultaneously with the powder total-scattering data, because the reciprocal space resolution of a calculated signal is inversely proportional to the configuration-box size. Fitting this diffuse scattering in three dimensions, as measurable using modern synchrotron and neutron instrumentation (*e.g.* Bosak *et al.*, 2012), requires computing speeds that are orders of magnitude faster than those permitted by the existing software. Besides nanostructured materials, the current speed limitations preclude adequate RMC modeling of crystal structures with large unit cells, as encountered, for example, in tungsten bronzes, layered compounds *etc.*

Here, we report on recent developments of *RMCPProfile* that overcome the computing-speed barriers, thereby extending the capabilities of this method to the nanoscale. The problem is twofold because, in addition to higher speed, fitting a PDF to large distances requires corrections for the limited resolution of the diffraction data, which causes progressive diminishing and broadening of PDF peaks with increasing interatomic separations. We have addressed both challenges. Speedup factors between one and two orders of magnitude have been achieved through a major optimization of the existing *RMCPProfile* software combined with the introduction of a new RMC algorithm. Concurrently, we have implemented accurate yet computationally efficient corrections for the instrument-resolution effects.

2. Computational speed in RMC refinements

We employed simulated data to benchmark the performance of the *RMCPProfile* software. The structural model adopted in

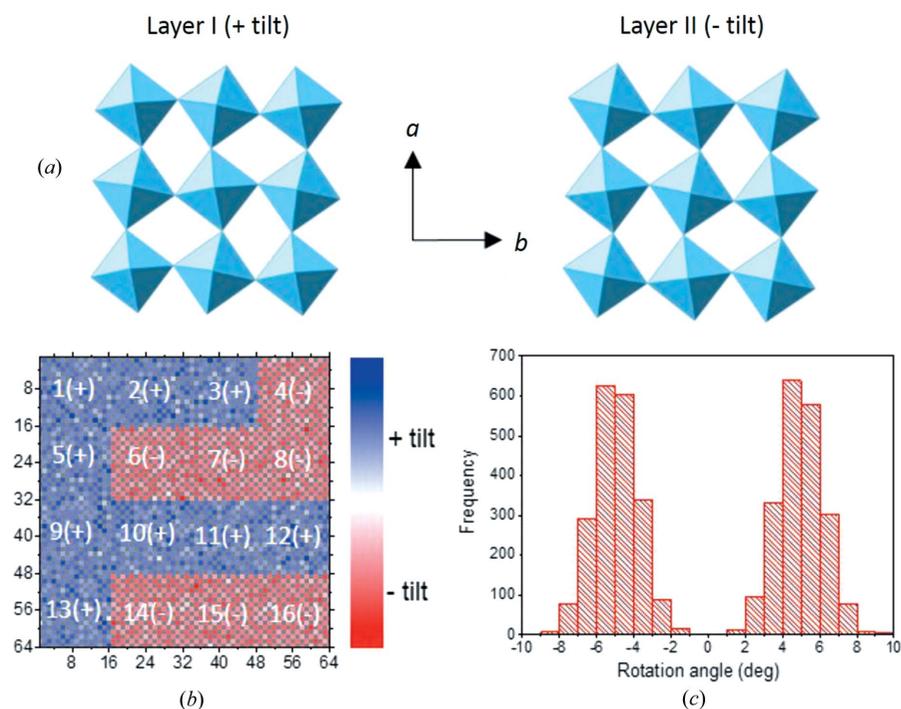


Figure 1

A schematic description of a perovskite structure with one-dimensional disorder of octahedral rotations. (a) This structure is composed of a random sequence of two types of layers, Layer I and Layer II, in which the corresponding octahedra are rotated out of phase around the c axis. (b) The 16 successive layers in this sequence are tiled side by side, with every other octahedron colored either blue (+ tilt) or red (– tilt); the remaining octahedra are colored in grey because the signs of their rotations are constrained by the connectivity of the octahedra. (c) The distribution of the tilting angles in this structure obtained using *GASP* (Wells & Sartbaeva, 2015), after the oxygen atoms have been additionally displaced according to a Gaussian distribution, as described in the text. The shades of blue and red colors reflect the magnitudes of the tilting angles.

these simulations is based on the perovskite structure, which features a three-dimensional network of vertex-sharing oxygen octahedra, having the composition CaTiO_3 . The initial configuration, which included an array of $64 \times 64 \times 64$ ideal cubic perovskite unit cells ($a = 4 \text{ \AA}$), has been modified by introducing octahedral rotations of 5° about one of the cubic axes, designated as the c axis (Fig. 1). For rigid octahedra, their connectivity imposes a chessboard rotation pattern within layers normal to the c axis, so that the neighboring octahedra are tilted in the opposite directions (Fig. 1a). However, along the c axis, octahedral rotations are unconstrained. We introduced a tilting disorder along this axis, with successive octahedra rotating randomly either clockwise or counterclockwise. The average symmetry of the resulting structure is tetragonal (the lattice parameters a and b and box dimensions have been adjusted accordingly), and the octahedral tilting pattern is illustrated schematically in Fig. 1(b). Here, 16 (out of 64) octahedral c layers (Type I and Type II) are tiled side by side into a 4×4 array, and within each plane every other octahedron is indicated using either blue (counterclockwise) or red (clockwise rotation) colored squares; the remaining half of the octahedra, the signs of rotations for which are determined by the connectivity constraints, are shown in grey. All the atoms (*i.e.* O, Ca and Ti) were additionally disordered following Gaussian distributions with the displacement para-

eters $U_{\text{O}} = 0.01 \text{ \AA}^2$, $U_{\text{Ca}} = U_{\text{Ti}} = 0.005 \text{ \AA}^2$. This model incorporates some challenges pertinent to the pseudosymmetric structures like NBT, because the octahedral tilting here also exhibits strongly distinct coherence lengths parallel and perpendicular to the c planes.

Ten independent atomic configurations of this kind were generated to obtain a statistically adequate representation of the tilting disorder and used to calculate the average neutron PDF $G(r)$, scattering function $S(Q)$ and Bragg profile, which served the role of experimental data. The resulting data were fitted using *RMCPProfile*, starting with atoms arranged as per an ideal perovskite structure, with the goal of reproducing key structural features of the target model. The number of data points for $G(r)$ and $S(Q)$ was 2842 and 2000, respectively. The minimum d spacing for the Bragg profile was set at 0.6 \AA , with 4116 reflections included in the fitted range.

