Reverse Monte Carlo methods

MARTIN T. DOVE1*, MATTHEW G. TUCKER1, STEPHEN A. WELLS1 and DAVID A. KEEN2

1 Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, U.K.
2 ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, U.K.; *e-mail: martin@esc.cam.ac.uk

Introduction

Most modelling techniques in the solid-state sciences are what would be called forward modelling. The starting point is an exact expression for a Hamiltonian or free energy, together with a way of using this to generate numerical quantities that can be compared with experiment. The starting expression could be a representation of the Schrödinger equation, or a numerical approximation to the forces between atoms. The models can also be based on differential equations to incorporate time-dependence (e.g. the Lagrangian of classical mechanics), or the partition function to obtain thermodynamic averages. The solutions to the models can be exact in some cases (particularly when the models are used to describe a system at zero temperature), but may frequently have some degree of statistical uncertainty. For example, in molecular dynamics, the dynamic equations are solved using discrete time steps over a limited time span, and in Monte Carlo simulations only a portion of the total phase space available to a system is sampled. Whatever approach is taken, the end point of forward modelling may be to provide new insights, to predict behaviour, or to interpret observations. A key point towards this end is to be able to reproduce some experimental data in order to check that the simulations are representative of reality, even though the true value of a simulation is to provide information, insight or understanding that cannot be extracted from experimental data.

In this chapter we will consider a second type of modelling technique, which is called inverse modelling. This is best described by comparison with forward modelling. In forward modelling a set of basic equations is used to generate atomic configurations from which numerical quantities can be calculated for comparison with experimental data. In inverse modelling, the starting point is a set of experimental data, and atomic configurations are generated by a procedure explicitly designed to give best agreement with experimental data. Inverse modelling bypasses the need for any representations of the interatomic forces; clearly by design good agreement with experiment is expected. As for forward modelling, inverse modelling has the objective to provide understanding
of phenomena, the focus in this case being the interpretation of experimental data and the search for insights behind the raw data.

Inverse modelling methods are currently restricted to generation of static configurations (noting that only the harmonic lattice dynamics and molecular dynamics methods of the set of forward modelling tools give dynamic information). In this chapter we will focus on using Monte Carlo methods with inverse modelling, the so-called Reverse Monte Carlo (RMC) methods. Our main example will be for the analysis of neutron total scattering data (that is, neutron scattering data that includes both the Bragg scattering and the diffuse scattering), but we will also consider an approach that is used for studies of cation ordering with NMR data.

Total scattering experiments

The phrase “total scattering experiment” refers to a measurement of the scattering of radiation by matter that covers all scattering vectors (i.e. all values of \( \sin \theta/\lambda \)) and includes scattering with all possible changes of energy of the radiation. It therefore encompasses elastic scattering, such as from Bragg peaks, which arises from the static or mean atomic scattering, and inelastic scattering, which arises from dynamic processes. The Fourier transform of the total scattering measurement provides information about the relative positions of atoms, which can usually only be interpreted over short distances (Billinge & Thorpe, 1998; Dove, 2002).

Until recently, total scattering experiments were mostly associated with studies of fluids or glasses (Chieux, 1978; Wright, 1993, 1994, 1997). In contrast, diffraction studies of crystalline materials tend to be primarily focused on the measurements of the Bragg peaks, with little concern for the shape of the background provided that it could be fitted by an appropriate polynomial (Pavese, 2002; Redfern, 2002). The Bragg peaks give information about the distributions of positions of atoms within the unit cell, and for many purposes this is exactly all the information that is required. Since fluids and glasses do not have long-range periodic order, there are no Bragg peaks. One of the exciting developments in crystallography over recent years has been the application of total scattering methods to crystalline materials (Billinge & Thorpe, 1998), particularly for crystalline materials that have a high degree of structural disorder. The subsequent coupling of total scattering measurements to modelling through the RMC method has extended the opportunities for studying disordered crystalline materials at an atomistic level.

The information contained within the Bragg scattering and total scattering can be appreciated by considering the basic scattering equations. The starting point is the static scattering function, \( S(Q) \) (Bacon, 1975; Dove, 2002):

\[
S(Q) = \frac{1}{N} \sum_{j,k} \langle b_j b_k \exp(iQ \cdot (r_j - r_k)) \rangle,
\]

where \( b_j \) is the scattering factor for atom labelled \( j \), \( r_j \) is the instantaneous position of this atom, and \( N \) is the number of atoms in the sample. \( Q \) is the scattering vector, defined as the change in wave vector of the neutron beam associated with the scattering process.
(note that the wavelength of the scattered beam can change through the scattering process). For coherent scattering (i.e. where all atoms of the same type scatter the same way), this can be rewritten as

\[ S(Q) = \frac{1}{N} \sum_{j,k} \overline{b_j b_k} \left\langle \exp(iQ \cdot [r_j - r_k]) \right\rangle, \]  

