Joint structure refinement of x-ray and neutron diffraction data on disordered materials: application to liquid water

2007 J. Phys.: Condens. Matter 19 335206

(http://iopscience.iop.org/0953-8984/19/33/335206)

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **19** (2007) 335206 (18pp) doi:10.1088/0953-8984/19/33/335206

Joint structure refinement of x-ray and neutron diffraction data on disordered materials: application to liquid water

A K Soper

ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK

Received 15 March 2007
Published 4 July 2007
Online at stacks.iop.org/JPhysCM/19/335206

Abstract

X-ray diffraction data on liquids and disordered solids often provide useful complementary structural information to neutron diffraction data. Interpretation of the x-ray diffraction pattern, which is produced by scattering from the atomic electrons rather than from the atomic nuclei as in the case of neutron diffraction, is, however, complicated by the Q-dependent electronic form factors, which cause the x-ray diffraction signal to decline rapidly with increasing Q, where Q is the wave vector change in the diffraction experiment. The problem is particularly important in cases such as water where there is a significant molecular polarization caused by charge transfer within the molecule. This means that the electron form factors applicable to the molecule in the condensed environment often deviate from their free atom values. The technique of empirical potential structure refinement (EPSR) is used here to focus on the problem of forming a single atomistic structural model which is simultaneously consistent with both x-ray and neutron diffraction data. The case of liquid water is treated explicitly. It is found that x-ray data for water do indeed provide a powerful constraint on possible structural models, but that the Q-range of the different x-ray data sets (maximum Q ranges from 10.8 to $\sim$17.0 Å^{-1} for different x-ray experiments), combined with variations between different data sets, means that it is not possible to rigorously define the precise position and height of the first peak in the OO radial distribution function. Equally, it is found that two different neutron datasets on water, although measured to a maximum Q of at least 30 Å^{-1}, give rise to further small uncertainties in the position of the hydrogen bond peaks. One general conclusion from the combined use of neutron and x-ray data is that many of the classical water potentials may have a core which is too repulsive at short distances. This produces too sharp a peak in r-space at too short a distance. A softer core potential is proposed here.

1. Introduction—definitions

In order to exploit the complementarity of neutrons and x-rays correctly for joint structure refinement of neutron and x-ray diffraction data, it is necessary to write down explicitly the equations that relate the structural quantities of interest, primarily the radial distribution functions, to the diffracted intensity. A wide range of definitions are possible [1–3], so it is important to specify which particular definitions will be used here.

1.1. Neutron diffraction

The neutron diffraction differential scattering cross section per atom for a system containing N atoms is described by an expression of the form [4]

$$I^{(n)}(Q) = \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} b_i b_j \exp(-\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_i - \mathbf{r}_j))\right\rangle \tag{1}$$

where b_i is the scattering length of atom i, $\mathbf{Q}$ is the wavevector transfer in the scattering experiment, and $\mathbf{r}_i$ is the position vector of atom i. The angle brackets represent an ensemble average over the system in question, and an average over the spin and isotope states of the nuclei, since the neutron scattering length can be dependent on both the spin and isotope state. Fortunately the spins and isotopes are not normally correlated with the atomic positions, unless the system is magnetic, so this average can be performed independently from the ensemble average. In addition, the ensemble average over pairs of distinct atoms can be replaced by the site–site radial distribution functions, $g_{\alpha\beta}(r)$, [4], since for a homogeneous liquid or glass or random crystalline powder the separation vector between atoms is not correlated with the direction of $\mathbf{Q}$. This leads to an immediate simplification of (1) to the form

$$I^{(n)}(Q) = I^{(n)}_{\mathrm{self}}(Q) + I^{(n)}_{\mathrm{distinct}}(Q) \tag{2}$$

where

$$I^{(n)}_{\mathrm{self}}(Q) = \sum_{\alpha} c_\alpha \langle b_\alpha^2\rangle \tag{3}$$

$$I^{(n)}_{\mathrm{distinct}}(Q) = \sum_{\alpha,\beta\geqslant\alpha} (2-\delta_{\alpha\beta})\, c_\alpha c_\beta \langle b_\alpha\rangle\langle b_\beta\rangle H_{\alpha\beta}(Q) \tag{4}$$

and where $\delta_{\alpha\beta}$ is the Kronecker δ-function, used here to avoid double counting, $c_\alpha = \frac{N_\alpha}{N}$ is the atomic fraction of atomic species α with N_α the number of atoms of that species, the angular brackets now represent spin and isotope averages, and $H_{\alpha\beta}(Q)$ is the partial structure factor [5] between atom types α and β defined by

$$H_{\alpha\beta}(Q) = 4\pi\rho_{\mathrm{a}}\int_0^\infty r^2(g_{\alpha\beta}(r)-1)\frac{\sin Qr}{Qr}\,\mathrm{d}r \tag{5}$$

with ρ_{a} the atomic number density. It is worth pointing out here that assuming corrections for inelasticity effects [6] have been made satisfactorily, then the self-scattering for neutrons, equation (3), is a constant, independent of Q.

For molecular systems it is common to split the distinct term into two terms, one intra-molecular, the other inter-molecular. The former consists entirely of pairs of atoms on the same molecule, while the latter consists of pairs of atoms on different molecules:

$$I^{(n)}_{\mathrm{distinct}}(Q) = I^{(n)}_{\mathrm{intra}}(Q) + I^{(n)}_{\mathrm{inter}}(Q) \tag{6}$$

where for harmonic or quasi-harmonic forces

$$I^{(n)}_{\rm intra}(Q) = \frac{1}{N}\sum_m \sum_{i\neq j} \langle b^{(m)}_i b^{(m)}_j\rangle \frac{\sin Qr^{(m)}_{ij}}{Qr^{(m)}_{ij}} \exp\left(-\frac{[\sigma^{(m)}_{ij}Q]^2}{2}\right), \tag{7}$$

and $I^{(n)}_{\rm inter}(Q)$ has exactly the same form as $I^{(n)}_{\rm distinct}(Q)$, equation (4) above, except that it is now understood that the $g_{\alpha\beta}(r)$ functions used in (5) consist only of correlations between atoms on different molecules. In (7) $b^{(m)}_i$ is the scattering length of the ith atom and $r^{(m)}_{ij}$, $\sigma^{(m)}_{ij}$ are the average separation and root-mean-square (rms) deviation of the (i, j) atom pair respectively, on the mth molecule.