The fitting was performed using two different model sizes: $16a \times 16a \times 16a$ with 20 480 atoms and $36a \times 36a \times 36a$ with 233 280 atoms. These configurations span interatomic distances of ~ 3 and $\sim 7 \text{ nm}$, respectively (because of

periodic boundary conditions imposed in *RMCPProfile*, the distance range is one-half of the smallest configuration-box dimension).

Fitting the simulated data using a 20 480-atom configuration and a current 32-bit Linux version of *RMCPProfile* running on typical computer hardware^{1,2} for nine days yielded a nearly perfect match, with statistical agreement indicators that would typically be regarded as satisfactory (Fig. 2a). Nevertheless, the patterns of octahedral rotations and distributions of the tilting angles remained grossly incorrect (Fig. 3a). A bimodal distribution of the tilting angles was recovered only after over a month of fitting (Fig. 3b), and it took more than four months for the configuration to attain a distribution with two well resolved peaks (Fig. 3c) that were reasonably, if still imperfectly, close to those expected for the target configuration.³ These slow but marked improvements in the quality of the structural model were accompanied by only tiny changes in the residual (Fig. 2d). On average, at least 10^3 moves per atom

¹ These computations have been performed using a cluster built with AMD Opteron 6136 processors, each having a base frequency of 2.4 GHz, eight cores, eight threads and 12 MB L3 cache.

² Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

³ Note that the exact sequence of the c layers in Fig. 1 does not have to be reproduced because it is random.

Table 1

Performance characteristics of the current distribution (Current) and the optimized (Optimized) and PCRMC versions of *RMCPProfile* obtained using a configuration of 20 480 atoms to fit $G(r)$ and the Bragg profile for the CaTiO_3 model with tilting disorder.

The PCRMC version has been tested using different numbers of moves (from 1 to 16) run in parallel (N_m). N_{th} refers to the number of computer threads used. The number of accepted moves required for convergence to a reasonably accurate solution was $\geq 2 \times 10^7$. For the 'Optimized' version, the fast Bragg computations account for a ~ 10 -fold increase in speed (out of the 29.78 total).

N_m	Current	Optimized	PCRMC						
	1	1	1	2	4	6	8	12	16
Time/move (ms)	12	0.403	0.571	0.324	0.248	0.21	0.181	0.166	0.158
N_{th}	1	8	8	8	8	12	16	24	32
Speedup factor	1	29.78	21.02	37.04	48.39	57.14	66.30	72.29	75.95
Time to 2×10^8 moves (h)	667	22	32	18	14	12	10	9	9

All the computations referenced in Tables 1 and 2 have been performed using a cluster built with Intel Xeon Processors E5-2687w v4, each having a base frequency of 3.0 GHz, 12 cores, 24 threads and 30 MB L3 cache.

were required to obtain a reasonably accurate representation of the target model. Therefore, in practice, when the actual structure is unknown and the data are not ideal, one would most likely terminate these refinements prematurely, producing an incorrect solution. A change from a 32-bit to a 64-bit application, which is available in the most recent distribution of *RMCPProfile*, yielded a 30% gain in speed. Expectedly, faster

which enabled faster single-thread computing. Second, we identified bottlenecks in the existing software and parallelized these computations using the Open Multi-Processing (OMP) Application Program Interface. In the parallelized version, each dataset included in the fit (*i.e.* PDF, Bragg *etc.*) can be calculated in a separate thread, which significantly reduces the total computing time; a further increase in computing speed

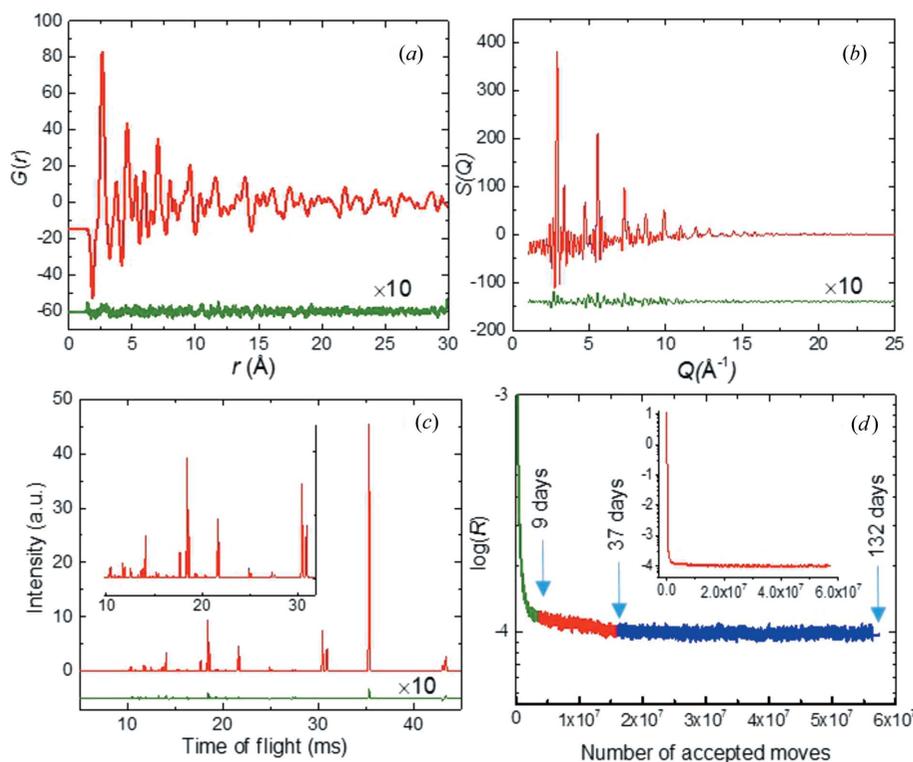


Figure 2

The experimental (red) and calculated (blue) (a) $G(r)$, (b) $S(Q)$ and (c) Bragg profile after nine days of RMC refinements, which correspond to the structural model presented in Fig. 1(a). Visual observation suggests a perfect match, as also reflected in the difference curves (green), which are shown multiplied by a factor of 10. (d) A magnified view of the dependence of the logarithm of the residual R on the number of accepted moves. The inset shows the same curve on the full scale. The changes in the residual between 37 and 132 days of fitting appear to be vanishingly small; however, the corresponding refined structures (Figs. 3b and 3c) differ markedly.

computer hardware can also considerably shorten the computing times, as demonstrated below (see the column 'Current' in Tables 1 and 2); however, even with all these improvements, the times for convergence remained prohibitively long.