(2)

where the overline represents the average over all atoms of the same type. The main point of this equation is that it shows how the intensity of scattering is determined directly by the instantaneous distances between atom positions, and does not directly contain information about the actual positions of individual atoms. For periodically ordered systems, the information about the positions of individual atoms is contained within the Bragg peaks. The equation for Bragg scattering is

\[ S_{\text{Bragg}}(Q) = \frac{1}{N} \left| \sum_j \overline{b_j} \left\langle \exp(iQ \cdot r_j) \right\rangle \right|^2. \]  

(3)

If we consider an atom to have a mean position, \( \bar{r}_j \), we can write the average as

\[ \left\langle \exp(iQ \cdot r_j) \right\rangle = \exp(iQ \cdot \bar{r}_j) \int p(r - \bar{r}_j) \exp(iQ \cdot r) dr, \]  

(4)

where \( p \) is a probability distribution function, and for a harmonic crystal it is a simple Gaussian function (Willis & Pryor, 1975), with a Fourier transform (i.e. the term in the integral) that is also a Gaussian.

The objective of total scattering experiments is to determine the distribution of interatomic distances. Indeed, the atomic structures of fluids and glasses can only reasonably be described in terms of the interatomic distances. Since fluids and glasses are isotropic, the scattering function is independent of the direction of \( Q \). The scattering function is therefore better described by averaging over all orientations of \( Q \), leading to

\[ S(Q) = \frac{1}{N} \sum_{j,k} \overline{b_j b_k} \frac{\sin(Q |r_j - r_k|)}{Q |r_j - r_k|}. \]  

(5)

This result is derived in the Appendix. It is useful to recast the formalism in terms of the distribution of interatomic positions rather than as a sum over all pairs of atoms. First we subtract out the terms where \( j = k \) to give

\[ S(Q) = i(Q) + \sum_m c_m b_m^2, \]  

(6)

where the first term will describe pairs of atoms and will be considered in more detail below, and where \( c_m \) is the proportion of atoms of type \( m \). The first term can be written as

\[ i(Q) = \rho_0 \int_0^{\infty} 4\pi r^2 G(r) \frac{\sin Qr}{Qr} dr, \]  

(7)

where the new function \( G(r) \) describes the distribution of interatomic distances:
and $\rho_0$ is the number of atoms of any type per unit volume. The individual pair distribution functions are defined as

$$g_{mn}(r) = \frac{n_{mn}(r)}{4\pi r^2 \rho_m dr},$$

(9)

where $n_{mn}(r)$ is the number of atoms of type $n$ lying within the range of distances between $r$ and $r + dr$ from any atom of type $m$, and $\rho_m = c_m \rho_0$. It is common to define

$$D(r) = 4\pi \rho_0 G(r)$$

(10)

so that

$$Q_i(Q) = \int_0^\infty D(r) \sin Qrdr.$$  

(11)

The reverse transform is then given as

$$D(r) = \frac{2}{\pi} \int_0^\infty Q_i(Q) \sin QrdQ$$

(12)

(Wright, 1993, 1994, 1997; Chieux, 1978; Dove, 2002). This transform provides the means by which the information about structure over short length scales, as encapsulated in the function $D(r)$, can be extracted from the experimental measurements.

The experimental task (Wright, 1993, 1994, 1997) is to obtain the best measurements of $Q_i(Q)$ from the total scattering data. It is not within the purpose of this paper to explain the experimental details; these have been documented elsewhere (Wright, 1993; Howe et al., 1989; Dove et al., 2002). However, it is essential to appreciate that it is important to determine $Q_i(Q)$ to a high value of $Q$ (typically 50 Å$^{-1}$) in order to achieve the best possible resolution in $D(r)$: the resolution $\Delta r$ is given as $2\pi/\bar{Q}_{\text{max}}$, where $\bar{Q}_{\text{max}}$ is the maximum value of $Q$ achieved in the measurement of $Q_i(Q)$. It is also important to appreciate that it is necessary to have an absolute measurement of $Q_i(Q)$ if the resultant $D(r)$ is to be interpreted quantitatively. It is essential that all sources of additional scattering and all sources of signal attenuation can be independently determined and taken into account in the treatment of the data (Wright, 1993; Howe et al., 1989; Dove et al., 2002).

It should be noted at this stage that there is a confusion in the literature in that different authors use different sets of symbols for the quantities discussed in this article, including the use of $G(r)$ for what we have called $D(r)$. This is a long-standing historic problem; Keen (2001) gives a good comparison of the different ways of labelling the fundamental quantities.

**Inverse methods for the determination of $D(r)$**

The forward Fourier transform of $Q_i(Q)$ to obtain $D(r)$ is straightforward in principle, but is subject to a problem that arises from the fact that the data for $Q_i(Q)$ extend only to some maximum value of $Q$. The effect of the finite range of $Q$ on the Fourier transform
is to introduce spurious ripple peaks into the computed $D(r)$ function. This problem can be avoided by multiplying $Q_i(Q)$ by a modification function $M(Q)$ that decreases smoothly to zero at $Q_{\text{max}}$. However, the Fourier transform of the modified form of $Q_i(Q)$ will then be convoluted with the Fourier transform of $M(Q)$, which means that the peaks in $D(r)$ will be artificially broadened. In order to overcome this problem an inverse Monte Carlo method can be used to obtain the $D(r)$ whose Fourier transform best matches the experimental $Q_i(Q)$ data (Soper, 1989; Pusztai & McGreevy, 1997).

