

RMCA

Version 3

A General Purpose Reverse Monte Carlo Code

R L McGreevy, M A Howe and J D Wicks

February 1993

1. Introduction

One of the major problems in the analysis of diffraction data on disordered systems has been the lack of any *general* method for producing structural models that agree *quantitatively* with the data. Most analysis is extremely qualitative and based on a few features of the data, for example peak positions and coordination numbers derived from radial distribution functions. Monte Carlo and Molecular Dynamics simulations based on an interatomic potential sometimes agree well with experiment (though comparison is normally made with radial distribution functions rather than structure factors), but usually the agreement is only qualitative and occasionally there are major differences. It is not obvious in most cases how the potential should be altered to improve the level of agreement; an iterative procedure is computationally extremely expensive and has only been applied in one or two instances.

The reverse Monte Carlo (RMC) method [4] overcomes these problems. It is a method for producing three dimensional models of the structure of disordered materials that agree quantitatively with the available diffraction data. No interatomic potential is required and data from different sources (neutrons, X-rays, EXAFS) may be combined. The structural model is actually fitted to the data and so there must be good agreement (given that the data do not contain significant systematic errors).

RMCA is a general purpose Fortran code for reverse Monte Carlo (RMC) modelling. This manual first describes the RMC method and then goes on to describe the RMCA programme.

2. Metropolis Monte Carlo (MMC)

RMC is a variation of the standard MMC procedure [1]. For those unfamiliar with such Monte Carlo methods it is useful to first introduce MMC. The principle is that we wish to produce a statistical ensemble of atoms (configuration) with a Boltzmann distribution of energies. Rather than simply generating and sampling configurations completely at random, which would be a very inefficient procedure, we make use of a weighted sampling procedure (Markov chain) that satisfies certain requirements.

- (a) variables $x^{(n)}$ are generated following a rule that requires the $(n+1)^{\text{th}}$ element to have a probability distribution $x^{(n+1)}$ that is only dependent on the distribution $x^{(n)}$ of the n^{th} element.
- (b) If $P(x \Rightarrow y)$ is the probability of reaching state y from state x , then $P(x \Rightarrow y)$ must permit movement to every state in the ensemble.
- (c) Microreversibility must be satisfied, that is $xP(x \Rightarrow y) = yP(y \Rightarrow x)$ when the system is in equilibrium

For an ensemble in which the number of particles, volume and temperature are fixed (NVT) this may be achieved by the following algorithm.

1. N atoms are placed in a cell with periodic boundary conditions, by which we mean that the cell is surrounded by images of itself. Normally a cubic cell is used,

although other geometries may be chosen. For a cube of side L the atom number density $\rho = N/L^3$ must equal the required density of the system. The probability of this particular configuration (old = o) is given by

$$P_o \propto \exp(-U_o/kT) \quad (1)$$

where U_o is the total potential energy, which may be calculated on the basis of a specified form of the interatomic potential, and T is the specified temperature.

2. One atom is moved at random. The probability of the new (n) configuration is

$$P_n \propto \exp(-U_n/kT) \quad (2)$$

and hence

$$P_n/P_o = \exp(-(U_n - U_o)/kT) = \exp(-\Delta U/kT) \quad (3)$$

3. If $\Delta U < 0$ the new configuration is accepted and becomes the next starting point. If $\Delta U > 0$ it is accepted with probability P_n/P_o , otherwise it is rejected and we return to the previous configuration.

4. The procedure is repeated from step 2.

As atoms are moved U will decrease until it reaches an equilibrium value about which it will then oscillate. The maximum size of the random move is normally adjusted so that the ratio of accepted to rejected moves in equilibrium is approximately unity. Configurations are considered to be statistically independent when separated by at least N accepted moves and are then saved. In this way an appropriate ensemble is generated.

3. RMC - the basic method

In RMC we assume that an experimentally measured structure factor $A^E(Q_i)$ contains only statistical errors that have a normal distribution (this will be discussed in more detail later). The difference between the real structure factor, $A^C(Q_i)$, which can be calculated from a model of the real structure, and that measured experimentally is then

$$e_i = A^C(Q_i) - A^E(Q_i) \quad (4)$$

and has probability

$$p(e_i) = \frac{1}{\sqrt{2\pi}\sigma(Q_i)} \exp\left(-\frac{e_i^2}{2\sigma(Q_i)^2}\right) \quad (5)$$

where $\sigma(Q_i)$ is the standard deviation of the normal distribution. The total probability of A^C is

$$\begin{aligned} P &= \prod_{i=1}^m p(e_i) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^m \exp\left(-\sum_{i=1}^m \frac{e_i^2}{2\sigma(Q_i)^2}\right) \end{aligned} \quad (6)$$

where m is the number of Q_i points in A^E and

$$\bar{\sigma} = \left(\prod_{i=1}^m \sigma(Q_i) \right)^{1/m} \quad (7)$$

In order to model the structure of a system using A^E we therefore wish to create a statistical ensemble of atoms whose structure factor satisfies the above probability distribution. Writing the exponent as

$$\chi^2 = \sum_{i=1}^m (A^C(Q_i) - A^E(Q_i))^2 / \sigma(Q_i)^2 \quad (8)$$

then $P \propto \exp(-\chi^2/2)$ and it can immediately be seen that $\chi^2/2$ in RMC is equivalent to U/kT in MMC. The algorithm for RMC is therefore as follows.

1. Start with an initial configuration with periodic boundary conditions. The positions of the N atoms may be chosen randomly, they may have a known crystal structure, or they may be a configuration from a different simulation or model.
2. Calculate the radial distribution function for this old configuration

$$g_o^C(r) = \frac{n_o^C(r)}{4\pi r^2 \Delta r \rho} \quad (9)$$

where $n_o^C(r)$ is the number of atoms at a distance between r and $r + \Delta r$ from a central atom, averaged over all atoms as centres. The configuration size L should in principle be sufficiently large that there are no correlations across the cell, so that $g(r > L/2) = 1$. The radial distribution function $g(r)$ is only calculated for $r < L/2$ and the nearest image convention is used to determine the atomic separations.

3. Transform to the total structure factor

$$A_o^C(Q) - 1 = \frac{4\pi\rho}{Q} \int_0^\infty r(g_o^C(r) - 1) \sin Qr dr \quad (10)$$

where Q is the momentum transfer.