It is possible to perform a direct Fourier inversion of $I^{(n)}_{\rm inter}(Q)$ to give the appropriately weighted sum of site–site radial distribution functions (RDFs):

$$\frac{1}{2\pi^2\rho_{\rm a}}\int Q^2 I^{(n)}_{\rm inter}(Q)\frac{\sin Qr}{Qr}\,{\rm d}r = \sum_{\alpha,\beta}(2-\delta_{\alpha\beta})\,c_\alpha c_\beta \langle b_\alpha\rangle\langle b_\beta\rangle(g_{\alpha\beta}(r)-1). \tag{8}$$

Note that the expressions (1)–(8) can be defined *per molecule*, instead of *per atom*, by replacing the number of atoms, N, with the number of molecules in the system, N_m, where appropriate (in (1) and (7) and in the definition of c_α), and by replacing the atomic number density with the molecular number density, ρ_m, in (5) and (8). The subtle but rather important distinction between the per-atom and per-molecule definitions (when it comes to trying to model a diffraction pattern) frequently goes unstated in the literature.

1.2. X-ray diffraction

For x-ray diffraction the expressions for the differential scattering cross section are entirely analogous to (1)–(7), within the standard approximation that the electron form factor, $f(Q)$, for each atom is spherically isotropic. This electron form factor replaces the neutron scattering length, b, in all of equations (1)–(7), but otherwise the expressions are the same. Of course there is no average over spin and isotope states in the case of x-rays. For example, the expressions for the self, intra-molecular and inter-molecular interference functions become, respectively,

$$I^{(X)}_{\rm self}(Q) = \sum_\alpha c_\alpha f^2_\alpha(Q), \tag{9}$$

$$I^{(X)}_{\rm intra}(Q) = \frac{1}{N}\sum_m \sum_{i\neq j} f^{(m)}_i(Q) f^{(m)}_j(Q)\frac{\sin Qr^{(m)}_{ij}}{Qr^{(m)}_{ij}} \exp\left(-\frac{[\sigma^{(m)}_{ij}Q]^2}{2}\right), \tag{10}$$

$$I^{(X)}_{\rm inter}(Q) = \sum_{\alpha,\beta}(2-\delta_{\alpha\beta})\,c_\alpha c_\beta f_\alpha(Q) f_\beta(Q) H_{\alpha\beta}(Q). \tag{11}$$

With this representation the molecular form factor is sometimes written as

$$I^{(X)}_{\rm mol}(Q) = I^{(X)}_{\rm self}(Q) + I^{(X)}_{\rm intra}(Q). \tag{12}$$

1.2.1. X-ray data normalization. The spherical electron form factors will be referred to here as independent atom form factors (IAFFs), and they are listed in various places [7]. Obviously the IAFFs are not spin or isotope dependent, but they are dependent on the magnitude of Q so that a direct Fourier inversion of the interference scattering to give the weighted sum of site–site radial distribution functions as given in (8) for neutrons is not possible. However, it is common to make this representation approximately, by replacing the x-ray differential cross section by

the normalized or reduced distinct differential cross section, which removes the Q dependence of the atom form factors approximately [1], but not exactly:

$$F^{(X)}(Q) = \frac{I^{(X)}(Q) - I^{(X)}_{\rm self}(Q)}{(\sum_\alpha c_\alpha f_\alpha(Q))^2}. \tag{13}$$

Since this quantity compensates for the rapid decline of the form factors at large Q, it is the one that is normally reported in an x-ray experiment. It is sometimes referred to as the 'pseudo-nuclear' x-ray structure factor [17].

For molecular systems—water is the example that will be used here—a different normalization to (13) is sometimes used, namely the data can be normalized to the molecular form factor [8, 9]:

$$F^{(X)}(Q) = \frac{I^{(X)}(Q) - I^{(X)}_{\rm mol}(Q)}{I^{(X)}_{\rm mol}(Q)}, \tag{14}$$

where the denominator can be calculated either from expressions such as (9) and (10) above, or via an *ab initio* calculation, as done in [10]. The aim in doing this is to attempt to remove any dependence within $F^{(X)}(Q)$ on the internal structure of the molecule, leaving only the molecule–molecule structure factor. Formally, however, the justification for this normalization is obscure [11], and it only really works for systems (like water) where one of the form factors (oxygen in this case) dominates the others (hydrogen in the case of water) so that the scattered x-ray intensity is only very weakly dependent on the orientational correlations between molecules [8], even in situations such as water where the orientational correlations are likely to be strong.

It should also be noted that all of the expressions (9)–(14) depend on whether per-atom or per-molecule definitions are in place. Normally if the data are normalized to the molecular form factor, (14), then it should be assumed that the per-molecule definition is in use, although this is not always stated explicitly.

The normalizations (13) or (14), although almost universally used in the presentation of x-ray data, overlook a fundamental property of the scattering process that seems not to be widely recognized. The single atom scattering, $I^{(X)}_{\rm self}(Q)$, represents the scattered intensity as a function of Q that would occur if all atomic correlations within the system were set to zero, that is the uniform atom background. In the parlance of reciprocal space, the single atom scattering represents the density of reciprocal lattice points as a function of Q for an x-ray experiment. It is about this curve that the structure factor for the material in question must oscillate. Moreover, since the total scattered intensity can nowhere be negative, we have the rule that $I^{(X)}_{\rm distinct}(Q) \geqslant -I^{(X)}_{\rm self}(Q)$ for all Q. Therefore it could be argued that the correct normalization for x-ray data is neither equation (13) nor (14), but in fact the correct normalization should be to the single atom scattering,

$$F^{(X)}(Q) = \frac{I^{(X)}(Q) - I^{(X)}_{\rm self}(Q)}{I^{(X)}_{\rm self}(Q)}, \tag{15}$$

since this represents the level about which the structural correlations oscillate. This is the definition that comes closest to forming a nuclear-like structure factor from x-ray diffraction data. We can say this because the interference intensity $F^{(X)}(Q) > -1$ for all Q values, which is not true if normalization (13) is used, while normalization (14) makes assumptions about the extent of orientational correlations between molecules.

This particular issue does not arise for neutron diffraction data since the scattered intensities by definition oscillate about a constant line, independent of Q, whatever method is used to normalize the data. Hence for neutrons the choice of method of normalization is not

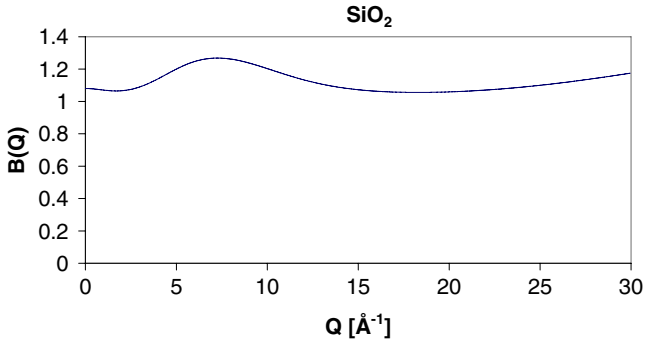

Figure 1. The ratio of single atom scattering to square of the mean form factor, $B(Q) = \frac{\sum_\alpha c_\alpha f_\alpha^2(Q)}{(\sum_\alpha c_\alpha f_\alpha(Q))^2}$, for SiO_2 using the independent atom form factors (IAFFs).

important provided we know what it is, since the net effect of different normalizations is simply an overall factor outside the differential cross section which is independent of Q.