The idea is straightforward. A trial form of $D(r)$ is set up on a set of discrete values of $r$ that spans a very wide range of $r$:

$$D_{\text{trial}}(r) = \begin{cases} -4\pi r \rho_0 \sum_{m,n} c_m c_n \vec{b}_m \cdot \vec{b}_n, & r < r_c \\ 0, & r > r_c \end{cases},$$

where $r_c$ is typically of the size of the shortest interatomic distance. This means that all the individual values of $g(r)$ are zero for $r$ up to some lower bound (which is usually just lower than the expected minimum interatomic distance), and 1 for larger distances (implying a completely random distribution of atoms). The Fourier transform of $D(r)$ can be compared with the experimental $Q_i(Q)$. Because the trial form of $D(r)$ spans a wide range of $r$, certainly well within the range for which $D(r) = 0$, there are no truncation ripples in the Fourier transform. The values of $D(r)$ can then be modified in order to improve the agreement with the experimental $Q_i(Q)$, and the ideal tool for this is the Monte Carlo method.

To generalise the discussion for later use, we write any experimental quantity as $y_{\text{exp}}$, and the corresponding calculated quantity as $y_{\text{calc}}$. We define an agreement factor as

$$\chi^2 = \sum_j (y_j^{\text{exp}} - y_j^{\text{calc}})^2 / \sigma_j^2,$$

where we sum over all data points (labelled by $j$), and $\sigma_j$ is a weighting factor that may correspond to the experimental uncertainty on $y_j$. Clearly the best calculation is that for which the value of $\chi^2$ is a minimum, as in any data fitting technique. In the Monte Carlo approach, the calculated values of $y$ are changed by random amounts. If the change lowers the value of $\chi^2$, the change is accepted. On the other hand, if the change causes $\chi^2$ to increase by an amount $\Delta \chi^2$, the change is not automatically rejected, but accepted with the probability

$$P = \exp(-\Delta \chi^2 / 2).$$

This ensures that the model does not get trapped in a local minimum, and instead the model will converge on the global minimum.

So far we have not made any allowance for the form of the experimental data. In practice $Q_i(Q)$ is measured for different banks of detectors over different ranges of $Q$, and each bank of detectors will be subject to a different resolution function. These are easily taken account of in the inverse modelling procedure (Tucker et al., 2001a), in a way that would not be possible when simply attempting to perform a direct Fourier transform of $Q_i(Q)$. 

Reverse Monte Carlo methods

63
Illustrative results of this inverse procedure for crystalline AlPO$_4$ are given in Figures 1 and 2 (Tucker et al., 2001a). Figure 1 shows a set of experimental values of $Q_i(Q)$ measured on different banks of detectors using the GEM diffractometer at the ISIS spallation neutron source (Williams et al., 1997), each with a different range of $Q$ and different resolution. Figure 2 shows the $D(r)$ function that fits each data set. It should be noted that the final form of $D(r)$ has high resolution and no truncation ripples; of note is the resolution of the two peaks associated with Al–O and P–O nearest-neighbour distances.

The Reverse Monte Carlo modelling of atomic structures from total scattering data

The Reverse Monte Carlo method uses the formalism outlined above to obtain atomic configurations that best match experimental total scattering data (McGreevy & Pusztai, 1988; McGreevy, 1995; Mellergård & McGreevy, 1999, 2000; McGreevy, 2001). In this case, the atomic coordinates are varied in a random manner in order to obtain the best agreement with experimental data. The experimental data, corresponding to the values of $y$ in the equation for $\chi^2$, can be $Q_i(Q)$ or $D(r)$. In fact, in our work we use both at the same time. We write the RMC $\chi^2$ in the following form:
where we define a separate $\chi^2$ for each set of data. The last term corresponds to the inclusion of data-based constraints (Keen, 1997, 1998), where $f$ may be a bond length or bond angle, with the required value obtained from the low-$r$ peaks in $D(r)$. The profile term was introduced by ourselves for the study of crystalline materials (Tucker et al., 2001b, 2002a, 2002b). The function $I_{\text{profile}}(t)$ describes the Bragg diffraction pattern. Our work is mostly based on time-of-flight neutron sources, so $I_{\text{profile}}$ is a function of neutron flight time $t$. This is the same $\chi^2$ that is minimised by a least-squares technique in Rietveld refinement.