4. Calculate the difference between the measured total structure factor $A^E(Q)$ and that determined from the configuration $A^C(Q)$

$$\chi_o^2 = \sum_i (A_o^C(Q_i) - A^E(Q_i))^2 / \sigma(Q_i)^2 \quad (11)$$

where the sum is over the m experimental points and $\sigma(Q_i)$ is the experimental error. In practice a uniform σ is normally used, since the distribution of systematic errors is normally unknown.

5. Move one atom at random. Calculate the new radial distribution function, $g_n^C(r)$ and total structure factor, $A_n^C(Q)$, and

$$\chi_n^2 = \sum_i (A_n^C(Q_i) - A^E(Q_i))^2 / \sigma(Q_i)^2 \quad (12)$$

6. If $\chi_n^2 < \chi_o^2$ the move is accepted and the new configuration becomes the old configuration. If $\chi_n^2 > \chi_o^2$ then the move is accepted with probability $\exp(-(\chi_n^2 - \chi_o^2)/2)$. Otherwise it is rejected.

7. Repeat from step 5.

As this process is iterated χ^2 will initially decrease until it reaches an equilibrium value about which it will fluctuate. The resulting configuration should be a three dimensional structure that is consistent with the experimental total structure factor within the experimental error. Statistically independent configurations may then be collected. In MMC configurations are normally assumed to be independent if separated by N accepted moves but in practice we usually use at least $5N$.

The algorithm used here is not strictly statistically correct, since we are actually sampling χ^2 (by varying A^C) and not the data (by multiple measurements of A^E). We should therefore use

$$\left(\frac{\chi_n^2}{\chi_o^2}\right)^{\frac{m}{2}-1} \exp(-(\chi_n^2 - \chi_o^2)/2) \quad (13)$$

in place of $\exp(-(\chi_n^2 - \chi_o^2)/2)$. However, given that most experimental data do not contain only statistical errors, and the loss of direct analogy with MMC, the former algorithm has been preferred.

The difference between RMC and MMC is simply that in RMC the difference between calculated and measured total structure factors (χ^2) is sampled, while in MMC the potential energy U is sampled. Otherwise the two algorithms are identical. It is particularly important that RMC uses a proper Markov chain, so that the final structure should be independent of the initial configuration. This makes the method an ab initio structural determination, rather than a refinement. However in some circumstances the method is deliberately used as a refinement. This involves only accepting moves that decrease χ^2 (i.e. $\sigma \rightarrow 0$) and corresponds to setting $T = 0$ in MMC.

Similar algorithms have been tried by a few others. The earliest examples are the work of Averbach and colleagues [2,3] who used models of only a few hundred atoms (all that could then be managed on the computers available), modelled $g(r)$, and used converging moves only (i.e. χ^2 must always decrease). This makes the results entirely dependent on the starting configuration and so the method did not find general applicability. In a later example Bertagnoli and colleagues [7] used an algorithm very similar to that for RMC, but only modelled the inter-molecular part of $g(r)$ for a series of molecular liquids.

4. Multiple data sets

The algorithm described in the previous section is specifically for modelling a single set of diffraction data which could be obtained using either X-rays, neutrons, or electrons. The fit may be to either the structure factor or to the radial distribution function, though the former is to be preferred because the distribution of errors in the latter may be highly non-uniform. In practice a fit may be made first to the radial distribution function, then to a subset of the total structure factor points, before being made finally to all the structure factor points. This considerably reduces the time required.

The RMC method is more general than this simple algorithm in that any set or sets of data which can be directly calculated from the structure can be modelled. It can be applied to isotopic substitution in neutron diffraction, or equivalently to anomalous scattering in X-ray diffraction, to EXAFS and possibly to NMR data. All data sets can be modelled simultaneously by adding the respective χ^2 values.

For a multicomponent system where the fit is to several different total structure factors (indicated by index k) we have

$$\chi^2 = \sum_k \chi_k^2 = \sum_k \sum_{i=1}^m (F_k^C(Q_i) - F_k^E(Q_i))^2 / \sigma_k(Q_i)^2 \quad (14)$$

For neutron diffraction

$$F_k^E(Q_i) = \sum_{\alpha} \sum_{\beta} c_{\alpha} c_{\beta} \bar{b}_{\alpha k} \bar{b}_{\beta k} (A_{\alpha\beta}(Q_i) - 1) \quad (15)$$

where c_{α} is the concentration and $\bar{b}_{\alpha k}$ the coherent scattering length for species α in sample k . $A_{\alpha\beta}(Q_i)$ are the partial structure factors. For X-ray diffraction

$$F_k^E(Q_i) = \sum_{\alpha\beta} c_{\alpha} c_{\beta} f_{\alpha k}^*(Q_i) f_{\beta k}^*(Q_i) (A_{\alpha\beta}(Q_i) - 1) \quad (16)$$

where

$$f_{\alpha k}^*(Q_i) = \frac{f_{\alpha k}(Q_i)}{\sum_{\alpha} c_{\alpha} f_{\alpha k}(Q_i)^2} \quad (17)$$

$f_{\alpha k}(Q_i)$ is the Q dependent form factor for X-rays of wavelength λ_k . The normalised value $f_{\alpha k}^*(Q_i)$ is used in the definition of $F_k^E(Q_i)$ for the usual case of the scattered intensity being measured with constant statistical error. For EXAFS data

$$F_{\alpha}^E(Q_i) = \sum_{\beta} 4\pi\rho \int r^2 (g_{\alpha\beta}(r) - 1) f_{\beta}(Q_i, r) dr \quad (18)$$

where we have replaced k with α in $F^E(Q_i)$ because the spectrum is measured at the absorption edge of species α . Here $f_{\beta}(Q_i, r)$ is the contribution to the EXAFS spectrum of a single atom of type β at a distance r which is calculated by one of

the standard EXAFS data analysis packages such as the SERC Daresbury programme EXCURV92. When fitting EXAFS data it is possible to Q -weight the spectra.

For simultaneous fitting of data sets obtained by different experimental techniques the separate χ^2 values are simply summed to give one value. The relative weighting of the different data sets is determined by the choice of the various σ values. Clearly the required computer time increases significantly if multiple data sets are fitted.

5. Constraints

Other information that cannot be used directly can be made use of in the form of constraints; this may include NMR, EPR, Raman scattering and chemical knowledge. The most commonly used constraint is on the closest distance of approach of two atoms. Because of systematic errors in the experimental data, and often because of the limited data range, the data would not forbid some atoms from coming very close together. However we know that this is physically unrealistic so an excluded volume is defined. Often realistic values for the closest approach distances can be determined from direct Fourier transformation of the measured total structure factors. It is usually obvious if an unsuitable choice has been made as spurious sharp spikes occur in $g(r)$ at low r .