2.1. Optimizing the performance of single-chain *RMCPProfile*

We enhanced the computing efficiency of *RMCPProfile* (hereafter, only 64-bit versions are considered) through several steps. First, the Fortran compiler parameters were optimized for the types of computations performed by this software and for the hardware used,

which enabled faster single-thread computing. Second, we identified bottlenecks in the existing software and parallelized these computations using the Open Multi-Processing (OMP) Application Program Interface. In the parallelized version, each dataset included in the fit (*i.e.* PDF, Bragg *etc.*) can be calculated in a separate thread, which significantly reduces the total computing time; a further increase in computing speed was obtained *via* the nested parallelization of calculations for individual datasets. Third, the algorithms used to calculate certain types of data have been revised for maximum computing efficiency. A major improvement in computing speed was obtained by exploiting the coincidences in diffraction peak positions during calculations of a Bragg pattern. In the new algorithm, the Bragg peaks are first grouped as per their unique Q -space positions, the intensities of peaks in each group are then added together and, finally, the combined intensities are multiplied by a profile function. (For our simulated tetragonal CaTiO_3 structure, the d spacings of 4116 reflections yielded 348 unique values.) For less distorted unit cells, the improvement in speed compared to the original algorithm, in which each generated peak was multiplied by a profile function separately, is nearly an order of magnitude. Note that using a Q multiplicity enables significantly more efficient calculations than could be achieved with an hkl multiplicity. The computing time for a Bragg profile depends weakly on the size of the atomic configuration.

All these optimizations combined resulted in a speedup of *RMCPProfile* by

a factor of 30 (Table 1), when tested using the same CaTiO₃ example with a configuration of 20 480 atoms while fitting the PDF and Bragg datasets. A configuration of 233 280 atoms (Table 2) could now be refined to a sensibly accurate solution [similar to that shown in Fig. 3(c)] within just a few days – a result that would be unachievable using the original non-optimized software version. Adding $S(Q)$, however, slows computations considerably. For a configuration with 233 280 atoms, the speedup factor relative to the original version is reduced to a factor of ten, compared to 24 achieved while using just PDF and Bragg; nevertheless, even with $S(Q)$ included, the speedup factors remain significant. The main reason for the substantial slowdown of calculations with $S(Q)$ for larger configurations is the increased size of the matrix of the Fourier transform (FT) coefficients, if the FT is performed over the entire distance range spanned by the configuration. The problem can be mitigated by outsourcing the matrix-multiplication operations to so-called computational accelerators [*e.g.* graphics processing units (GPUs), many integrated core (MIC) architectures *etc.*]; in our tests, using GPU computing increased the speed of these operations by at least a factor of six.

Table 2

Performance characteristics of the current distribution (Current) and the optimized (Optimized) and PCRCM versions of *RMCPProfile* obtained using a configuration of 233 280 atoms to fit $G(r)$ and the Bragg profile for the CaTiO₃ model with the tilting disorder.

The PCRCM version has been tested using different numbers of moves (from 1 to 16) run in parallel (N_m). N_{th} refers to the number of computer threads used. The number of accepted moves required for convergence to a reasonably accurate solution was $\geq 2 \times 10^8$.

N_m	Current	Optimized	PCRCM						
	1	1	1	2	4	6	8	12	16
Time/move (ms)	30	1.25	1.34	0.865	0.67	0.506	0.436	0.387	0.426
N_{th}	1	8	8	8	8	12	16	24	32
Speedup factor	1	24	22.38	34.68	44.77	59.288	68.81	77.52	70.42
Time to 2×10^8 moves (h)	1667	69	74	48	37	28	24	22	24

All the computations referenced in Tables 1 and 2 have been performed using a cluster built with Intel Xeon Processors E5-2687w v4, each having a base frequency of 3.0 GHz, 12 cores, 24 threads and 30 MB L3 cache.

2.2. Parallel-chains RMC algorithm

While the above improvements of *RMCPProfile* appear to be transformative, even these speeds can be insufficient for fitting three-dimensional single-crystal diffuse scattering near Bragg peaks or for probing nanostructures. Opportunities for further improvements of computing efficiency using the existing serial (single Markov chain) RMC algorithm are limited because increasing the number of threads in parallel computations beyond a certain threshold yields no gain. One way to address this challenge is by parallelizing the RMC algorithm itself as described below.