$$
\chi^2_{\text{RMC}} = \sum_m \chi^2_m
$$

$$
\chi^2_Q(Q) = \sum_k \sum_j \left[ Q_{\text{calc}}(Q,j)_k - Q_{\text{exp}}(Q,j)_k \right]^2 / \sigma^2_k(Q)
$$

$$
\chi^2_{D(r)} = \sum_j \left[ D_{\text{calc}}(r,j) - D_{\text{exp}}(r,j) \right]^2 / \sigma^2(r)
$$

$$
\chi^2_{\text{profile}} = \sum_k \sum_j \left( I_{\text{profile}}(j)_k - I_{\text{exp}}(j)_k \right)^2 / \sigma^2_k(j)
$$

$$
\chi^2_f = \sum_t \left( f_{\text{calc}}(t)_t - f_{\text{req}}(t) \right)^2 / \sigma^2_t
$$

**Fig. 2.** Pair distribution function $D(r)$ for AlPO$_4$ obtained from the $Qi(Q)$ data of Figure 1 using the inverse transformation method described in the text (Tucker et al., 2001a).
The starting point in an RMC study is an initial configuration of atoms. When modelling crystalline materials, this configuration will have atoms in their average crystallographic positions, and will contain several unit cells (perhaps of order of $10^3$ unit cells). If modelling non-crystalline materials, an initial algorithm will be required to generate a random distribution of atoms without unreasonably short interatomic distances. Atoms are selected at random, and moved small random distances. If the move reduces $\chi^2_{\text{RMC}}$, it is accepted. If the move increases $\chi^2_{\text{RMC}}$, it is accepted with the probability

$$P_{\text{RMC}} = \exp(-\Delta \chi^2_{\text{RMC}} / 2)$$

This algorithm is the same as that used to extract $\mathcal{D}(r)$ from the experimental $\mathcal{Q}(Q)$ as described above. Moreover, the RMC method allows us to account properly for the effects of experimental resolution (Tucker et al., 2002a, 2002b).

Part of the reason for including as large a range of experimental data as possible in the RMC analysis is associated with the fact that the RMC method is effectively a method based in statistical mechanics. As a result, an RMC simulation will evolve to maximise the amount of disorder (entropy) in the configurations. Thus the RMC simulation will give the most disordered atomic configurations that are consistent with the experimental data. There may be a range of configurations that match the data, with different degrees of disorder. Only by maximising the range of experimental data can this problem be minimised. In our work, we include data that act as constraints on both the short-range and long-range structures, together with constraints on the topology of the SiO$_4$ tetrahedra. With regard to the latter, the $\mathcal{D}(r)$ data show that the average coordination number of the silicon atoms is 4. We believe all silicon atoms have a coordination number of 4, but the average allows some degree of fluctuation. The bond constraints impose the interpretation that there are no fluctuations from the average coordination number. From a number of tests, we believe that our RMC simulations do not generate artificial disorder. For example, we find that our configurations become ordered for data obtained at lower temperatures, with no sign of any disorder at the lowest temperatures (this point will be illustrated with some of the examples given in the following section).

Examples of RMC modelling from total scattering data

Cristobalite

Cristobalite is an interesting example for RMC modelling since it is known that the high-temperature cubic phase has substantial disorder of the positions of the oxygen atoms. In the average structure deduced by simple crystallographic analysis, the oxygen atoms lie exactly halfway between their two neighbouring silicon atoms. Apart from the fact that this implies linear Si–O–Si bonds, whereas it is more usual for the angle subtended at the oxygen atom to be closer to 145°, the Si–O bonds in this description are anomalously short, and the refined thermal displacement parameters suggest a significant amplitude of motion of the oxygen atoms in directions perpendicular to the Si–O–Si linkage (Schmahl et al., 1992; Swainson & Dove, 1993). In fact, such motions of the oxygen
atoms would bend the Si–O–Si bond and increase the Si–O distance, giving a more
typical local structure. It is probable that the motions of the oxygen atoms will involve
rotations of the SiO₄ tetrahedra. How this could be achieved has been subject to some
controversy, but our total scattering studies (Dove et al., 1997; Tucker et al., 2001c) have
tended to confirm the ideas contained within the “Rigid Unit Mode” model discussed
below. The results are also consistent with molecular dynamics simulations (Swainson
& Dove, 1993, 1995); several studies we report in this article have also been studied by
the molecular dynamics simulation method, providing an example of the
complementarity of forward and inverse methods.

Figure 3 shows configurations of the SiO₄ tetrahedra in cristobalite. The cubic
phase has considerable orientational disorder of the tetrahedra without clear distortions.
Moreover, it is clear that this disorder is sustained without the formation of short-range
structural correlations that extend much beyond a single ring of tetrahedra (such as the
formation of domains of a lower-symmetry phase). The change in orientational disorder
on cooling into the tetragonal low-temperature phase is clear to see. There are a number
of ways to quantify the changes in the structure on changing temperature. In Figure 4 we
show the probability distribution function for the Si–Si–Si angle. This function has a
single broad peak for the configurations of the high-temperature phases, but there are

Fig. 3. Atomic configurations of cristobalite obtained from RMC analysis of neutron total scattering data
(after Tucker et al., 2001c). The configurations are shown as linked tetrahedra representing the SiO₄ groups.
two significant side peaks in the configuration of the low-temperature phase. This shows that there are significant changes in the short-range structure caused by the phase transition. More details of the results of this study are given in Tucker et al. (2001c).