While the closest distance of approach constraint is very simple, it is very powerful when used in conjunction with a fixed density. For many materials the dominant effect determining the structure is packing, and hence to include information on atomic sizes (these are minimum sizes rather than, for example, ionic radii) implicitly in the model severely limits the number of structures that are consistent with the data.

The second most commonly used constraint is on the coordination of the atoms. A coordination number $n_{\alpha\beta}$ is defined as being the number of atoms of type β between two fixed distances of one of type α . Normally the lower fixed distance is the closest distance of approach of the two types of atom (or equivalently zero). If we define the proportion of atoms of type α in the configuration with a particular coordination as f_{RMC} , and the desired proportion with such a coordination as f_{req} then we can add an additional term to χ^2 :

$$\chi^2 = \dots + (f_{\text{req}} - f_{\text{RMC}})^2 / \sigma_c^2 \quad (19)$$

Obviously multiple coordination constraints can be applied by adding additional terms. The parameter σ_c in this case simply acts as a weighting of the coordination constraint relative to the data. If $\sigma_c \approx 0$ it is effectively impossible for atoms with the constrained coordination to change it; this can be used to mimic the effect of covalent bonding. In many cases hard sphere Monte Carlo simulation with such coordination constraints, that is RMC with no data, can be used to produce structures with suitable topology prior to fitting the data.

This is a constraint on coordination numbers of individual (although unspecified) atoms. It is also possible to constrain average coordination numbers in the same

way. Average coordination numbers can be obtained from EXAFS data, and using the coordination number constraints rather than fitting the spectra is an alternative way of using such data.

6. RMC - why use it?

There are numerous methods of structural modelling, from the simplest ‘hand built’ or ‘hand drawn’ models to conventional MC or MD simulations. However RMC has several advantages.

1. RMC uses all the available structural data, not just particular features, in a quantitative rather than qualitative manner. Many models that use particular features, for example peak positions and coordination numbers from radial distribution functions, can be misleading.
2. RMC is potential independent. If a potential exists that, when used in an MC or MD simulation, also produces quantitative agreement with experimental structure factors, then this is obviously equally as good as RMC (and one would hope that the results were similar!). However few potentials provide such quantitative agreement, and in some cases it has not yet proved possible to produce potentials that provide qualitative agreement. Also most simulations are compared to experiment at $g(r)$ stage because the configurations are too small to allow transformation to $A(Q)$. For the reasons discussed above it is important for good structural modelling that comparison be made at the $A(Q)$ stage.
3. Because RMC models a three dimensional structure $g^C(r)$ and $A^C(Q)$ must correspond to a possible physical structure; the model is subject to the simple but powerful constraints of fixed density and excluded volume (minimum atomic sizes). However $g^E(r)$ derived by conventional methods may contain errors which mean that it *could not* correspond to a possible physical structure; that is, it is internally inconsistent. In the case of multicomponent systems there is no requirement that the partial radial distribution functions derived by conventional methods are consistent with one another, while in RMC they must be consistent and physically possible. This constraint improves the separation of partials in cases where the separation matrix is poorly conditioned. It also means that some information on partials can be obtained for underdetermined cases, where none can be obtained by conventional methods.
4. Different types of data, for example neutron and X-ray diffraction, can be modelled. The different data sets can have different Q ranges, spacings, resolutions etc. It is also easy to include additional constraints on the structure; these could be from other experimental information (e.g. NMR) or could be some other knowledge of, or assumptions about, the system (e.g. chemical bonding ideas).
5. RMC is easily adapted to different physical problems.

7. Uniqueness

The three dimensional structure produced by RMC is *not* unique, it is simply a model that is consistent with the data and any additional constraints. Other methods that produce structures which are equally consistent with the data are equally valid and there is no way of determining which is ‘correct’ *in the absence of any additional information*. One possible disadvantage of RMC is that it tends to produce the most disordered structure that is consistent with the data and constraints, that is the configurational entropy is maximised. However this is counteracted by the ability to include additional constraints, which means that additional ordering can be imposed and a range of consistent structures investigated; those that are found to be inconsistent can then be discounted.

In the special case of a system for which the interatomic potential is purely pairwise additive there is a theoretical justification for the determination of the three dimensional structure from a one dimensional $g(r)$ or $A(Q)$ [11]. Given that the potential uniquely determines the structure

$$\phi(r) \Rightarrow g^{(2)}(r_1, r_2), g^{(3)}(r_1, r_2, r_3), g^{(4)}(r_1, r_2, r_3, r_4) \dots \quad (20)$$

where $g^{(2)}(r_1, r_2) = g(r)$ and $g^{(n)}(\dots)$ are the n -body correlation functions, then for a pairwise additive potential there is a functional relationship between $\phi(r)$ and $g(r)$ such that

$$g^{(2)}(r_1, r_2) \Rightarrow \phi(r) \quad (21)$$

that is that $g(r)$ uniquely determines $\phi(r)$. (This is not to say that we can write down the relationship, but merely that one exists.) If $g(r)$ determines $\phi(r)$ and $\phi(r)$ determines the structure, then $g(r)$ determines the structure

$$g(r) \Rightarrow g^{(3)}(r_1, r_2, r_3), g^{(4)}(r_1, r_2, r_3, r_4) \dots \quad (22)$$

Theoretical tests have shown that, in such cases where the potential is purely pairwise additive, the RMC method works satisfactorily [21].

The potentials in real systems are never purely pairwise additive (though such potentials are used in the majority of MC and MD simulations). However the above result does indicate that a precisely measured $g(r)$ or $A(Q)$ does contain a great deal of information about the three dimensional structure. RMC is one possible way of attempting to extract this information. In cases where there are significant three-body terms in the potential then constraints may be used to take account of them. In the case of molecular liquids for example, molecules can be included explicitly in the model.

8. Applications of RMC

There are now numerous different applications of RMC modelling. Here they are summarised, with references, both by subject and technique.