All the x-ray diffraction data quoted in this paper have been renormalized according to (15), and not according to (13) or (14) as in previous work. This is because the data are to be used directly in the empirical potential structure refinement (EPSR) process to be described, and this requires the minimum of spurious Q-dependent structure to be supplied in the scattering data. To see the effect of using (15) instead of (13) in the normalization of x-ray diffraction data, figure 1 shows the ratio of $B(Q) = \frac{I_{\rm self}^{(X)}(Q)}{(\sum_\alpha c_\alpha f_\alpha(Q))^2}$ for SiO_2, using the independent atom form factors. It can be seen that this ratio is certainly not flat with Q, and so could generate spurious structure in r-space in any Fourier transforms that might be performed on the diffraction data.

1.2.2. Modified atomic form factors (MAFFs). Another feature of x-ray diffraction, which has been widely discussed in the case of water [8, 9, 12], is the effect of covalent bonding on the electron form factors. For heavy elements this is a minor effect since bonding affects only the valence electrons. For lighter atoms such as carbon and oxygen, however, a significant fraction of the total electron complement can be involved in bonding interactions, which may have the effect of distorting the free atom form factors at low Q. To circumvent this problem, various approximations have been proposed, [8, 12, 13]: the one adopted here is that proposed by Head-Gordon *et al* [9], which is to replace the IAFF for atom α with a modified atomic form factor (MAFF):

$$f'_\alpha(Q) = \left[1 - \frac{q_\alpha}{f_\alpha(0)} \exp(-Q^2/2\delta_\alpha^2)\right] f_\alpha(Q), \tag{16}$$

where q_α is the fractional electron charge shifted onto or off the ith atom, and δ_α is a width function. Both q_α and δ_α have to be determined by comparison with calculations based on the free molecule. Obviously the values of q_α have to preserve electrical neutrality so that $\sum_\alpha q_\alpha = 0$. For a free water molecule this representation was able to give an excellent account [9] of the calculated water molecule form factor [10]. In the work presented here the charges that are used to define the MAFFs for water are the same as those used in the empirical potential structure refinement simulations, making the MAFFs and the intermolecular potential energy function mutually self-consistent.

1.3. Comparison of x-ray and neutron experiments

A fundamental aspect of any diffraction experiment on a liquid or disordered material is the need to put the diffraction data on an absolute scale of cross section. This absolute scale is imperative if we are to say anything quantitative about the structure of the material in question, such as the coordination number. It is also essential when we wish to compare a structural model of the material with the diffraction data. For the present purposes the action of putting the diffraction data on an absolute cross-section scale will be referred to as 'calibration' and should be distinguished from the 'normalization' referred to in the previous section concerning the removal of the Q dependence of the electron form factors, even though the actions of calibration and normalization are frequently performed in a single operation.

For neutrons the calibration is normally achieved via a standard scattering sample, usually vanadium, which by a good fortune of nature has an almost zero coherent scattering length [14]. This means that the scattering pattern from vanadium is dominated by the self-scattering. There are only very weak Bragg peaks from the structure of the vanadium which are weakly visible in the diffraction pattern and which can be readily compensated for. The data are normalized to the vanadium sample and since there is no Q dependence to the neutron scattering lengths, the vanadium calibration serves to eliminate a large fraction of the uncertainties that might arise from the detector efficiency correction, incident spectrum uncertainties, and different detector solid angles at different scattering angles. It does not of course eliminate standard corrections such as attenuation, multiple scattering and inelasticity effects, but at least the data are on an absolute scale by measurement rather than by calculation [15].

For x-rays there is no such incoherent sample that can be employed. Instead the data analysis must proceed with a great deal of care and attention to all the details [1, 2, 8, 9, 16]. At the end the data have to be calibrated either by comparison with the single atom scattering [1, 2, 17] or by comparison with the molecular form factor [8, 9, 17]. Usually the processes of normalization and calibration are achieved in a single refinement operation [2, 17], using a formula of the form

$$F^{(X)}_{\text{distinct}}(Q) = \frac{(cR^{(X)}(Q) - I^{(X)}_{\text{calib}}(Q))}{I^{(X)}_{\text{norm}}(Q)}, \tag{17}$$

where c is a calibration constant to be determined by minimizing the standard deviation of $F^{(X)}_{\text{distinct}}(Q)$, $R^{(X)}(Q)$ are the corrected but uncalibrated x-ray counts, $I^{(X)}_{\text{calib}}(Q)$ is either the single atom scattering (9) or the molecular form factor (12), plus the Compton scattering if present, and $I^{(X)}_{\text{norm}}(Q)$ is the denominator of one of equations (13)–(15). Obviously there will in practice be many variations in detail on this basic procedure [1].

2. Incorporating both x-ray and neutron diffraction data into structural modelling

The basic idea for structure refinement that will be adopted here follows the empirical potential structure refinement (EPSR) scheme [18, 19]. This method has been used extensively to study the structure of water, molecular liquids and aqueous systems, and so is relevant to the discussion here. References [18, 19] and the references therein give much of the detail of the method, which will not be repeated here. Suffice to say that a standard Monte Carlo simulation of the system in question is performed using a potential typically obtained from the literature. This potential, which is normally a site–site potential, is then perturbed by the diffraction data, taking account of the weights outside each partial structure factor according to (4). Hitherto the method has been applied almost exclusively to neutron diffraction data, but recently some work in which neutron and x-ray data were jointly refined together was

Table 1. Lennard-Jones and charge parameters used in the EPSR models described in this article.

Atom	ε (kJ mol^{-1})	σ (Å)	q (e)	Dipole moment μ (D)
O	0.3	3.2	−1.0	2.89
H	0.0	0.0	+0.5	—

Table 2. Parameters used to define the exponential repulsive core potential, equation (18).