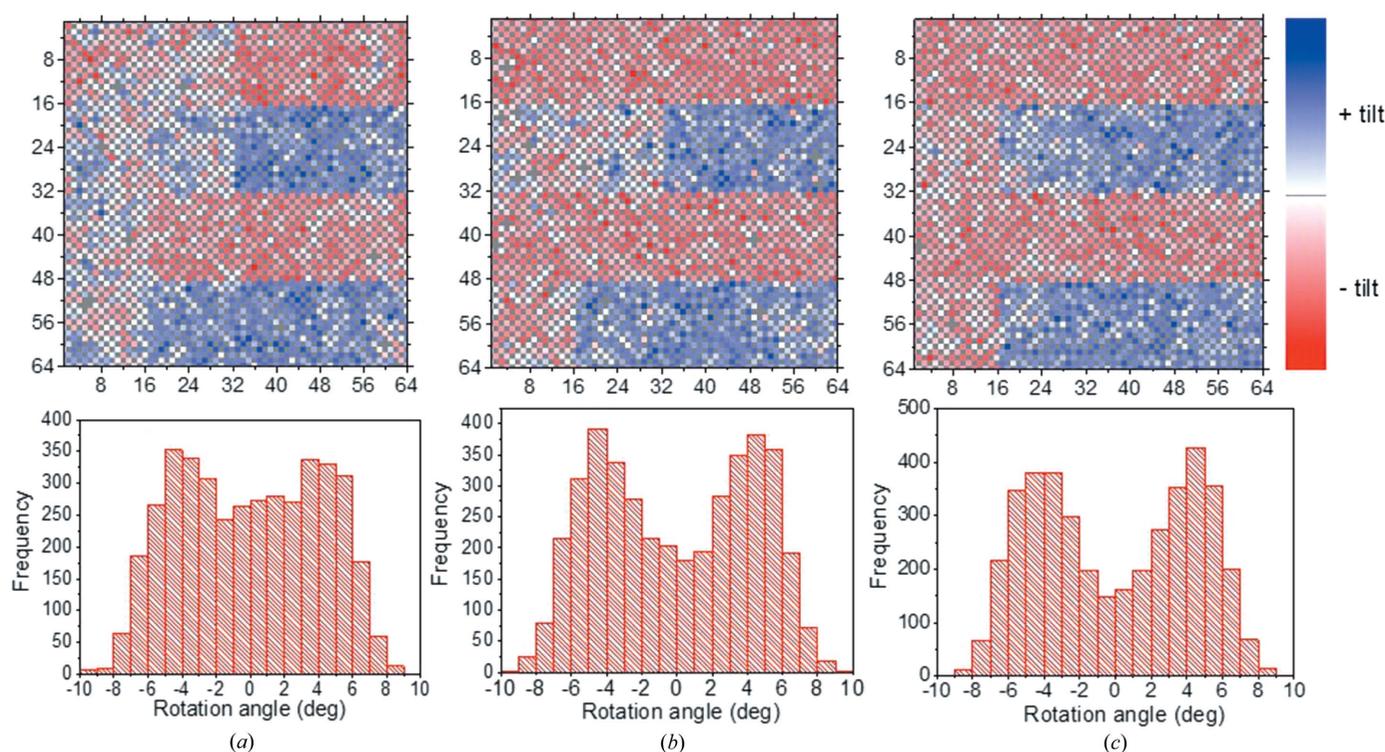


Figure 3

A schematic rendering of the same structure as shown in Fig. 1, refined using the current distribution of the 32-bit version of *RMCPProfile* after (a) nine days, (b) 37 days and (c) 132 days of fitting time. Over 30 days were required to achieve a model that has a degree of rotational ordering resembling that in the target structure (*cf.* Figs. 1b and 1c), if using typical contemporary hardware. The magnitudes and signs of octahedral rotations are coded according to the same color scheme as used in Fig. 1. The white color represents rotation angles that are close to zero.

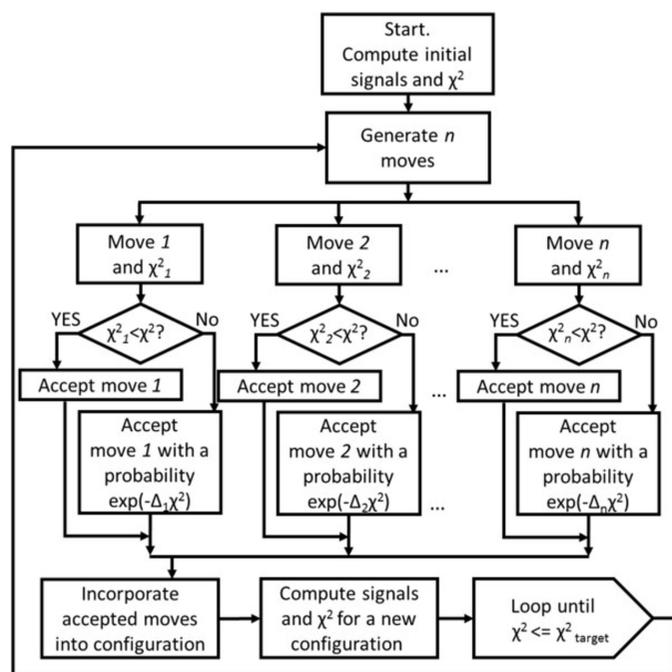


Figure 4
A flow-chart diagram describing the PCRMC algorithm.

Unlike the current *RMCPProfile*, which considers a single atomic move at a time, the new algorithm processes multiple moves in parallel, as in Markov chain Monte Carlo with inter-chain adaptation (Solonen *et al.*, 2012). We will refer to this algorithm as parallel-chains RMC (PCRMC). Several atoms, separated by a distance larger than a user-defined exclusion radius,⁴ are selected at random and moved in random directions similarly to a standard RMC procedure (Fig. 4). For each of these moves, signals of interest are calculated and compared with experimental data, and the total residual χ^2 is used to determine the move's acceptance. The relevant signals are then recalculated for an updated configuration, which incorporates all the accepted moves from the considered bunch, and again compared with the experimental data to obtain a new total residual. This total residual is used as a reference for evaluating the changes produced by the next bunch of moves. The process continues until convergence.

As each move in a generated bunch is evaluated separately for acceptance, if several of these moves are accepted, the interatomic distance histogram [and the signals calculated from it, such as $G(r)$ and $S(Q)$] for the resulting configuration must be corrected to account for distances that separate the moved atoms. This correction adds computing time, partially reducing the gain from considering several moves in parallel. Nevertheless, the PCRMC algorithm yields significant speedup factors, as demonstrated using the same CaTiO_3 example with a tilting disorder.

⁴ In the current implementation, the exclusion radius must be set to a value greater than or equal to one-half of the interatomic distance. The upper limit for this radius depends on the size of an atomic configuration and the number of parallel moves used. Here, we used a value of 3 Å.

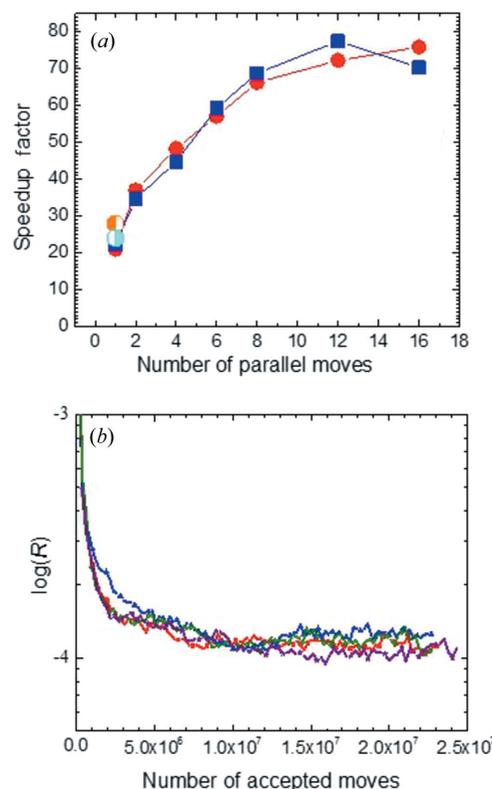


Figure 5
(a) The dependence of the speedup factor for PCRMC (relative to the standard non-optimized version) on the number of moves in a bunch for the test configurations with 20 480 (circles, red) and 233 280 (squares, blue) atoms. The speedup factor for the optimized serial RMC is also shown for comparison (left half-filled circle, orange – 20 480 atoms; right half-filled circle, cyan – 233 280 atoms). (b) The dependence of the residual for $G(r)$ on the number of accepted moves for the PCRMC fits. Different curves correspond to different numbers of atomic moves (from 1 to 16) run in parallel. The curve for the optimized serial version of *RMCPProfile* is included for comparison. All the curves are similar, making their color coding redundant. These two plots (a), (b) demonstrate that the PCRMC algorithm drastically reduces the time per move, while requiring the same total number of accepted moves for convergence as the serial RMC.