1. RMC method [4,8,11,17,18,21,27,28,33-35,37,48].
2. Neutron diffraction [4-6,8-16, 18-20,22-29,31-34,36-38,40-44,47].
3. X-ray diffraction [8,9,14,24,28,33,39,45,46].
4. EXAFS [17].
5. Use of constraints [19,22,28,33,34].
6. Combination of experimental techniques [8,14,24,28,33,34,36].
7. Elemental liquids: condensed inert gases [4,23,32], liquid metals [20,32], liquid semiconductors [32], molecular liquids [9,32].
8. Binary liquids: liquid metal alloys [19,43,44], liquid semiconductors [6], molten salts [5,15].
9. Aqueous solutions [10].
10. Covalent glasses: silicates [14], Ag^+ fast ion conducting glasses [25,31,36], amorphous carbon [24].
11. Metallic glasses [38-42,45-47].
12. Magnetism in metallic glasses [22].
13. Structural disorder in crystals: Ag^+ fast ion conductors [12,13,26,29], C_{70} [30].

9. RMC - simulation details

9.1. Configuration size and shape

When starting from the initial configuration $g(r)$ must be calculated. This involves a summation of order N^2 . However for each particle move it is only necessary to calculate the change in $g(r)$ corresponding to the moved particle, which is a summation of order N . This is the same in MMC, but not in Molecular Dynamics (MD) where all moves are of order N^2 (unless the potential is truncated). For this reason MC simulations may involve much larger configurations than are used in MD. Generally we use $N > 1500$, and have used $N \approx 30000$. The size of simulation is important when modelling $A^E(Q)$ since $g^C(r)$ may only be calculated up to $r = L/2$. In order to be able to transform $g^C(r)$ directly to $A^C(Q)$ we require that $g^C(r > L/2) = 1$. Any significant deviations from this, either due to long range correlations or statistical fluctuations, will cause truncation ripples at low Q in $A^C(Q)$. Size is also relevant when modelling $g^E(r)$, because this determines the statistical fluctuations in $g^C(r)$ and hence the effective value of $\sigma(r)$.

Sometimes, if there are long range correlations, it is not possible to make the box large enough for $g^C(r > L/2) = 1$ to be a good approximation. However, there is a way around this problem. The effect of the finite box size on the calculated structure factor is that in the sine Fourier transform that calculates $F^C(Q)$ the argument $g(r) - 1$ is multiplied by a step function that is unity for $r < L/2$ and zero for $r > L/2$. Using the convolution theorem for Fourier transforms it is possible to show that this is equivalent to convolution of $F^C(Q)$ to give

$$\begin{aligned} & \frac{1}{\pi} \int_{-\infty}^{\infty} F^C(Q') \frac{\sin L/2(Q - Q')}{Q - Q'} dQ' \\ &= \frac{1}{\pi} \int_0^{\infty} F^C(Q') \left[\frac{\sin L/2(Q - Q')}{Q - Q'} - \frac{\sin L/2(Q + Q')}{Q + Q'} \right] dQ' \end{aligned} \quad (23)$$

In order to compare with the experimental data we should therefore perform the same convolution before using it as input to the RMC programme. A program CONVOL is available for this purpose. Alternatively one could model the radial distribution function using the program MCGR, and then fit this.

9.2. Closest approach distance of two atoms

For perfect data the distances of closest approach of pairs of atoms are determined by the low r cut-off in $g^E(r)$. However for imperfect data, particularly when $A^E(Q)$ is significantly truncated at the maximum Q value, the closest approach may not be well defined. For this reason it is usually sensible to specify allowed distances of closest approach, in other words to define an excluded volume. This also saves considerable time since moves which would result in atoms being too close together can be rejected before calculation of the change in $g^C(r)$. For good data the specified closest approaches may be somewhat lower than realistic values but for poor data they need to be more carefully chosen. If the values are too large then this is usually apparent because the resulting $g^C(r)$ has a sharp cut-off instead of decreasing more gradually to zero. If they are too low $g^C(r)$ may have a sharp spike in the low r region.

While this is a very simple constraint on the structure it is also very powerful, since the imposition of both an excluded volume and a fixed density restricts possible configurations. One could also view it as the imposition of a hard sphere repulsive potential. In the case of, for example, a two component system where the hard sphere radii are sufficiently different this constraint allows one to obtain some information on all three partial radial distribution functions or structure factors from one or two total structure factors, while three are required for a conventional solution. This is valuable in cases where suitable isotopes are not available for neutron diffraction. Since the resulting structure is then dependent on the choice of radii one should, if possible, make a choice based on other experimental information rather than treating them as free parameters.

9.3. Maximum size of random move

The maximum size, δ , of the random move determines the ratio of accepted to rejected moves, but also determines the amount that the structure may change with each move. If δ is too small then nearly all moves will be accepted but the structure will change little, while if it is too large then few moves will be accepted and the average structural change will also be small. If we attempt to choose δ such that the ratio of accepted to rejected moves is approximately one, as is often done in MMC, this usually leads to a value $\delta < 0.1\text{\AA}$. The average structural change per move is usually maximised for $0.1 < \delta < 0.3\text{\AA}$ with an acceptance/rejection ratio approximately 0.5, so this range is normally used.

When starting from a structure that is significantly different from the ‘real’ structure it is possible that certain atoms become ‘trapped’ in some local arrangement. This is a local minimum in the minimisation procedure, rather than the required global minimum. One way around this problem is to run the simulation for a while with a large value of δ , for instance up to $10\text{\AA}$. While hardly any moves will be accepted those that are may be sufficient to get the configuration out of the local minimum. (The terms ‘local’ and ‘global’ are used here in an unconventional manner; ‘global’ refers to any minimum that satisfies the required fitting criterion and ‘local’ to any minimum that does not. We specifically do not wish to attain the conventional global minimum, the single structure that is closest to the data, since the data contain errors.)

If the packing fraction (ratio of excluded volume to total volume) is high, as with many metallic glasses, then it is necessary to use small moves so that a sufficient number are accepted. The convergence of the RMC procedure may then be very slow. One way to ‘speed up’ the process is to artificially decrease the cut-offs for some time, i.e. to decrease the packing fraction, and then to increase them again later when a better fit to the data has been achieved.

See subsection 9.8 for a discussion of move sizes in relation to the use of coordination constraints.

9.4. r spacing and Q range

When modelling $A^E(Q)$ the minimum r spacing is determined by the maximum Q value, Q_m . The real space resolution is then $2\pi/Q_m$; one requires approximately five points over this range so an r spacing of $2\pi/(5Q_m)$ is appropriate. For simple liquids where structure in $A^E(Q)$ extends out to $Q \approx 10\text{\AA}^{-1}$ this makes a spacing of approximately $0.1\text{\AA}$ suitable, whereas for molecular liquids or glasses with structure out to $Q \approx 40\text{\AA}^{-1}$ a spacing of $0.025 - 0.05\text{\AA}$ is suitable. However it should be noted that decreasing the r spacing for a fixed number of particles increases the statistical error in $g^C(r_j)$, so it may be necessary to increase the model size. Having a large number of r points also increases the transform time if structure factors are being modelled, and increases the size of the coefficients file if EXAFS data are being modelled. Some compromise may be necessary if all of these factors are taken into account.