The important insight from the RMC study has been to compare the structures of the two phases of cristobalite, showing that there is considerable orientational disorder in the $\beta$-phase. The question of how a network structure can sustain this disorder had previously been addressed by our Rigid Unit Mode (RUM) model (Giddy et al., 1993; Hammonds et al., 1996). RUMs are normal modes of vibration whose eigenvectors do not involve any distortions of the SiO$_4$ tetrahedra. Because they do not cause any high-energy deformations, they have low frequency and hence high amplitude (examples of animations of RUM motions, including those for $\beta$-cristobalite, have been posted as mpeg files on http://www.esc.cam.ac.uk/movies). A normal mode analysis of the RUM spectrum of cristobalite has shown that there are many more RUMs in the high-temperature phase than in the low-temperature phase; this is quite common and can be explained in terms of the additional symmetry in the high-temperature phase (Giddy et al., 1993; Swainson & Dove, 1993; Hammonds et al., 1996). In the case of cristobalite, we had earlier postulated that the superposition of many RUMs would be capable of creating a dynamically disordered state in which the SiO$_4$ tetrahedra do not undergo significant distortions (Swainson & Dove, 1993). This appears to be supported by the RMC simulations. Moreover, calculations of the three-dimensional patterns of diffuse scattering from the RMC configurations (Tucker et al., 2001c) are consistent with the prediction of the RUM model (Swainson & Dove, 1993; Hammonds et al., 1996), as well as with electron diffraction data (Hua et al., 1988).

Fig. 4. Si–Si–Si angle distribution function for cristobalite obtained from the RMC analysis (after Tucker et al., 2001c). The key point from these results is that the function has a single peak for the disordered $\beta$-phase but a central peak and two side peaks for the ordered phase. This result shows that there is a structural difference between the two phases even over the length scale of three linked SiO$_4$ tetrahedra.
Quartz

Measurements of the crystallographic Si–O distance in quartz on heating through the phase transition suggest that a similar type of disorder also occurs in quartz. The distance between the mean silicon and oxygen nearest-neighbour positions decreases significantly on heating up to the phase transition, whereas the mean instantaneous Si–O distance obtained from $D(r)$ data (Fig. 5, Tucker et al., 2000, 2001d) increases gradually on heating with no changes associated with the phase transition. The comparison of the two distances suggests that there will be similar orientational disorder of the SiO$_4$ tetrahedra that is allowed to set in as a result of the phase transition. Almost simultaneously with our RMC study, Muser & Binder (2001; see also Kihara, 2001) performed a molecular dynamics simulation that gave results that are fully consistent with the interpretation of the phase transition that emerges from our RMC study (Tucker et al., 2000, 2001d).

RMC configurations for three temperatures are shown in Figure 6. The lowest temperature shows a well-ordered structure. The second temperature is still within the low-temperature phase, but much closer to the phase transition. It is clear that the average structure is that of the low-temperature phase, but there is considerable orientational disorder. Finally, the highest temperature configuration is in the high-temperature phase, and it can be seen that the average structure is associated with a high degree of orientational disorder of the SiO$_4$ tetrahedra. The onset of disorder can be characterised through evaluation of the distribution of Si–Si–Si angles. The distribution function is shown in Figure 7. Two of the peaks in the distribution function broaden and coalesce. The angles of the midpoints of these peaks and their widths are also shown in Figure 7. Two features are apparent. First, the distribution of angles clearly shows the

![Fig. 5. Comparison of the temperature dependence of the mean instantaneous Si–O bond length in quartz as obtained from the analysis of the $D(r)$ data with the distances between the mean Si and O positions obtained from Rietveld structure analysis (after Tucker et al., 2000, 2001d). The latter quantity is seen to decrease on heating due to the neglect of correlated motions in the analysis, whereas the true bond length obtained from the total scattering measurement has a small but positive coefficient of thermal expansion.](image-url)
Fig. 6. Atomic configurations of quartz obtained from RMC analysis of neutron total scattering data (after Tucker et al., 2000, 2001d). The configurations are shown as linked tetrahedra representing the SiO$_4$ groups. The three configurations represent a highly ordered low-temperature state, a partially disordered state at a temperature just below that of the phase transition, and an orientationally disordered state representative of the high-temperature phase.
effects of the phase transition. Second, the widths of the peaks become comparable in size to the separation of peaks at low temperatures. The picture that emerges from this and additional analysis is that the changes in the structure associated with the phase transition are on top of a large “background” of orientational disorder that is established on heating up to the phase transition.