Atom pair	$r_{\alpha\beta}$ (Å)	γ (Å)	$C_{\alpha\beta}$ [a] (kJ mol^{-1})
O–O	2.00	0.3	0.00 ± 0.00
O–H	1.35	0.3	2.72 ± 0.35
H–H	1.70	0.3	0.26 ± 0.08

[a] This shows the average values and rms standard deviation for the six simulations reported here.

reported [20]. This work performed a joint refinement of structural models for water against both x-ray and neutron diffraction data using EPSR. In all the cases reported in this paper the assumed molecular geometry has an O–H bond length of 0.984 Å with rms deviation 0.074 Å, and the H–H distance is 1.55 Å with rms deviation 0.11 Å. The simulations are performed with 1000 molecules, with the box sizes set appropriately to give an atomic number density of 0.1002 atoms Å^{-3}. The temperature is set to 300 K in all the simulations. The reference potential parameters were modified from the 'SPC/E' potential for water [21] and for completeness they are listed in table 1. The reference potential is given by the standard Lennard-Jones plus Coulomb term, plus an additional exponential repulsive term which is designed to soften the Lennard-Jones repulsive core:

$$U_{\alpha\beta}^{(\mathrm{ref})}(r) = 4\varepsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r}\right)^{6}\right] + \frac{q_\alpha q_\beta}{4\pi\varepsilon_0 r} + C_{\alpha\beta}\exp\left(\frac{1}{\gamma}(r_{\alpha\beta} - r)\right), \tag{18}$$

where $C_{\alpha\beta}$ is adjusted iteratively so that there are no pairs of atoms of type α, β at separations of $r < r_{\alpha\beta}$. The hardness of this repulsive term is controlled by the value of γ, which was set to 0.3 Å for all the simulations shown here. The values of $C_{\alpha\beta}$ and $r_{\alpha\beta}$ used in the simulations are given in table 2.

The data are introduced to the structure refinement simulation via the so-called 'weights' equations such as (4) or (11). These can be written in general form [19]:

$$D_i(Q) = \sum_{j=1,N} w_{ij} H_j(Q), \tag{19}$$

where $D_i(Q)$ represents the ith set of data, the index j runs over the N partial structure factors in the system, and the weights matrix, w_{ij}, is given for neutrons by $w_{ij} = (2 - \delta_{\alpha\beta})\, c_\alpha c_\beta \langle b_\alpha^{(i)}\rangle\langle b_\beta^{(i)}\rangle$, where j runs over all the N pairs of α, β values, and for x-rays by $w_{ij} = \frac{(2-\delta_{\alpha\beta})\, c_\alpha c_\beta f_\alpha(Q) f_\beta(Q)}{I_{\mathrm{norm}}^{(X)}(Q)}$. A set of such weights is defined for each i of M sets of measured diffraction data. In general there are fewer sets of data than partial structure factors, $M < N$, so inversion of the weights matrix is indeterminate. However, as pointed out in [19], even when $M \geqslant N$ one does not necessarily want to rely totally on the data since they will often contain a systematic error which has the possibility of introducing artefacts into the simulated structure. Thus it was proposed to use the simulation itself as a form of additional 'data' so that the weights matrix can be inverted under all conditions. This was done by means of a 'feedback' factor, f, such that for the data a modified set of weights is defined, namely $w'_{ij} = f w_{ij}$, for $1 \leqslant i \leqslant M$, and $w'_{ij} = (1 - f)\delta_{(i-M),j}$; for $M < i \leqslant (M + N)$, giving rise to an overdetermined weights matrix. The value of f lies in the range $0 < f < 1$.

The inverse of this matrix, w_{ji}^{-1}, is found by least squares by requiring that the $(M+N)\times(M+N)$ matrix formed from $P_{ii'} = (\sum_{j=1,N} w'_{ij} w_{ji'}^{-1} - \delta_{ii'})$ has a minimum norm. Note that with the definitions used here and provided $f < 1$, an inverse of this modified weights matrix can always be found. As f approaches unity the emphasis on the data increases, along with the increasing risk of artefacts being introduced into the reconstructed structure. As f is made smaller, the risk of artefacts decreases, as does, however, also the ability to fit the data. Hence a value which minimizes the artefacts, but gives best fit to the data, has to be chosen, usually by inspection. For all the cases discussed in this paper the feedback factor was set to 0.8.

To include x-ray data in the EPSR method, it is necessary to calculate the inversion of the x-ray weights matrix alongside the neutron weights matrix. However, because the x-ray weights are Q dependent, the inversion matrix will also be Q dependent, there being a different inversion for each value of Q. Since the inversion matrix is used to calculate the perturbations to the site–site potential function in real space, the Q dependence of the inversion matrix will translate to an awkward convolution in r-space. Therefore in the work reported here and in [20] the inversion of the weights matrix at $Q = 0$ is used to calculate the perturbation to the potential in r-space. Provided the x-ray data are supplied in the form of the pseudo-nuclear structure factors, using the single atom normalization (15), then the approximation in doing this will not be too severe, because the adopted normalization method will remove the main form factor Q dependence in the x-ray intensity. The forward calculation from RDF to x-ray intensity in EPSR is of course accomplished with the correct Q dependence of the x-ray form factors. Obviously for simulations on other systems the present protocol may not be appropriate, so a study of the use of the $Q = 0$ inversion would be needed.

As discussed in [19], a useful feature of the inverted weights matrix is that it allows us to reconstruct the site–site partial structure factors from the data, even when there are insufficient distinct datasets to do this exactly. This is achieved via the equation

$$H_j^{(E)}(Q) = \sum_{i=1,M} w_{ji}^{-1} D_i(Q) + \sum_{i=M+1,M+N} w_{ji}^{-1} H_{i-M}(Q), \qquad \text{for } 1 \leqslant j \leqslant N. \tag{20}$$

With x-ray data present the inverse weights matrix is calculated at each Q value so that these estimated partial structure factors are still valid.

Note that due to the likely presence of systematic effects in the diffraction data a perfect fit to the data is often difficult to achieve. In that case the best fit is one which approaches the data most closely. Left to itself, the empirical potential in EPSR will grow indefinitely, if there are residual systematic differences between simulation and data which cannot be fitted by any distribution of atoms. To prevent this happening, it is often necessary to impose an upper limit on the amplitude of the empirical potential. This amplitude is defined by the quantity [19]

$$\bar{U} = 4\pi\rho \sum (2 - \delta_{\alpha\beta})\, c_\alpha c_\beta \int r^2 |U_{\alpha\beta}(r)| g_{\alpha\beta}(r)\, \mathrm{d}r. \tag{21}$$

In all the simulations reported here this amplitude was limited to 50 kJ mol^{-1}.