The minimum Q value that can be modelled is given by $Q_{min} = 2\pi/L$. If you try to fit to smaller Q values then the effects are unpredictable. For example, if a much smaller Q_{min} is used then this can lead to distinct density fluctuations, of period $2\pi/L$, in the configuration. Note that weak fluctuations of this period can be seen in many simulations, not just in RMC.

9.5. experimental error: σ

The RMC algorithm assumes that we have only statistical errors. In practice this is not true, but the whole procedure is not thereby invalidated. A three dimensional structure that is consistent with the experimental data within some measure of the error can still be produced, though this measure is now less well defined.

A real experimental structure factor $A^E(Q)$ will contain both statistical and systematic errors. While one might expect statistical errors to be small where $A^E(Q)$ is large, and vice-versa, in practice the requirement to perform container and background corrections in many experiments means that statistical errors are often quite uniformly distributed. In many X-ray experiments counting times are chosen to deliberately produce a uniform distribution. Since we often have no knowledge of the likely distribution of systematic errors it is usually simplest to assume a constant value of σ at all Q , though σ may differ between different data sets. However there have been cases in which large values of σ have been used in particular Q ranges where it is known that there were errors in the data. By setting $\sigma(Q)$ at an extremely large value these errors can effectively be ignored.

When $A^E(Q)$ is transformed to $g^E(r)$ the errors are redistributed. There are also statistical errors inherent in $g^C(r)$. Comparison at the $g(r)$ stage will therefore not produce precisely the same result as comparison at the $A(Q)$ stage. This is of course more exaggerated when additional errors are introduced in $g^E(r)$ by truncation of $A^E(Q)$. It is also worth noting that certain features of $A^E(Q)$, in particular ‘pre-peaks’ or ‘first sharp diffraction peaks’ in some glasses and liquids at $Q \approx 1\text{\AA}^{-1}$, correspond to small, long period modulations of $g^E(r)$. It is possible that such real space structure can be ‘ignored’ to a great extent if the modulation amplitude is comparable to the chosen value of $\sigma(r)$. For these reasons it is strongly recommended that RMC is used for modelling $A^E(Q)$ wherever possible. However it is also possible that in cases where there are well defined peaks in $g(r)$, for instance covalent bonding, and correspondingly oscillations in $A^E(Q)$ out to large Q , the high Q structure factor is not fitted as well as the low Q part. The peaks in $g^C(r)$ will then be lower and broader. To overcome this both $A^E(Q)$ and $g^E(r)$ (preferably obtained by an inverse method such as the MCGR program) can be fitted simultaneously.

From the above discussion it is clear that the precise value of σ is not known in any particular case; it may therefore be considered as a parameter of the simulation. If we make the analogy with MMC that $\chi^2 = U/kT$ then σ corresponds to kT . Under normal circumstances we would use a value of σ that was approximately 1% of the amplitude of the input data (a value typical of experimental error). However if it is believed that the simulation has run into a local minimum then, as is common

procedure in other simulations, we would increase the value of σ , (analogous to increasing the temperature). After running the simulation for a while σ would then be decreased to its original value. Alternatively if we deliberately wished to find a local minimum closest to a particular starting configuration then we can effectively set $\sigma = 0$ by only accepting moves which decrease χ^2 .

We have generally found for disordered structures that the global minimum in χ^2 is relatively broad and little manipulation of σ or d is required to reach it. However for more ordered structures, such as crystals, this would not be the case and global minimisation of χ^2 would only be achieved by simulated annealing with σ as the control parameter.

9.6. Renormalisation

Experimental data will normally contain small normalisation errors in the form of multiplicative and additive constants. It is possible to take account of such errors within the RMC algorithm. This can be particularly important when dealing with isotopic substitution neutron diffraction data when the relative normalisation of different structure factors must be correct. The required multiplicative factor which minimises χ^2 is

$$\frac{\sum_i A^E(Q_i)A^C(Q_i)}{\sum_i A^E(Q_i)^2} \quad (24)$$

It is recommended that such renormalisation only be performed as a refinement when a reasonable fit has already been achieved and the renormalisation factor is close to 1. If the required value differs significantly from 1 then the experimental data should obviously be checked.

The additive factor which minimises χ^2 is

$$\frac{1}{m} \sum_i (A^E(Q_i) - A^C(Q_i))^2 \quad (25)$$

It is generally safe to use this factor for all data though again if the value is very different from that expected the original data should be checked.

When both factors are used simultaneously the formulae are more complex [6].

9.7. Definitions of structure factors etc.

For the purposes of RMCA the input data, or structure factors, are defined in the following way. Data which are not defined in this way should be modified accordingly.

$$G(r) = \sum_{\alpha=1}^n \sum_{\beta=\alpha}^n \gamma_{\alpha\beta} (g_{\alpha\beta}(r) - 1) \quad (26)$$

where $\gamma_{\alpha\beta}$ are coefficients which are specified. If you wish to model a partial radial distribution function, e.g. $g_{12}(r)$, then all coefficients except γ_{12} must be set to zero

and 1 must be subtracted from the data. It is not possible to subtract an offset value from $G(r)$ within the program since it must necessarily tend to zero at large r .

$$S(Q) = \sum_{\alpha=1}^n \sum_{\beta=\alpha}^n \gamma_{\alpha\beta} (A_{\alpha\beta}(Q) - 1) \quad (27)$$

where $\gamma_{\alpha\beta}$ are coefficients which are specified. Partial structure factors can be modelled in the same way as partial radial distribution functions. Note that $G(r)$ is the direct transform of $S(Q)$.

$$F(Q) = \sum_{\alpha=1}^n \sum_{\beta=\alpha}^n \gamma_{\alpha\beta}(Q) (A_{\alpha\beta}(Q) - 1) \quad (28)$$

where $\gamma_{\alpha\beta}(Q)$ are Q dependent coefficients which are given in the same file as $F(Q)$. Note that the direct transform of $F(Q)$ is *not* $G(r)$, because of the Q dependence of the coefficients. Normally for X-ray data we define

$$\gamma_{\alpha\beta}(Q) = \frac{c_{\alpha} c_{\beta} f_{\alpha}(Q) f_{\beta}(Q)}{\sum_{\alpha} c_{\alpha} f_{\alpha}(Q)^2} \quad (29)$$

where $f_{\alpha}(Q)$ is the form factor for species α . $F(Q)$ then tends to a constant value at high Q and the offset and renormalisation options can be used.