Our hypothesis is that the phase transition facilitates the generation of a large number of RUMs in the high-temperature phase. Calculations of the RUM spectrum of

Fig. 7. Si–Si–Si angle distribution functions for quartz over a wide range of temperatures obtained from the RMC analysis (after Tucker et al., 2000, 2001d). The key point from these results is that the function has a single peak for the disordered β-phase but a central peak and two side peaks for the ordered phase. This result shows that there is a structural difference between the two phases even over the length scale of three linked SiO₄ tetrahedra. The important point is that a pair of peaks in the low-temperature phase coalesces on heating into the high-temperature phase. The temperature dependence of the midpoints of these peaks together with their widths are shown on the right-hand side.
quartz have shown that there are many more RUMs in the high-temperature phase than
in the low-temperature phase (Vallade et al., 1992; Hammonds et al., 1996), as in
cristobalite. This idea is supported by earlier single-crystal inelastic neutron scattering
measurements on quartz, and by inelastic neutron scattering experiments on powdered
samples of other silica phases.

We have recently developed a method for the real-space analysis of RUM motions
based on the mathematical technique of geometric algebra (Wells & Dove, 2002; Wells
et al., 2002). This tool allows us to easily quantify the distortions (local fluctuations) of
structures in terms of RUM motions and motions involving deformations of the SiO₄
tetrahedra. Analysis of the RMC configurations of quartz has allowed us to determine the
fractional component of the atomic displacements that can be associated with RUM
motions. One set of results is shown in Figure 8. This shows the temperature dependence
of the total (i.e. summed over all atoms) mean-squared change in atomic coordinates
from one configuration to another, with the changes being decomposed into motions
involving RUMs and those involving O–Si–O bond-bending and Si–O bond stretching
distortions of the SiO₄ tetrahedra. It is clear from Figure 8 that the dynamic fluctuations
of the structure in the high-temperature phase are primarily due to RUMs, the number of
which increases on heating up to the phase transition. The change in the RUM spectrum
associated with the phase transition was earlier predicted from calculations performed

![Image of a graph showing analysis of the fluctuations of atomic positions between different RMC configurations of quartz. The fluctuations are quantified by the mean square difference in atomic positions between different configurations, scaled by the number of atoms. The top plots show the total mismatch between two configurations, the second plot down shows the residual after subtracting the rigid unit mode motions, and this is divided into the effects of bond-bending and bond-stretching within individual tetrahedra in the lower two plots. The effect of the phase transition is clearly seen in the change of slope of the overall curve (Wells & Dove, 2002; Wells et al., 2002).](image-url)
with the RUM model, and anticipated by the inelastic neutron scattering measurements of Boysen et al. (1980), which showed how some of the modes that are RUMs have a significant increase in their frequencies on cooling below the phase transition. Figure 9 shows the data for the RUM component scaled by temperature, which should scale as an average inverse-squared frequency. If the idea is that the fluctuations are dominated by RUMs in the high-temperature phase, with the number of RUMs decreasing in the low-temperature phase, we would expect that the average inverse-squared frequency should decrease on cooling below the phase transition, which is exactly what is seen in Figure 9. Moreover, changes in the average inverse-squared frequency can be fitted against the order parameter for the phase transition (which, for simplicity, is treated as having the dependence of a tricritical phase transition, Carpenter et al., 1998). We also remark that, following the RMC work on cristobalite discussed above, the calculated three-dimensional patterns of diffuse scattering from quartz from the RMC configurations are consistent with the predictions of the RUM model and with experimental data (Tucker et al., 2001d).

Other polyhedral systems

Similar studies can be performed on other materials containing structural polyhedra. We now briefly describe studies of the two ceramics ZrW$_2$O$_8$ and ZrP$_2$O$_7$. The first of these materials was recently recognised as having isotropic negative thermal expansion over a wide range of temperatures (Mary et al., 1996; Evans, 1999; Evans et al., 1999). The

Fig. 9. The temperature dependence of the RUM mismatch contribution to the data shown in Figure 8 scaled by temperature. This plot gives the inverse of an effective mean frequency, which is expected to increase on cooling below the phase transition because of the loss of RUM flexibility (Wells et al., 2002).
second material has a small positive coefficient of thermal expansion, but is a member of a family of structures of which some members have negative thermal expansion (Evans, 1999). The structures of \( \text{ZrW}_2\text{O}_8 \) and \( \text{ZrP}_2\text{O}_7 \) are similar to each other in the overall structures of networks of the \( \text{ZrO}_6 \) octahedra and \( \text{WO}_4 \) or \( \text{PO}_4 \) tetrahedra, the significant difference being that pairs of \( \text{WO}_4 \) tetrahedra in \( \text{ZrW}_2\text{O}_8 \) have one non-bridging \( \text{W–O} \) bond in each tetrahedron, which are replaced by a \( \text{P–O–P} \) linkage in \( \text{ZrP}_2\text{O}_7 \). This tightens up the structure of \( \text{ZrP}_2\text{O}_7 \) relative to that of \( \text{ZrW}_2\text{O}_8 \).