Of course it is always possible that if a given structure refinement is not able to reproduce a given set of diffraction data in its entirety, then that does question the validity of the structural model extracted. The corollary of this is that if even if there are no differences remaining between data and fit, that cannot simply be used to imply the model is necessarily valid, since the model may be fitting systematic effects in the data which have nothing whatever to do with the real structure. Unless we know precisely the magnitude of systematic effects, we have to live with this uncertainty. The best one can do is find a model which fits the data as closely as possible. The most important benefit of EPSR/RMC methods is that they enable us to explore the range of models that may be consistent with a given set of diffraction data. This gives us some idea of how sound our structural conclusions are. We then may exclude some of those

Table 3. Neutron weightings outside the site–site partial structure factors for water (barns/sr/atom) (1 barn = 10^{-28} m^2).

Mole fraction of D_2O in H_2O	O–O	O–H	H–H
0.00	0.0374	−0.0964	0.0621
0.25	0.0374	−0.0293	0.0058
0.50	0.0374	0.0378	0.0096
0.75	0.0374	0.1049	0.0735
1.00	0.0374	0.1721	0.1978

models on the basis of some other information (e.g. bonding constraints, lack of hydrogen bonding, etc).

3. Results: application to liquid water at 25 °C

3.1. Neutron diffraction data

The neutron diffraction data for water presented here were obtained on the small angle neutron diffractometer for amorphous and liquid samples (SANDALS) at the ISIS pulsed neutron source. There are two datasets reported here, namely the data that were originally shown in [20] and which formed the basis for the analyses presented in [22]. The set consisted of pure heavy water, pure light water, and mixture ratios, 25%, 50% and 75% D in H, referred to here as water (old). Subsequently the data for H_2O, D_2O and a 50:50 mixture of H_2O and D_2O were remeasured in a recent experiment on SANDALS, here referred to as water (new). Between the two sets of experiments, the SANDALS detectors were completely refurbished, with the result that they became significantly more stable, and the effects of detector deadtime could be corrected for more reliably. This results in the two sets of measurements being relatively independent from one another. The appropriate neutron weightings outside the HH, OH and OO partial structure factors are shown in table 3 for each of these samples. In every case the samples are contained in a flat plate container made from a 'null' alloy of $Zr_{0.324}Ti_{0.676}$ which has the useful property of being a random alloy and having a zero mean coherent scattering length due to the negative scattering length of Ti compared to Zr. This means that the background scattering from this material is almost featureless, and so after subtraction cannot contribute to the coherent scattering from the sample. The wall thickness was 1 mm in every case, and the sample thickness was also 1 mm, to help reduce the effects of multiple scattering and attenuation in the samples. The data were corrected for attenuation and multiple scattering using standard techniques and put on an absolute scale of differential cross section by comparison with the scattering from a vanadium sample, as described in section 1.3.

Figure 2 shows both data sets, and the corresponding EPSR refinement using the reference potential defined by equation (18) with the parameters as given in tables 2 and 3. The corresponding radial distribution functions (RDFs) are shown in figure 3. It will be seen that (a) the fits to the diffraction data are not perfect: there are small discrepancies which do not go away however much the empirical potential in EPSR is refined; and (b) that the resulting RDFs from the different datasets show overall good agreement with one another, with differences only in detail. In particular, the position of the second peak in the OO RDF appears shifted to small r with the more recent data, and the OH and HH first peaks are in slightly different positions. Closer inspection of the data reveals small differences between corresponding new and old datasets, which may explain the differences seen in r-space. It is important to realize for example that any change in the detector response could affect the extent to which the data are

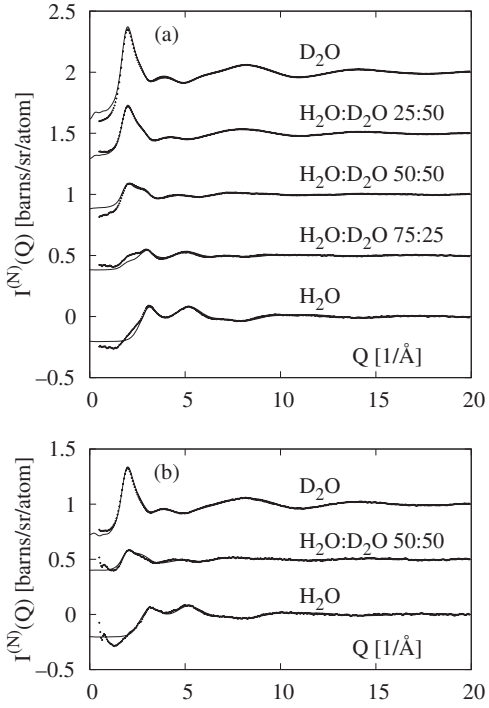

Figure 2. EPSR fits to the old (a) and new (b) neutron diffraction data from liquid water at ambient pressure and 25 °C. For (a) there were five datasets consisting of mole fractions of 0, 0.25, 0.5, 0.75, and 1.0 D_2O in H_2O, from bottom to top respectively. For (b) there were three data sets consisting of mole fractions of 0, 0.5, and 1.0 D_2O in H_2O, from bottom to top respectively. The lines show the EPSR fits to the data which are shown as points. The separate datasets are shifted by 0.5 barns/sr/atom vertically for clarity.

sensitive to inelastic neutron scattering. This sensitivity to external changes in the instrument configuration demonstrates the difficulties inherent in neutron diffraction experiments of the kind described here, that is those containing significant amounts of hydrogen. The OO RDF seems particularly vulnerable in this case because it represents such a small fraction of the total diffraction pattern, as indicated by the weightings in table 3.

Nonetheless, and despite those caveats, the overall agreement between the two simulations is remarkably good and demonstrates the value of performing some form of structure refinement when attempting to derive structural information from a set of diffraction data: the process of structure refinement has helped to eliminate some spurious structure which otherwise might have occurred if the data had been interpreted without reference to a structural model. Of course the analysis performed here assumes that heavy and light water, and mixtures of the two, all have identical structures, which is almost certainly not true. Therefore another cause of the discrepancies between data and fit could also be related to the fact that H_2O and D_2O may not be structurally identical, as shown recently [23]. This will hopefully form the subject of a future study.