For EXAFS data the coefficients are both Q and r dependent, and the structure factor is defined as

$$E_{\alpha}(Q) = \sum_{\beta} 4\pi\rho \int r^2 (g_{\alpha\beta}(r) - 1) f_{\beta}(Q, r) dr \quad (30)$$

The values of $f_{\beta}(Q, r)$ are given in a separate file from the data.

9.8. Use of coordination constraints

Coordination constraints are one of the most valuable and instructive ‘tools’ in the RMCA program. A description of how they may be used is best given by some examples.

9.8.1. Molecular liquids.

A model of water can be constructed as follows. The H-O-H molecule can be contained in a sphere of radius ?Å centred on the O atom. We start with a random arrangement of O atoms at an appropriate density and run a hard sphere simulation (RMC with no data) and with a constraint on the closest approach of ?Å, until no atoms are too close together. Two H atoms are then added to each O atom with relative positions (x,y) and (-x,y). The resulting configuration is now a set of aligned H-O-H molecules. The hard sphere simulation is run again, with a constraint that each

O is coordinated to two H between ? and ?Å, to destroy the molecular alignment. The constraint is then maintained and the data fitted. The molecules will remain ‘bonded’ but are flexible; the H-O-H bond angle should be determined by the data. If a bond angle constraint is wanted then this can be done by requiring each H to have one H neighbour (that in the molecule) within appropriate distances. Because the constraint on the O-H bond length is severe only small move sizes (of order 0.05Å) can be used and convergence is then slow.

9.8.2. Network glasses.

Si in silicate glasses are coordinated to four O. The SiO_4 tetrahedra share O at each corner (bridging oxygens) with other tetrahedra. In pure SiO_2 all oxygens are bridging; when alkali oxide is added some corner O are no longer shared and these are known as non-bridging. Tetrahedra with n bridging O are known as Q_n species, and their relative proportions can be found from MAS NMR data.

To model such glasses we make an initial configuration of random Si at an appropriate density. Coordination constraints of 4,3 and 2-fold coordination (corresponding to the proportions of Q_n species) are applied between suitable distances and a hard sphere simulation is run until the constraints are satisfied. In the later stages of this process it may be necessary to move over- and under-coordinated atoms around the configuration by hand, otherwise the process may take a very long time. O atoms are added at the mid-point of each Si-Si bond, and single O atoms are added close to all Si atoms which are then coordinated to less than four O. The resulting configuration now has the correct topology, but the Si-O bond lengths are all wrong. A hard sphere simulation is run, with appropriate constraints on Si-O coordination to maintain the Si-O bonds, until all closest approach constraints are satisfied. The data can then be fitted. Because the constraints are extremely severe only small move sizes (of order 0.05Å) can be used and convergence will be slow.

9.9. Metropolis Monte Carlo

The RMCA program may also be used for conventional MMC simulation with either hard spheres or a potential. Since the imposition of closest approach distances for atoms is equivalent to having hard spheres, for a hard sphere simulation these distances should be chosen as appropriate and no experimental data specified, i.e. the number of experimental data sets is zero. In most cases it would be recommended that a suitable hard sphere simulation be run before any attempt is made to fit to the data.

When a potential is used it is defined in a table at the same r interval as used to define $g(r)$. This makes the calculation of the energy very fast, but is not a suitable method for a ‘proper’ MMC simulation. In this case it is only intended that the potential be used to produce an initial configuration, or as an additional constraint for the RMC procedure.

9.10. Efficient use of RMC

The time taken by the RMCA program depends on the number of atoms in the configuration, the number of r points used and the number of data points and data sets. For any particular application the time required can vary from hours to months, so it is sensible to have a strategy which will reduce this as much as possible. The general approach should be as follows.

1. Use parameters appropriate for the problem, i.e. do not use enormous configurations (though always more than 1000 atoms) etc.
2. Create an initial configuration which satisfies all constraints. If only closest approach constraints are being used then this may be possible using the MOVEOUT program. Otherwise RMCA must be run with no experimental data. If coordination constraints are being used then this first step may take a long time.
3. If using diffraction data only then initially fit to $G(r)$ or $g_{\alpha\beta}(r)$ (preferably obtained using MCGR). This is better than fitting directly to the structure factor as the transform from r to Q is expensive. Fit to all sets of diffraction data available. Fitting to single sets and then adding others does not generally save time, unless the data sets contain very similar information anyway. This can be assessed using the program PARTIALS.
4. Fit to a subset of Q points in the structure factors, e.g. at $0.1\text{\AA}^{-1}$ intervals, unless you are fitting to diffraction data from crystals when the resolution provided by the experimental Q spacing is required.
5. Add in EXAFS data if it is being used. Since EXAFS only provides information on short range order it is generally safe to obtain a good fit to diffraction data first. However this will not always be the case and it may then be necessary to relax the fit to the diffraction data before the EXAFS data can be fitted satisfactorily.
6. Finally fit to all the experimental data points, though do not use more than are justified by the experimental resolution.

10. The RMCA program

10.1. Running the program

The current version of RMCA (version 3) is written as a general purpose program for modelling multi-component systems using experimental diffraction data as a constraint. When RMCA is run it must be supplied with a name used for all its data files. This is usually done (depending on the computer on which the programme is installed) on the command line, for example

```
rmca cscl
```

would run the RMCA program with files called `csc1` with various extensions. This name is used for a number of files used by the program. The general program parameters are supplied in `name.dat`. The first time of running there must be a file `name.cfg` containing the positions, or configuration, of the atoms. It is only necessary to calculate $g(r)$ once at the beginning of the run, thereafter only the change in $g(r)$ needs to be calculated. To facilitate this the histograms representing $g(r)$, as well as a binary representation of the atomic coordinates, are written out to a file `name.his` when the program terminates. On the next run these histograms, and the atomic coordinates will be read from `name.his` rather than from the configuration file unless any parameters have been changed which require the $g(r)$ histogram to be recalculated. Saved configurations will be written to `name.sav` and the output results will be written to `name.out`. Information on the status of the program is written to `name.log`.