The optimal number of moves in a bunch and the achievable gains in speed vary with the size of the atomic configuration and the types of data included in the fit (Fig. 5). For example, if fitting the $G(r)$ and Bragg profiles for the 233 280-atom configuration, a speedup factor of ~ 45 (Table 2) is obtained with only four parallel moves, whereas if $S(Q)$ is included, at least 12 moves are required to achieve the same level of performance. Currently, without $S(Q)$, the PCRMC algorithm can yield speedup factors of ~ 80 . In principle, including $S(Q)$ in the fit should be less important if the PDF is fitted to longer distances. However, we expect that after the $S(Q)$ calculations are implemented on a GPU speedup factors of the order of ~ 80 will be achievable even while fitting $S(Q)$.

The actual maximum number of moves that could be enabled in PCRMC refinements depends on the size of the atomic configuration, as the moved atoms must be separated by distances greater than a certain exclusion radius, and on the number of threads available in the hardware, logical or physical cores. The required number of cores, N_C , is calculated

as the product $N_C = N_m N_{\text{th/exp}}$, where N_m is the number of moves generated in parallel and $N_{\text{th/exp}}$ is the number of threads used in the calculations of each dataset. Thus, a fit with $N_m = 4$ and $N_{\text{th/exp}} = 4$ will require 16 cores. An option for computing several datasets in parallel has also been implemented, which is projected to yield additional gains in speed; however, using this option would require a proportionally larger number of cores according to $N_C = N_m N_{\text{th/exp}} N_{\text{exp}}$, where N_{exp} is the number of experimental datasets calculated in parallel.

3. Corrections for instrument resolution

$G(r)$ is obtained as the sine Fourier transform of the normalized X-ray/neutron scattering function $S(Q)$. Finite reciprocal-space resolution of a diffractometer causes progressive damping of $G(r)$ peaks with increasing r (Egami & Billinge, 2012). In the case of a Gaussian instrumental profile with a Q -independent peak width, w , the intensities of the PDF peaks are modified by a factor $\exp(-w^2 r^2/2)$. However, for most measurement setups, the resolution varies with Q , which leads to additional broadening of the PDF peaks. If the instrumental Gaussian broadening of diffraction peaks increases linearly with Q (i.e. $w = \alpha Q$), the $G(r)$ peak widths, σ_{ij}^{tot} , become $\sigma_{ij}^{\text{tot}} = \sigma_{ij}^2 + \alpha^2 r_{ij}^2$, where σ_{ij} corresponds to the spread of interatomic distances r_{ij} in the material (Egami & Billinge, 2012). The r -space effects of such a Gaussian instrumental resolution function, which can be modeled using parameters w and α , are incorporated into the *PDFGUI* software (Farrow *et al.*, 2007), which is designed for Rietveld-style structural refinements using a PDF. Most neutron time-of-flight (TOF) instruments (e.g. Polaris and GEM at ISIS, and NOMAD at SNS) currently available for PDF measurements exhibit medium resolution, the effects of which on a PDF become significant above 2–3 nm. The resolution of X-ray PDF setups with the currently available area detectors is worse, impacting PDFs at even shorter distances.

Typically, in RMC refinements, one would fit both $G(r)$ and $S(Q)$, which emphasize the local and longer-range structures differently (Tucker *et al.*, 2007). As indicated above, the calculated $S(Q)$ is obtained as the FT of $G(r)$, determined directly from atomic coordinates; that is, this $S(Q)$ represents a scattering function that has been effectively convoluted with a configuration box. For the box sizes currently used (<6 nm), the Bragg peaks in such a calculated $S(Q)$ are considerably broader than those encountered in a typical experimental diffraction pattern; therefore, in *RMCPProfile*, the experimental $S(Q)$ is also convoluted with a box function prior to comparison with the calculated data. The instrumental broadening of diffraction peaks is significantly smaller than that resulting from a box size and can be largely ignored. However, as configuration sizes become larger, instrumental and box-related broadenings become comparable, and therefore the resolution effects must be accounted for.

In practical cases, the instrumental profile shape and its dependence on Q are more complex than those described by the above Gaussian model, and so are the r -space effects of

instrument resolution. TOF neutron data frequently combine contributions from several detector banks, each having its own resolution. The instrumental broadening in $S(Q)$ thus depends on the Q ranges of individual banks included in the blending.