The negative thermal expansion in \( \text{ZrW}_2\text{O}_8 \) has been understood in terms of rotational RUMs that cause the structure to be dragged inwards as the vibrational amplitudes increase with increasing temperature (Pryde et al., 1996; Heine et al., 1999). The large rotational motions are seen in the RMC configuration of \( \text{ZrW}_2\text{O}_8 \) shown in Figure 10.

Fig. 10. RMC configuration of \( \text{ZrW}_2\text{O}_8 \) showing orientational disorder of the \( \text{ZrO}_6 \) octahedra and \( \text{WO}_4 \) tetrahedra. The projection is viewed down [001].
The situation with ZrP$_2$O$_7$ is slightly more complicated. In this case the main driving force for atomic motion is similar to that in $\beta$-cristobalite, namely that the P–O–P linkage in the average structure is linear, whereas it might be expected that this linkage will subtend an angle closer to 145°. The bent linkages can be clearly seen in the RMC configuration shown in Figure 11. Analysis of the bond-angle distribution is shown in Figure 12; this shows that there is a wide spread of angles within the P–O–P linkages, with the most probable angle being close to 145°.

Molecular systems

The examples of the two crystalline phases of silica are instructive because they give new insights about the relationship between short-range and long-range order when heating through a displacive phase transition. RMC with neutron total scattering is similarly useful in the study of orientationally disordered molecular materials and their phase transitions. Here we cite an example from a study of the orientational

Fig. 11. RMC configuration of ZrP$_2$O$_7$, showing bonds rather than atoms or polyhedra in order to highlight the bent P–O–P linkages. The P atoms are those on tetrahedral sites, and the Zr atoms are those on octahedral sites, as indicated by the number of bonds coming out of each site. All the O atoms are located between two tetrahedral sites or between an octahedral and tetrahedral site. Most of the P–O–P linkages are bent; the straighter bonds appear straight because of the viewing direction.
order–disorder phase transition in NaNO₃. The primary reason for performing this study is as an analogy to calcite. It is believed that for temperatures above 1260 K the carbonate molecular ions in calcite become orientationally disordered, ostensibly through rotations of the carbonate groups about their 3-fold axes in order to be consistent with a change in site symmetry from 3 to 3̅ (Dove & Powell, 1989; Dove et al., 1992; Harris et al., 1998a), but in fact molecular dynamics simulations have shown that there is tumbling of the molecular group about all axes. This effect can be seen in animations produced from the molecular dynamics simulations, posted as mpeg files on http://www.esc.cam.ac.uk/movies. Unfortunately calcite decomposes (CaCO₃ → CaO + CO₂) on heating up to the phase transition, so it has not been possible to obtain structural data for temperatures above the phase transition (the details of the phase transition have been inferred from changes in structure on heating up to the phase transition). On the other hand, NaNO₃ undergoes a similar type of phase transition, and some insights into what happens at the phase transition in calcite may be obtained from studies of the phase transition in NaNO₃ (Schmahl & Salje, 1989; Swainson et al., 1997; Harris et al., 1998b). We performed a series of neutron total scattering measurements on NaNO₃ at various temperatures, and configurations for two temperatures are shown in Figure 13. These clearly show tumbling about all axes of the molecular nitrate group, exactly as predicted in the molecular dynamics simulations of calcite.

![Fig. 12. P–O–P angle distribution function from the RMC configuration of ZrP₂O₇.](image-url)
Other types of RMC modelling

RMC is primarily used in conjunction with total scattering data, but in principle other types of data can be used. One type of data comes from NMR measurements. For example, $^{29}\text{Si}$ magic-angle spinning NMR can provide information about how many Si cations have $n$ neighbouring Al cations, where $n$ varies from 0–4. Using just these few data, it is possible to use inverse Monte Carlo methods to generate topological configurations of Al and Si cations over a network of tetrahedral sites. This approach was initially developed because of difficulties in obtaining the number of Al–O–Al linkages directly from the experimental data. It was found that small errors in the NMR data could produce rather large uncertainties in the calculated values of the number of Al–O–Al linkages, in some cases even producing negative numbers.

One of us (Dove & Heine, 1996; Dove, 1997; see also Peterson, 1999) has implemented the inverse modelling approach to study Al/Si ordering in aluminosilicates using the NMR data. The example of Al/Si ordering in cordierite, $\text{Mg}_2\text{Al}_4\text{Si}_5\text{O}_{18}$, provides a good illustrative example. In the high-symmetry disordered phase, there are two distinct tetrahedral sites. A Si cation in each of these sites can have one of five distinct environments, which are defined by the number of neighbouring Al cations (which, as noted above, can vary between 0–4). The NMR spectra will show a distinct peak for each of these different environments, with an intensity that is proportional to the number of Si cations with that particular environment. There are therefore up to five peaks associated with each tetrahedral site, and there are two non-overlapping groups of peaks associated with the two distinct tetrahedral sites. As samples order, the relative strengths of these peaks change, particularly as the number of distinct environments in the fully ordered structure is only a small subset of the total number possible. In the original NMR data published by Putnis and coworkers (Putnis $et$ $al.$, 1985, 1987; Putnis & Angel, 1985) there are only four peaks associated with each tetrahedral site, the missing peak in each case corresponding to the Si cations having no Al neighbours. It will be noted below that this point is significant.