10.2. The configuration file

The configuration is normally cubic, but in general any parallelepiped can be used. If a crystalline system is being modelled then obviously the box dimensions must be chosen to accomodate an integral number of unit cells in each direction, with the number of cells being chosen to make the dimensions as equal as possible. The atomic coordinates are defined in terms of the box coordinates and are normalised to limits of ± 1.0 . The box coordinates are then defined in terms of the laboratory coordinates by the matrix given in the top of the configuration file. The values in this matrix are strictly only required to be relative; their absolute values are recalculated from the density every time the program is run and these are then written into the new configuration file. However some programs for analysis of configurations do not check the density (and may produce erroneous results) so it is sensible to always use absolute values. For a cubic box the diagonal terms in the matrix are equal to half the box length ($L/2$) in Å.

The numbers of different types of atoms and the order of their coordinates are given in the top of the configuration file. The atomic coordinates then follow in a single list. The choice of a format in which the system is not explicitly identified in the configuration file (other than by the title) is made deliberately; configurations can then be easily modified or used as starting configurations for different systems with a minimal amount of editing. The same format of configuration file is also used for molecular systems, where both coordinates and Euler angles define the molecules. An example configuration file is shown in Appendix I.

10.3. The data file

The RMCA data file `name.dat` can be thought of as being divided into various sections. The description that follows is given section by section. An example data file is shown in Appendix II.

10.3.1. General parameters.

title	character*80	A title for the run.
rho	real	Number density of system in atoms per cubic Angstrom.
rcut	real array	The closest allowed approach of two particles. There should be a value for each partial $g(r)$. There are thus $n(n+1)/2$ values for a n -component system. As in all other parts of the programme the partials are given in the order 1-1, 1-2, ..., 1- n , 2-2, ..., 2- n , ... n - n . See section 9.2 for a description of the use of this parameter.
delta	real array	The maximum move for each type of particle. Recommended values are in the range 0.05-0.5. A value of zero is allowable, in which case the program will not attempt to move those particles. See section 9.3 for a description of the use of this parameter.
dr	real	The spacing for calculating $g(r)$. See section 9.4 for a description of the use of this parameter.
moveout	logical	This should be set <code>.true.</code> if, and only if, there are some particles that do not satisfy the cut-off restrictions and you want to preferentially choose these to be moved. 10% of the time one of these particles will be chosen, and 10% of the time any particle will be chosen.
ncoll	integer	The number of configurations to collect after convergence. The program will stop when ncoll configurations have been collected.
iprint	integer	Determines how often a summary will be written to the standard output. It will be written after every iprint moves generated, except that this will only occur when a move is accepted. iprint should not be too small or otherwise time will be wasted in continually writing to the <code>.log</code> file and this will become very large.
timelim	real	The time the programme should run for, in minutes.
timesav	real	The interval at which the results should be saved to the output files (they are always saved when the programme ends anyway). If the programme is left running for a very long time it is best to save the results every now and then, perhaps once an hour. The results should not be saved too often as this simply wastes time writing large files.

10.3.2. Parameters referring to all data sets.

nexpt	integer array	Four values, being the number of experimental data sets of different types. The first is the number of $G(r)$ constraints, the second the number of $S(Q)$'s, the third the number of $F(Q)$'s, (for the purpose of this programme the difference between $S(Q)$ and $F(Q)$ is that the former has constant coefficients of partial structure factors whereas the latter has Q dependent coefficients as may be the case with x-ray data), and the fourth the number of EXAFS data sets. Definitions of these quantities in terms of partial structure factors and radial distribution functions are given in section 9.7.
-------	---------------	--

10.3.3. Parameters for each $G(r)$ constraint.

file	character*80	File containing the experimental data.
nx1,nx2	integers	The indices of the first and last data point to be used. If nx2 is larger than the number of data points the last data point to be used will be the last data point given. If the r value corresponding to nx2 is larger than $L/2$ then nx2 will be reduced appropriately.
c	real	A constant to be subtracted from the data on input.
coeff	real array	The coefficients of each of the partial $g(r)$'s.
sigma	real	The standard deviation to be used.
renorm	logical	.true. if data renormalisation is required.

10.3.4. Parameters for each $S(Q)$ constraint.

file	character*80	File containing the experimental data.
nx1,nx2	integers	The indices of the first and last data points in the experimental data file. If nx2 is larger than the number of data points the last data point to be used will be the last data point given.
c	real	A constant to be subtracted from the data on input.
coeff	real array	The coefficients of each of the partial structure factors.
sigma	real	The standard deviation to be used.
renorm	logical	.true. if data renormalisation is required.
offset	logical	.true. to subtract a constant from the data.

10.3.5. Parameters for each $F(Q)$ constraint.

file	character*80	File containing the experimental data and the coefficients of the partial structure factors.
------	--------------	--

nx1,nx2	integers	The indices of the first and last data point to be used. If nx2 is larger than the number of data points the last data point to be used will be the last data point given.
c	real	A constant to be subtracted from the data on input.
sigma	real	The standard deviation to be used.
renorm	logical	.true. if data renormalisation is required.
offset	logical	.true. to subtract a constant from the data.

10.3.6. Parameters for all EXAFS data (only present if there is some).

rmax	real	Maximum r value to be used.
weight	integer	Data are to be weighted by Q^{weight} . The experimental data should be supplied unweighted, as weighting is done within the program. Data and fits written in the .out file will have the required Q weighting.

10.3.7. Parameters for each EXAFS constraint.

file	character*80	File containing the experimental data.
nx1,nx2	integers	The indices of the first and last data point to be used. If nx2 is larger than the number of data points the last data point to be used will be the last data point given. When EXAFS data is modelled together with diffraction data the values of nx1 and nx2 <i>must</i> correspond to Q values which are a subset of the <i>same</i> Q values as in the data files for the diffraction data, and similarly in the file of EXAFS coefficients, as the Q values will be taken from the diffraction data files.
nx3,nx4	integers	The indices of the first and last data points to be used in calculation of χ^2 for the EXAFS data. nx3 and nx4 must lie within the range of nx1 and nx2. These additional parameters are given for EXAFS data so that different ranges of the EXAFS data can be modelled without having to recalculate the EXAFS coefficients.
edge	integer	Index number of the particle of the absorption edge being used.
file	character*80	File containing the EXAFS coefficients.
sigma	real	The standard deviation to be used.
renorm	logical	.true. if data renormalisation is required. It is not recommended that this option be used with EXAFS

		data since renormalisation by a negative number, corresponding to a phase shift of π , may occur.
offset	logical	.true. to subtract a constant from the data.