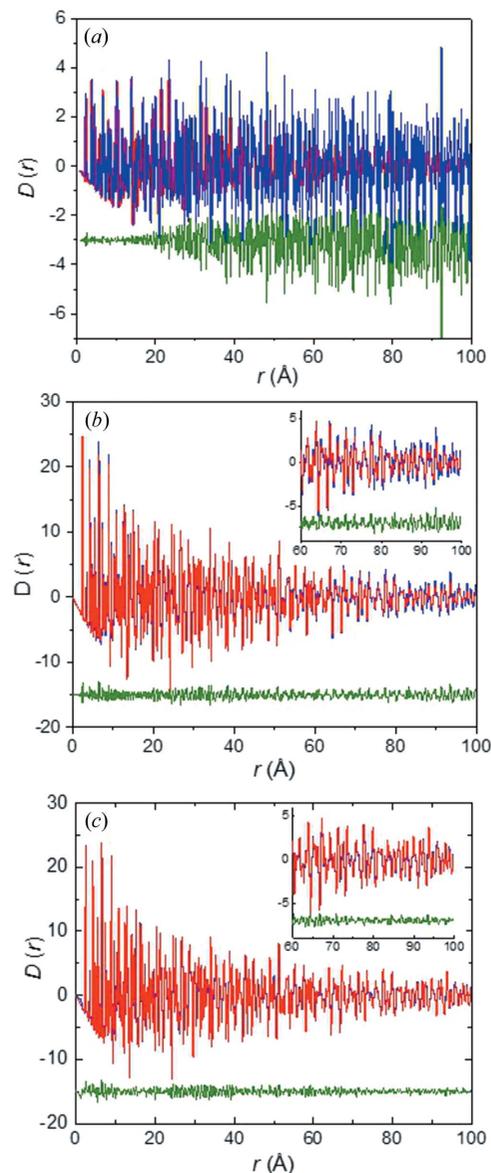


Figure 6

The experimental (red) and calculated (blue) PDFs of nickel with the difference curve (green) shown below. The data are presented in the form of $D(r)$, which is used by *PDFGUI* and is related to $G(r)$ as $D(r) = G(r)4\pi\rho_0 r/(b)^2$, where ρ_0 is the number density of the material and b is the neutron scattering length. $D(r)$ emphasizes the high- r range stronger than $G(r)$. (a) The calculated curve was obtained by first fitting the experimental $D(r)$ in *PDFGUI* to $r_{\text{max}} = 40 \text{ \AA}$ without the instrument resolution correction and then recalculating it over the entire distance range with the structural parameters fixed at their refined values. (b) The calculated curve was obtained by fitting the experimental data in *PDFGUI* over the entire distance range with both structural and resolution parameters (Q_{damp} and Q_{broad}) varied. (c) The calculated curve corresponds to a fit in *RMCPProfile*, presented in Fig. 7(b), which has been converted into the $D(r)$ form to facilitate comparison with the output of *PDFGUI*. Clearly, using the exact resolution function, as now implemented in *RMCPProfile*, improves the fit considerably, especially at large r .

To illustrate the effects of Q -space resolution on RMC refinements, we performed test fits using the total-scattering data collected for silicon (NIST SRM 640c) and nickel powder samples on Polaris at ISIS. The experimental data were reduced in the *GUDRUN* software to obtain $S(Q)$ and $G(r)$. Figs. 6(a) and 6(b) display results of fitting the $G(r)$ for silicon in *PDFGUI* up to $r_{\max} = 10$ nm, both without accounting for the resolution of the instrument and using a Gaussian model of the resolution function, the parameters of which, Q_{damp} and Q_{broad} , were adjusted during the fit. Clearly, this model fails to describe the experimental resolution effects over the entire distance range and the fit overestimates the atomic displacement parameter. Therefore, in *RMCPProfile* we implemented a more elaborate approach that uses parameters of the actual resolution function to correct the calculated $G_{\text{calc}}(r)$ prior to its comparison with experimental data. $G(r)$ is related to $S(Q)$

as $G(r_i) = \mathbf{F}S(Q_i)$, where \mathbf{F} is the matrix of the FT coefficients calculated on a discrete $m \times n$ grid, with m and n representing the number of Q and r points, respectively. Since both functions are defined on discrete grids, $S_{\text{calc}}(Q)$ can be obtained from $G_{\text{calc}}(r)$ according to $S_{\text{calc}}(Q_i) = \mathbf{F}^{-1}G_{\text{calc}}(r_i)$, as it is implemented in *RMCPProfile*.

The convolution of $S_{\text{calc}}(Q)$ on a discrete grid with an instrument resolution function is equivalent to multiplying it by another matrix \mathbf{B} , which can be generated from the profile coefficients obtained by fitting the data for a standard reference sample in a Rietveld software, like *GSAS* (Larson & Von Dreele, 1994):

$$S'_{\text{calc}}(Q_i) = \mathbf{B}\mathbf{F}^{-1}G_{\text{calc}}(r_i), \quad (1)$$

where $S(Q_i)$ and $G(r_i)$ are the components of the vectors \mathbf{S} and \mathbf{G} .