3.2. X-ray diffraction data

Following on from the earlier work of Narten and collaborators [8] (referred to here as Narten), there have been a number of recent efforts to measure the structure of water with x-rays. The

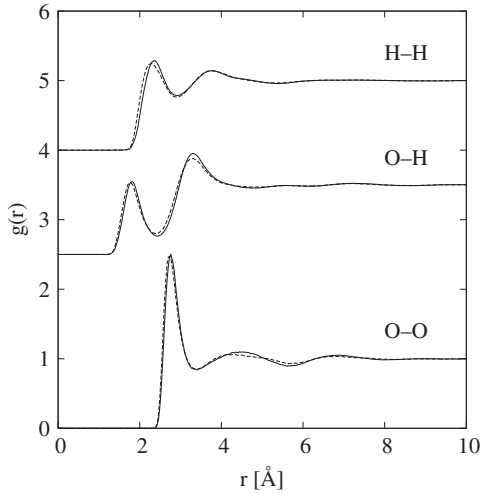

Figure 3. O–O (bottom), O–H (middle) and H–H (top) radial distribution functions for water as estimated from the structure refinements of figure 2. The refinement of the old data is shown as the solid line, while that of the new data is shown as the dashed line.

list used here is not exhaustive, but includes the most recent measurements by Head-Gordon and co-workers [24] (referred to here as PCCP), and a recent measurement of the difference in structure between heavy and light water [23] (referred to here as HASYlab). With the exception of [23] these different datasets were originally calibrated and normalized against the molecular form factor, but it is not readily practical to incorporate x-ray data normalized to the molecular form factor into the EPSR procedure, since the molecular structure can fluctuate in the course of the EPSR simulation. (In the analysis shown in [20] it was assumed that the x-ray data corresponded simply to the OO partial structure factor, which is an approximation as described in section 1.2). Therefore for the present refinements, and following the discussion given in section 1.2.1, the x-ray data supplied by the various authors were renormalized according to equation (17), with $I^{(X)}_{\rm norm}(Q) = \sum_\alpha c_\alpha f^2_\alpha(Q)$, using the MAFFs given by equation (16), the charges being the same as those used in the reference potential, table 2, and with the value of δ set to 2.2 Å^{-1} for both hydrogen and oxygen atoms.

Figure 4 compares the results of these three distinct x-ray experiments after renormalization. It can be seen that although there is general agreement between the three datasets, there are, as with the neutron data, differences in detail, including in the high-Q region. However, in spite of these differences, there does actually seem to be quite good overlap between the different datasets. It was in this high-Q region where it was proposed that the effect of any asymmetry in the first peak of $g_{\rm OO}(r)$ might become apparent [20] in the x-ray data.

In the important Q region 2–3 Å^{-1} the Narten and PCCP datasets overlap quite well, but the HASYlab data are significantly above these other data. Clearly there are discrepancies here that arise from the different methods of data analysis employed. It should be emphasized that the purpose of the HASYlab data, [23], was primarily to look at the *differences* in structure between heavy and light water, rather than the structure itself, so that provided background effects are the same for both heavy and light water, they would cancel when taking the difference. The extent to which the discrepancies make any significant difference to the outcome of the structure refinement process will be discussed below. Overall, the differences seen here are not unexpected: they represent three truly independent measurements on different

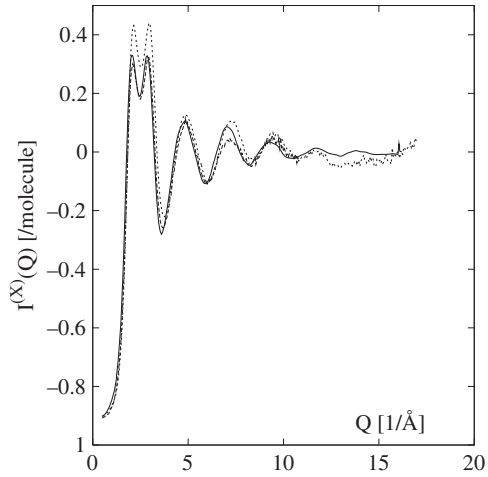

Figure 4. Comparison of the three x-ray diffraction datasets, renormalized to the single atom scattering as described in the text, namely Narten (solid line), PCCP (short dashed line), and HASYlab (dotted line).

x-ray sources, with different resolutions, detector responses, etc. As already seen above,if the neutron data had been measured three times on different sources a similar spread in results might have been observed. However, figure 4 does highlight the difficulty of obtaining accurate diffraction data from disordered systems on an absolute scale of intensity.

3.3. Joint refinement of x-ray and neutron data

A separate joint structure refinement was performed against each x-ray dataset in turn, using the old water neutron data. The fits to the different sets of data are shown in figures 5(a)–(c). Both the Narten and PCCP data appear to be fitted quite well, but there are discrepancies with the HASYlab data that appear to be related to a Q-dependent background on which the data may sit, as already mentioned. All attempts to further improve the fits to the data were unsuccessful. In addition a low-Q rise appears in the fit to the HASYlab data which does not appear in the other datasets, even though they all use the same low-Q cutoff.

Figure 6 shows the radial distribution functions from these three structure refinements. As we saw with the neutron data alone, small differences are introduced by using different datasets, although the overall agreement in terms of the shape of the RDFs between the different datasets is quite good, suggesting that although there are still some uncertainties in the RDFs for water, by using combined neutron and x-ray data it is possible to get a reasonably consistent description of the water structure. Again it should be emphasized that these results are obtained assuming that the structures of D_2O and H_2O are identical, which is almost certainly not true at the level of the differences seen here between the different datasets.

Figures 7 and 8 show the fit and RDFs respectively for the HASYlab x-ray data combined with the new water data. Previous comments about the fit to the HASYlab data apply here as well. Small discrepancies in the RDFs are observed between the fits with the new and old water data, figure 7, although these tend to be smaller than those seen when the neutron data are fitted on their own, figure 3. This implies there is significant added value of fitting both x-ray and neutron data simultaneously, rather than just the neutron data on their own.