10.3.8. Parameters for coordination constraints.

ncoord	integer	The number of coordination constraints.
--------	---------	---

10.3.9. Parameters for each coordination constraint (one line per constraint).

typec	integer	The type of the central particle.
typen	integer	The type of the neighbour particles.
rcoord	real	The two distances between which to calculate the coordination number.
coordno	real	The desired coordination number.
coordfrac	real	The fraction of the central particles desired to have this coordination number.
sigmac	real	Effectively a parameter weighting this constraint relative to others and the fit to the data.

10.3.10. Parameters for average coordination constraints (one line per constraint).

ncoord	integer	The number of average coordination constraints.
--------	---------	---

10.3.11. Parameters for each average coordination constraint.

typec	integer	The type of the central particle.
typen	integer	The type of the neighbour particles. Note that for the average coordination number constraints, unlike the constraints on coordination number distribution, there is no difference in the constraint obtained by swapping neighbour and central particles except in the value of coordination number required, as the constraint is simply obtained by integration of $g_{ij}(r)$ and $g_{ji}(r) = g_{ij}(r)$.
rcoord	real	The two distances between which to calculate the coordination number.
coordno	real	The desired average coordination number.
sigmac	real	Effectively a parameter weighting this constraint relative to others and the fit to the data.

10.3.12. Parameters concerning a potential.

usepot	logical	.true. if using a potential. The remaining parameters in this section are only present if we are.
--------	---------	---

<code>temp</code>	<code>real</code>	The absolute temperature.
<code>eunits</code>	<code>real</code>	The energy units the potential is supplied in (in other words the conversion factor to Joules).
<code>weight</code>	<code>real</code>	The weighting factor for the potential. To use the programme for a conventional Metropolis Monte Carlo simulation this should be 1.
<code>file</code>	<code>character*80</code>	The data file containing the potential tabulated at intervals of dr .

11. The experimental data files

The files containing the experimental data are in the DATA format as defined for the NDP series of programs. The number of points is given on the first line, the second line contains a title or other information, and the subsequent lines contain the Q , $S(Q)$ or r , $g(r)$ values in two columns. An example file is shown in Appendix III.

If the coefficients of the partial structure factors are to be included in the files, i.e. for $F(Q)$ data as defined in subsection 9.7, then they come after the $F(Q)$ values in the order 11,12,...1n,22,23,...2n,...nn. For X-ray data files in this format can be produced from files in the DATA format using the program XCOEFF. This has the option to normalise the form factors in three different ways, depending on how the original data have been defined. An example file is shown in Appendix III.

Note that if more than one set of experimental data is supplied they must all be defined at the same Q , or r , points. You only need to use a subrange of these points so it is possible to use data sets that cover different Q ranges provided they are defined at the same Q points where they overlap. The NDP program REBIN can be used to produce data sets satisfying this requirement.

12. Other programmes available

There are three other RMC codes available. RMCX fits to single crystal diffuse scattering data, RMCM uses semi-rigid molecules and RCMAG models magnetic diffuse scattering. RMCPow, for crystal structure refinement based on powder diffraction data (Bragg scattering), is under development.

13. Installing the RMC programmes

13.1. Source code

The program is written in a slightly extended Fortran-77 and compiles under VAX Fortran as well as some other versions of the language. Non-standard features used are names longer than 6 characters, use of underscore character, use of `DO WHILE` and `DO...ENDDO` constructs. Most of the code is contained in one file. The configuration file reading and writing routines and some machine-dependent timing routines are found in two other files.

The configuration of atoms is stored in a file with a standard format suitable for systems containing any number of components and for molecular systems. As this file will need to be read by any analysis programs it is convenient to keep the subroutines for reading and writing it separate from the main programme. So that the identical code for the main programme can be used on any of the machines on which we wish to run it all the machine dependent parts are contained in separate subroutines. A random number generator must be supplied. In VAX Fortran the function `ran(seed)` generates a random number in the interval 0 to 1. Other versions of Fortran may have equivalents. If not code for such a function is included.

The programme is written to be general purpose and allocate array space when it runs. The total array space available is specified by a parameter statement near the beginning of the code and this may need to be changed according to the computer system being used.

13.2. Compiling and running the programme

The code is contained in the following files:

<code>rmca.for</code>	- The main programme
<code>rmc_cfg.for</code>	- The configuration reading and writing routines
<code>rmc_vax.for</code>	- The machine dependent routines, VAX version
<code>ran.for</code>	- A random number generator (not needed for VAXes)

all of which should be compiled and linked (after any necessary alterations) in whatever manner is appropriate for your computing system. It is intended that the programme be run in such a way that the data file name can be supplied on the command line. The Unix version of the machine dependent routines allows this with no further ado. On VAX/VMS systems it is necessary to define a symbol to run the programme in the following way:

```
rmca := $user$disk : [user.rmc]rmca.exe
```

and then the programme can be run by saying, for example

```
rmca ybco
```

13.3. Example files:

There are some example files available which demonstrate the use of the program and which can be used for testing. The example is molten copper. The configuration is small, only 250 atoms, partly so that the programme runs quickly and partly because larger files can be delayed during transfer by E-mail. The example consists of the following files:

<code>cu.dat</code>	- input data
<code>cusq.dat</code>	- experimental structure factor
<code>cu.cfg</code>	- configuration of 250 atoms
<code>cuconv.cfg</code>	- converged configuration

The file `cu.cfg` contains a suitable starting configuration which can be used to run the programme. Convergence is rapid for such a small configuration. An example configuration after convergence is in `cuconv.cfg`. Note that the agreement with the experimental data is not exceptionally good at convergence because of the small size of the configuration; this is because the oscillations in $g(r)$ have not died out by the largest distance at which $g(r)$ can be calculated, i.e. half the box size. However this suffices as a quick example. You will find that if you use a larger configuration (say 1000-2000 atoms) it is possible to fit the experimental data very well indeed.

14. Using RMCA

RMCA is *not* a 'black box' program. It *will not* tell you the 'correct' structure of any material and it *will not* produce wonderful results from poor data. What RMCA *will* do is allow you to explore, in a controlled manner, what information on the three dimensional structure can actually be obtained from the experimental data provided. To do this it must be used in a sensible and thoughtful manner. The information obtained must also be used in an unbiased fashion. If the RMC model does not agree with a preferred theory then this should not simply be ignored. If RMC modelling produces significantly different structures that agree with the same data then this is not a fault of the method, but rather it illustrates the inadequacy of the data. The different RMC models should then be used to predict whether different, or more accurate, experiments might distinguish between the model. In this way progress can be made.