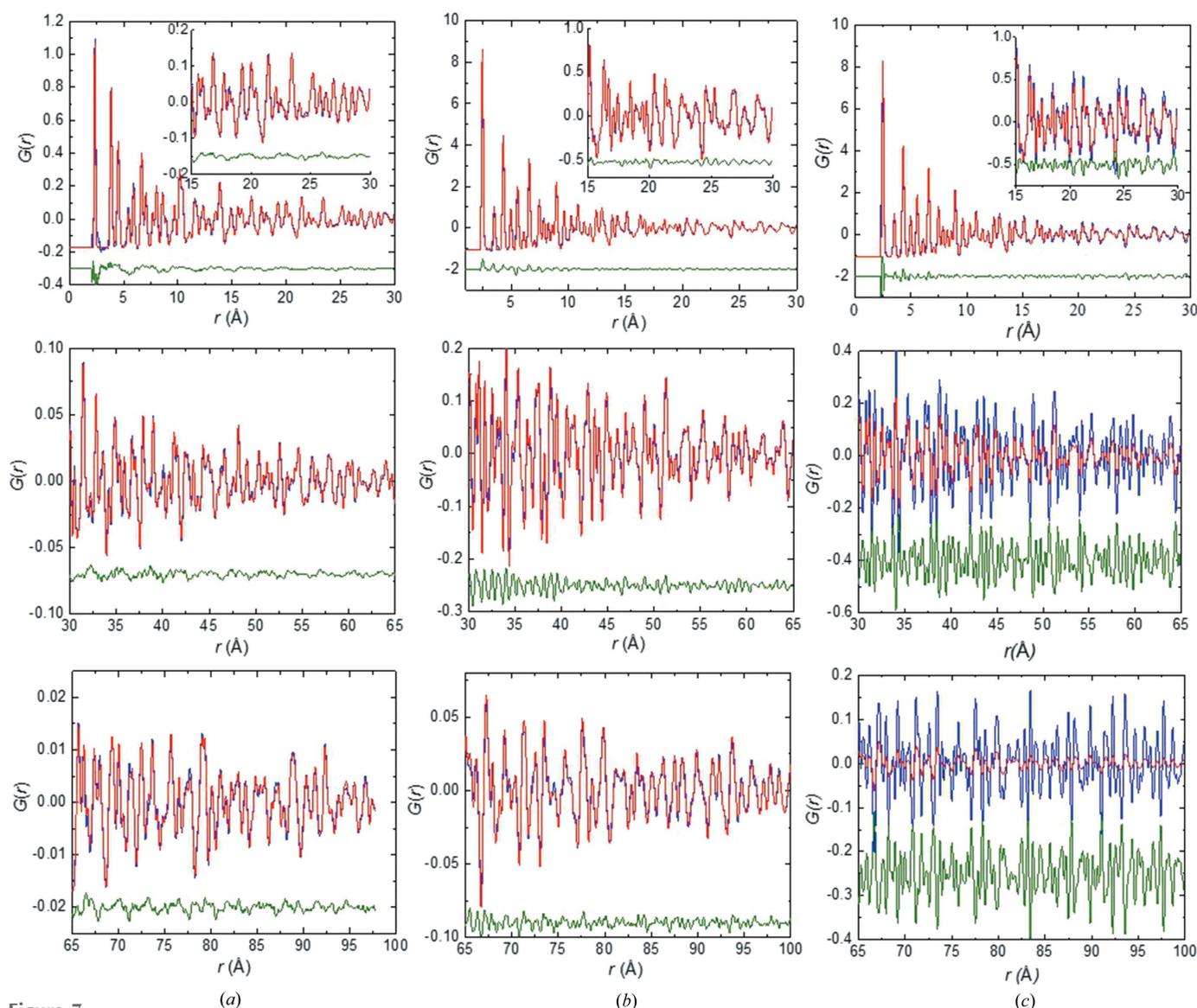


Figure 7 (a), (b) Portions of the neutron experimental (red) and calculated (blue) $G(r)$ for (a) Si and (b) Ni obtained by fitting the experimental data in *RMCPProfile* with the calculated signals corrected for the instrument resolution during these fits as described in the text. The difference is shown in green. (c) A fit for Ni without the resolution correction. In this case, forcing the configuration to improve the fit at larger r results in grossly incorrect values of the atomic displacement parameter.

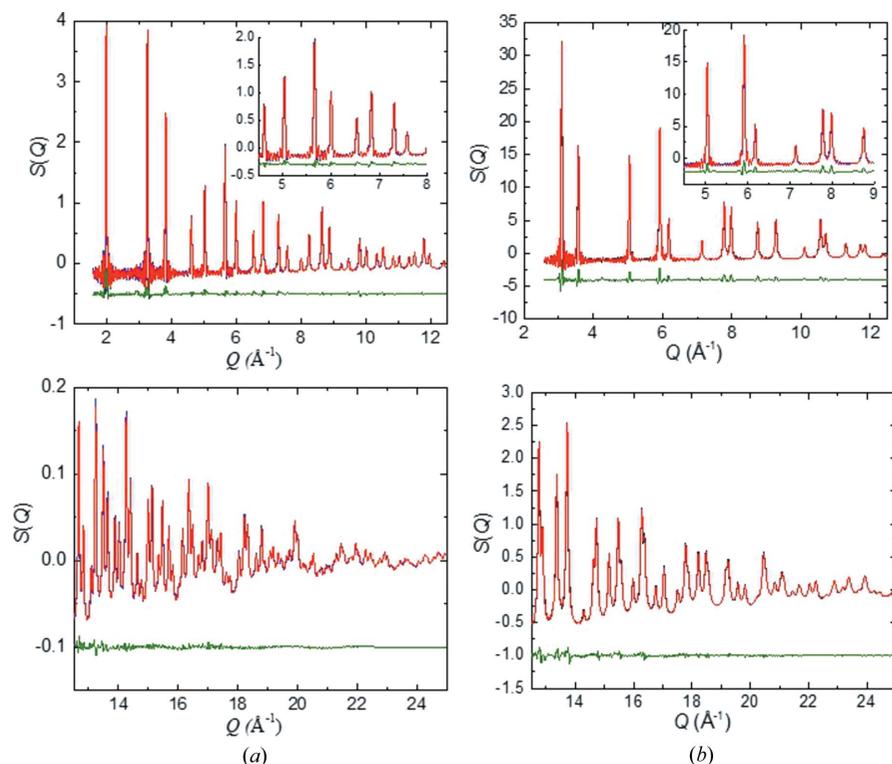


Figure 8
Portions of the neutron $S(Q)$ for (a) Si and (b) Ni fitted in *RMCProfile* simultaneously with the $G(r)$ shown in Fig. 7(a) and 7(b). The calculated $S(Q)$ have been corrected for resolution during these fits.

The resulting resolution-corrected $S_{\text{calc}}(Q)$ can be transformed back into r space by multiplying it by \mathbf{F} as

$$G'_{\text{calc}}(r_i) = \mathbf{F}\mathbf{B}\mathbf{F}^{-1}G_{\text{calc}}(r_i), \quad (2)$$

where $G'_{\text{calc}}(r)$ is the PDF that has been corrected for the instrument resolution. This operation is equivalent to multiplying $G_{\text{calc}}(r)$ directly by the product $\mathbf{R} = \mathbf{F}\mathbf{B}\mathbf{F}^{-1}$, which can be determined prior to a fit. The matrix \mathbf{R} is such that it can be approximated by a band matrix, which enables significant savings in computing time. If $S(Q)$ is included in a fit, the resolution-corrected $S'_{\text{calc}}(Q)$ must be matched to the experimental data. Since the $\mathbf{B}\mathbf{F}^{-1}$ matrix has the same dimensions as \mathbf{F}^{-1} , accounting for the resolution effects in Q space adds no extra computing time.⁵

In our procedure, an instrument resolution function is first determined for a reference sample using a full profile fitting (e.g. a Le Bail fit) in a standard Rietveld software. For TOF neutron data, the fit must be performed for all the relevant detector banks. The resulting profile coefficients are then fed into a utility program, which generates the resolution matrices $\mathbf{B}\mathbf{F}^{-1}$ and \mathbf{R} to be used by *RMCProfile*. This program accounts for the specific Q ranges of individual banks that are used to generate a blended $S(Q)$.

⁵ Our approach differs from that previously adopted in the *MCGRtof* program (Tucker *et al.*, 2001), which recovers the resolution-corrected $G(r)$ by fitting its Fourier transform, convoluted with the instrument resolution function, to the experimental $S(Q)$. In the present method, both the resolution correction and the fit of $G(r)$ are performed directly in r space.

Figs. 7(a) and 8(a) illustrate the results of a fit of $G(r)$ and $S(Q)$ for silicon using the data collected on Polaris, with the resolution function determined and accounted for in *RMCProfile* as described above. Satisfactory agreement could be achieved up to distances of 10 nm, returning a Gaussian-shaped isotropic atomic displacement parameter of 0.006 Å, which is consistent with the values reported in the literature. The tests demonstrated that using the exact resolution function is critical for reproducing the correct shape of an atomic probability density distribution (also known as an atomic displacement parameter) if performing fits over $r > 4$ nm, where the resolution effects become particularly significant. Naturally, the importance of accurately reproducing the resolution function grows for structures with narrow diffraction peaks.