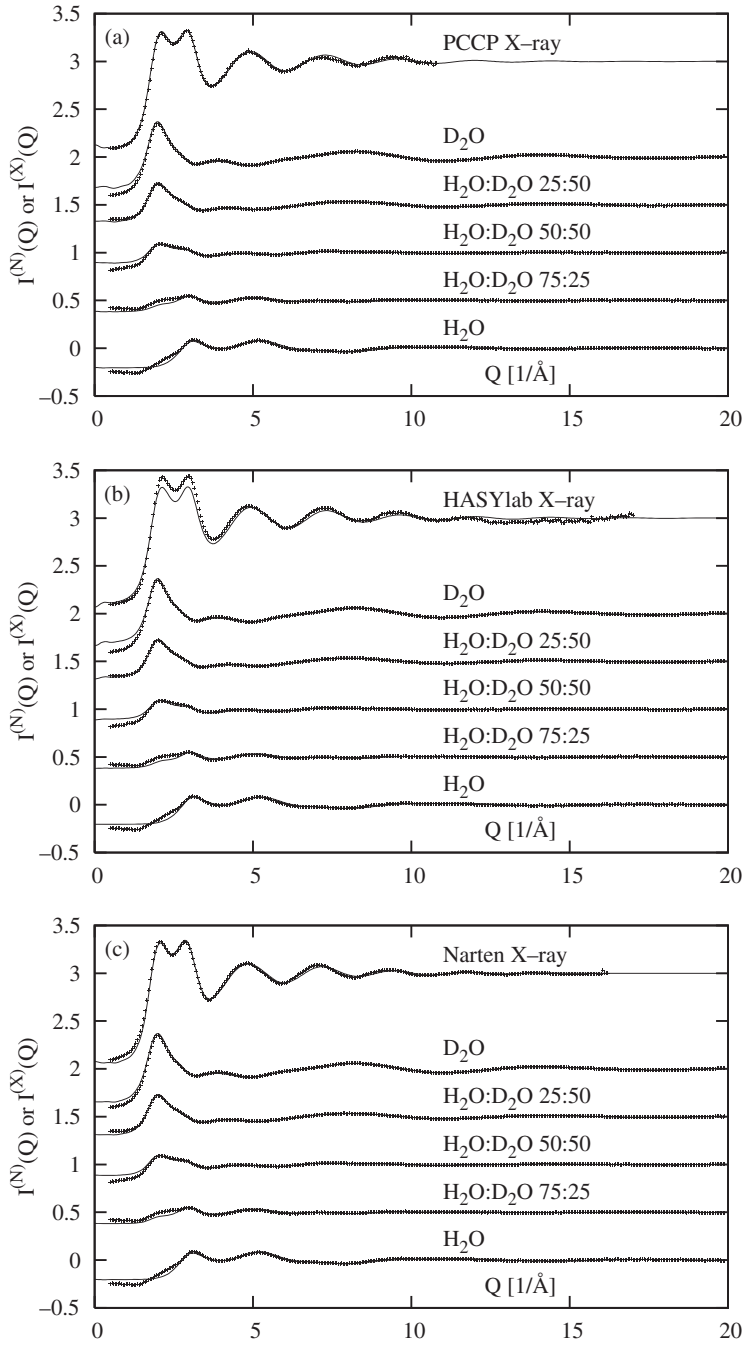

Figure 5. Joint structure refinement of x-ray (figure 4) and old neutron diffraction data on water (figure 2(a)) using the modified atom form factors (MAFFs, $q_O = -1.0e$, $q_H = 0.5e$) to describe the x-ray data. As previously the dots show the data and the lines show the fits to the data. (a) corresponds to the fit to the neutron plus PCCP x-ray data, (b) corresponds to the fit to the neutron plus HASYlab x-ray data, and (c) corresponds to the fit to the neutron plus Narten data. Each neutron curve is shifted successively upwards by 0.5 barns/sr/atom, while the x-ray curves have been shifted upwards by 3.0, for clarity.

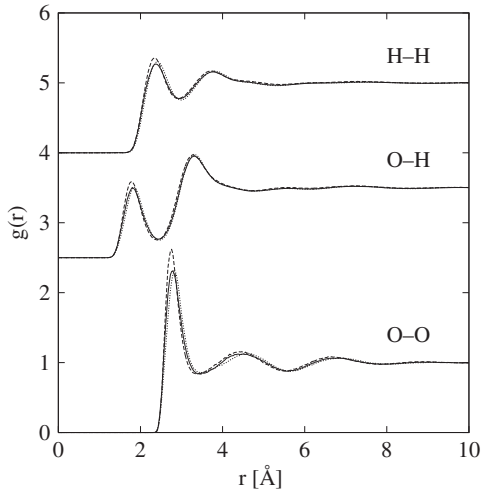

Figure 6. O–O (bottom), O–H (middle) and H–H (top) radial distribution functions for the simulations shown in figure 5. The solid line corresponds to the fit to the PCCP x-ray data (figure 5(a)), the dashed line to the fit to the HASYlab x-ray data (figure 5(b)), and the dotted line to the fit to the Narten x-ray data (figure 5(c)).

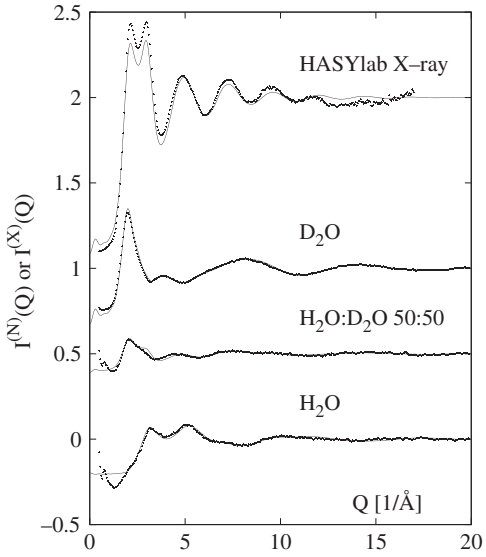

Figure 7. Joint structure refinement of HASYlab x-ray diffraction data (figure 4) and new neutron diffraction data on water (figure 2(b)) using the modified atom form factors (MAFFs, $q_O = -1.0e$, $q_H = 0.5e$) to describe the x-ray data.

4. Discussion

A significant feature of the radial distribution functions shown in figures 3 and 8 is the height of the first peak in $g_{OO}(r)$. This is found to be significantly smaller than traditional estimates based on simple charge potentials for water [21, 9, 22, 24], and is more in line with the original estimates of Narten [8] and from direct Fourier inversion of the neutron diffraction

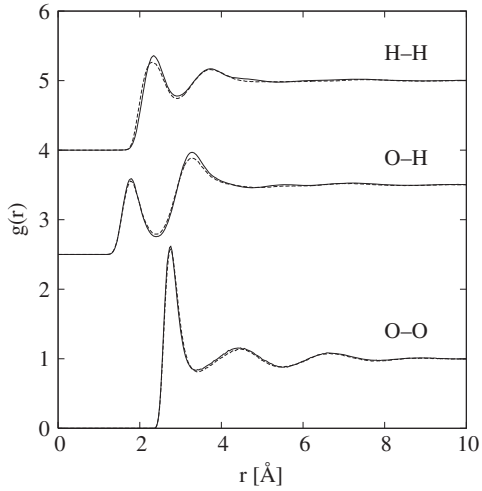

Figure 8. O–O (bottom), O–H (middle) and H–H (top) radial distribution functions derived from the simulations shown in figure 5(b) (solid line, HASYlab x-ray plus old neutron diffraction data) compared with those derived from the simulations shown in figure 7 (dashed line, HASYlab x-ray plus new neutron diffraction data).