Fig. 13. Atomic configuration of NaNO$_3$ at two temperatures, showing the loss of orientational order of the molecular NO$_3$ groups on heating above the phase transition, and the three-dimensional tumbling motions of the NO$_3$ groups.
NMR data were obtained for samples annealed at three fixed temperatures for different lengths of time, with ordering growing with longer annealing times. The inverse Monte Carlo method was set up with a topological arrangement of tetrahedral sites, with either an initially fully ordered or random distribution of Al and Si cations. The positions of different cations were swapped at random using a standard Monte Carlo algorithm based on comparing the calculated NMR spectra (which simply reflects the number of Si cations in each environment) with the experimental data. In all cases the final configurations gave good agreement with the experimental data, but for the disordered states (low annealing times) the calculated spectra showed a small but significant (and consistent) strength for one of the missing peaks. Subsequent NMR experiments have now shown that this peak is indeed not zero. However, assigning a zero value to this peak, and then renormalising the overall spectra to give the required number of tetrahedral sites, gave errors that fed through into the final calculation of the number of Al–O–Al linkages, giving for one pair of tetrahedral sites a negative number. On the other hand, the inverse Monte Carlo method, being based on actual configurations of cations, ensures that the number of linkages must be positive at all times. In this way, the errors associated with a wrong measurement of a peak strength feed into the overall result in a more even manner. The time dependence of the number of Al–O–Al linkages for one suite of samples calculated from the RMC configurations is shown in Figure 14, where it is compared with the number obtained by direct analysis of the raw data. The latter quantity is artificially low because it has an unphysical negative contribution that arises from the small errors in the data described above.

**Fig. 14.** Calculated numbers of Al–O–Al linkages in cordierite for different annealing times. The filled circles were obtained from an RMC analysis of the NMR data, and the open circles were obtained from direct analysis of the raw data (Dove & Heine, 1996; Dove, 1997).
Conclusions

The objective of this article, coming as it does within a collection of articles of other types of modelling, has been to show that there is considerable benefit in using data-based simulation methods. Such methods form a useful complement to the more traditional modelling tools, which are instead based on energy functions (empirical or quantum mechanical), and which are compared with experimental data at the conclusion of the simulation. We have shown how inverse modelling techniques can be used with diffraction or NMR data.

Appendix

In this appendix we derive the equation for the scattering of radiation from an isotropic material. This means that we assume that any interatomic vector \( r \) is found for all orientations, which in turn means that we need to average over all relative orientations of \( r \) and \( Q \). We write \( r_{jk} = |r_j - r_k| \), and \( Q = |Q| \), and calculate the orientational average for one vector as

\[
\langle \exp(iQ \cdot (r_j - r_k)) \rangle = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta \exp(iQ r_{jk} \cos \theta) = \frac{1}{2} \int_{-1}^1 \exp(iQ r_{jk} x) dx = \frac{\sin(Q r_{jk})}{Q r_{jk}}.
\]

Thus for all atoms we obtain

\[
S(Q) = \sum_{jk} \overline{b}_j \overline{b}_k \sin(Q r_{jk}) / Q r_{jk} = \sum_j \overline{b}_j^2 + \sum_{j<k} \overline{b}_j \overline{b}_k \sin(Q r_{jk}) / Q r_{jk} ,
\]

where we separate the terms involving the same atoms (the self terms) and those involving interference between different atoms.

We can express the equation using pair distribution functions rather than perform a summation over all atom pairs. We define \( g_{mn}(r) \) as the probability of finding a pair of atoms of types \( m \) and \( n \) with separation between \( r \) and \( r + dr \). This function will have peaks corresponding to specific sets of interatomic distances. For example, in a material containing SiO\(_4\) tetrahedra there will be a peak corresponding to the Si–O bond at \( \sim 1.6 \) Å and a peak corresponding to the O–O bond at \( \sim 2.3 \) Å. \( g(r) \) will be zero for all \( r \) below the shortest interatomic distance, and will tend to a value of 1 at large \( r \). Thus we can rewrite \( S(Q) \) as
\[ S(Q) = \sum_{m} c_m \bar{b}_m^2 + i(Q) + S_0, \]
\[ i(Q) = \rho_0 \sum_{m,n} c_m \bar{b}_m \bar{b}_n \int_{0}^{\infty} 4\pi r^2 (g_{mn}(r) - 1) \frac{\sin Qr}{Qr} dr, \]

where \( c_m \) and \( c_n \) are the proportions of atoms of type \( m \) and \( n \) respectively, and \( \rho_0 \) is the number density. \( S_0 \) is determined by the average density, and gives scattering only in the experimentally inaccessible limit \( Q \rightarrow 0 \).

Acknowledgements

We are grateful to EPSRC for financial support.

References