Figs. 7(b) and 8(b) display a fit of $G(r)$, $S(Q)$ and the Bragg profile for Ni using a resolution function determined for Si. [For both Si and Ni, the same Q ranges were included for the individual banks while obtaining $S(Q)$.] Satisfactory agreement between the experimental and calculated

signals over the entire r and Q ranges confirms that a resolution function determined using a reference sample can be applied to a sample of another material measured at a different time; note that changing the bank ranges used to generate the merged $S(Q)$ for Ni would require regenerating the resolution function from the Si data for these new bank ranges. Fig. 6(c) displays the same result as shown in Fig. 7(a) but in the $D(r)$ form, which better emphasizes the high- r range. The fits for Si and Ni have been performed using atomic configurations of 373 248 and 740 772 atoms, respectively, and required ~ 1 week of computing time to generate $\sim 10^8$ moves, which for these mono-elemental materials were deemed sufficient for convergence.

4. Summary

The *RMCProfile* software, which has become a *de facto* standard for structure refinements using atomistic models and an RMC algorithm, has been optimized and further developed to drastically increase its computational speed. In addition to major optimization of the existing code *via* extensive parallelization of bottlenecks and redevelopment of the Bragg-profile computations, we outfitted *RMCProfile* with the new parallel-chains RMC algorithm, which altogether can speed the calculations up by nearly two orders of magnitude. With these improvements, structural refinements using atomic configurations that contain of the order of 500 000 atoms and

sample distances up to 10 nm become feasible. We have developed and implemented efficient procedures to account for the effects of the finite instrument resolution on both the reciprocal- and real-space forms of the total-scattering data, so that a PDF could be fitted to distances much longer than previously possible. For example, satisfactory fits up to 10 nm have been demonstrated for the PDFs of the reference samples. The ‘fast’ *RMCPProfile* opens new opportunities for the RMC structure-refinement method by extending its applications to nanostructures, including materials with pseudosymmetry, structures with large unit cells *etc.* We expect the new version to be available to the users of *RMCPProfile* via the *RMCPProfile* project page (<http://www.rmcpfile.org>) soon.

Acknowledgements

The authors are grateful to Helen Playford (ISIS, UK) for providing the neutron total scattering data for silicon and nickel and to Matt Tucker (ORNL) for his help with data processing.

References

- Aksel, E., Forrester, J. S., Nino, J. C., Page, K., Shoemaker, D. P. & Jones, J. L. (2013). *Phys. Rev. B*, **87**, 104113.
- Bosak, A., Chernyshov, D., Vakhrushev, S. & Krisch, M. (2012). *Acta Cryst. A* **68**, 117–123.
- Burns, G. & Scott, B. A. (1973). *Solid State Commun.* **13**, 423–426.
- Comes, R., Lambert, M. & Guinier, A. (1968). *Solid State Commun.* **6**, 715–719.
- Duncan, H. D., Dove, M. T., Keen, D. A. & Phillips, A. E. (2016). *Dalton Trans.* **45**, 4380–4391.
- Egami, T. & Billinge, S. J. L. (2012). *Underneath the Bragg Peaks, Structural Analysis of Complex Materials*. New York: Pergamon.
- Farrow, C. L., Juhas, P., Liu, J. W., Bryndin, D., Božin, E. S., Bloch, J., Proffen, Th. & Billinge, S. J. L. (2007). *J. Phys. Condens. Matter*, **19**, 335219.
- Jones, G. O. & Thomas, P. A. (2002). *Acta Cryst.* **B58**, 168–178.
- Keeble, D. S., Barney, E. R., Keen, D. A., Tucker, M. G., Kreisel, J. & Thomas, P. A. (2012). *Adv. Funct. Mater.* **23**, 185–190.
- Keen, D. A. & Goodwin, A. L. (2015). *Nature*, **521**, 303–309.
- Krayzman, V., Levin, I., Woicik, J. C., Proffen, Th., Vanderah, T. A. & Tucker, M. G. (2009). *J. Appl. Cryst.* **42**, 867–877.
- Larson, A. C. & Von Dreele, R. B. (1994). *General Structure Analysis System (GSAS)*. Report LAUR 86–748. Los Alamos National Laboratory, New Mexico, USA.
- Levin, I., Krayzman, V. & Woicik, J. C. (2014). *Phys. Rev. B*, **89**, 024106.
- Levin, I., Krayzman, V., Woicik, J. C., Bridges, F., Sterbinsky, G. E., Usher, T.-M., Jones, J. L. & Torrejon, D. (2016). *Phys. Rev. B*, **93**, 103106.
- Levin, I. & Reaney, I. M. (2012). *Adv. Funct. Mater.* **22**, 3445–3452.
- McGreevy, R. L. (2001). *J. Phys. Condens. Matter*, **13**, R877–R913.
- Playford, H. Y., Owen, L. R., Levin, I. & Tucker, M. G. (2014). *Annu. Rev. Mater. Res.* **44**, 429–449.
- Rödel, J., Jo, W., Seifert, K. T. P., Anton, E. M., Granzow, T. & Damjanovic, D. (2009). *J. Am. Ceram. Soc.* **92**, 1153–1177.
- Senn, M. S., Keen, D. A., Lucas, T. C. A., Hriljac, J. A. & Goodwin, A. L. (2016). *Phys. Rev. Lett.* **116**, 207602.
- Solonen, A., Ollinaho, P., Laine, M., Haario, H., Tamminen, J. & Jarvinen, H. (2012). *Bayesian Anal.* **7**, 715–736.
- Tucker, M. G., Dove, M. T. & Keen, D. A. (2001). *J. Appl. Cryst.* **34**, 780–782.
- Tucker, M. G., Keen, D. A., Dove, M. T., Goodwin, A. L. & Hui, Q. (2007). *J. Phys. Condens. Matter*, **19**, 335218.
- Wells, S. & Sartbaeva, A. (2015). *Mol. Simul.* **41**, 1409–1421.
- Xu, G., Shirane, G., Copley, J. R. D. & Gehring, P. M. (2004). *Phys. Rev. B*, **69**, 064112.
- Zhang, N., Yokota, H., Glazer, A. M., Ren, Z., Keen, D. A., Keeble, D. S., Thomas, P. A. & Ye, Z.-G. (2014). *Nat. Commun.* **5**, 5231.