data [25]. However, there is more to this discrepancy than simply the peak height. Figure 9(a) compares the simulated OO radial distribution function from all six EPSR simulations with the corresponding $g_{OO}(r)$ function obtained by Fourier inversion of the estimated OO partial structure factor from using equation (20) in each of the different simulations. Table 4 lists the positions and heights of the first peak in this function from the data and from the simulation. In every case the simulated peak is higher and at shorter distance than implied by the data, even though the observed shifts are within what is allowed by the current Fourier resolution. This is true even in the cases of HASYlab and Narten, where the x-ray data extend beyond $Q = 15$ Å^{-1}. Whilst the neutron data by themselves are relatively insensitive to the height and position of this peak, the x-ray data appear to define it more precisely. The conclusion, consistent across all the datasets, is that this peak is lower and at larger distances than is conventionally assumed. Unless there is something fundamentally wrong with the x-ray datasets, which seems unlikely in the present circumstances, there seems to be little escaping from the conclusion that traditional estimates of the radial distribution functions for water overestimate the sharpness of the first peak in $g_{OO}(r)$, and give the near-neighbour distance for water at too small an r value. In the present instance it is the exercise of trying to fit the x-ray dataset to higher Q values that led to the softer potential given by equation (18), since fitting in the Q range up to $\sim$10 Å^{-1}, as is done in [9, 16] and [24], apparently is not able to establish accurately the sharpness of this peak.

Additionally it seems from figure 9(a) and table 4 that the *position* of this peak tends to be underestimated by classical potentials. Since the most recent interpretations of the neutron diffraction data [22] are based on computer simulation using a classical potential with charges (whereas earlier interpretations, [8] and [26], were based on Fourier inversions of the diffraction data), and since the neutron data do not appear to be overly sensitive to the precise position and sharpness of this peak (compare [26] with [25] for example), it seems likely that these more recent interpretations may have been overly biased by the hardness of Lennard-Jones core potential used in these simulations. Furthermore, it will also be seen that the position of

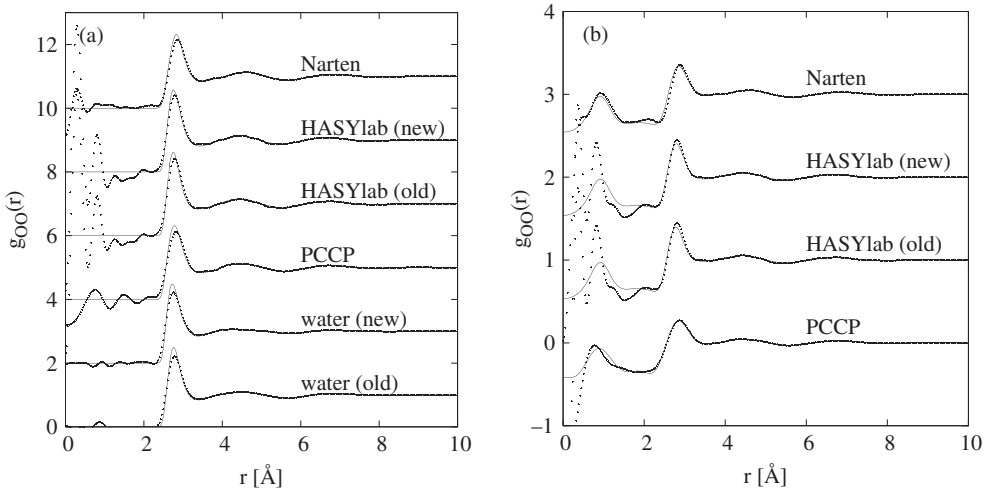

Figure 9. (a) Comparison of the simulated (lines) and estimated (dots) $g_{OO}(r)$ functions for the six EPSR simulations discussed in this paper. A summary of the different first peak positions and heights is given in table 4. Note how in every case the simulation appears to make the rising edge of the $g_{OO}(r)$ function too steep compared to the data. (b) Fourier transform of the fits to the x-ray data in the four simulations that invoke these data (lines), compared to the Fourier transform of the renormalized data themselves (dots), using a number density of 0.1002 atoms Å^{-3}. Both Fourier transforms are performed over the Q range of the diffraction data in each case.

Table 4. Comparison of position and height of first peak in $g_{OO}(r)$ between estimates from the data and estimates from the simulations (figure 9(a)).

	Estimated from the data		Estimated from the simulation	
Data set	Peak position (Å)	Peak height	Peak position (Å)	Peak height
Water (old)	2.78	2.23	2.76	2.50
Water (new)	2.76	2.22	2.73	2.47
PCCP	2.82	2.12	2.78	2.31
HASYlab	2.79	2.42	2.75	2.62
HASYlab (new)	2.78	2.40	2.76	2.58
Narten	2.86	2.15	2.83	2.32

the *second* peak in $g_{OO}(r)$ also has some uncertainty between different datasets, particularly when only the neutron data are used in the reconstruction, figure 9(a). This uncertainty is much reduced when the x-ray data are included, confirming the constraining effect that the x-ray diffraction data can have on the structure refinement.

Of course in the comparison shown in figure 9(a) it could be claimed that the Fourier transform of the estimated OO partial structure factor is going to be affected by the truncation in the diffraction data, while the simulated $g_{OO}(r)$ will not be affected by truncation effects. Figure 9(b) shows, for the four simulations that invoke the x-ray diffraction data, the Fourier transform of the renormalized data, together with the Fourier transform of the simulated fit to those data *over the same Q range as the data*. Here we again see that the simulation still tends to slightly underestimate the first peak position in all four cases. The trend with the peak heights is not so clear however, with the Narten and PCCP simulations apparently getting the peak height nearly the same as the Fourier transformed data, while the HASYlab

simulations apparently underestimating the peak height compared to the data. Since there are issues with systematic effects with HASYlab data that have already been alluded to, it is not clear whether this discrepancy is a consequence of the simulation procedure itself, or has to do with an inherent defect in the data. Comparing figure 9(a) with 9(b), it is clear that the value of $Q_{\max}$ for any given dataset can have a major impact on our understanding of the precise position and heights of peaks in the radial distribution functions of water.

There is a separate question here of whether EPSR is really up to the task of producing an accurate model of the structure of water based on diffraction data. This is a matter over which no doubt opinions will differ, but what does clearly emerge in the analysis presented here is that, whatever the shortcomings of the EPSR method itself, there really are significant uncertainties in both neutron and x-ray diffraction data obtained from different sources, and these are currently at least as significant as any failings in the EPSR method. The primary conclusion that there is a fundamental need for diffraction data with much improved reliability at all Q values, but especially at high Q values in particular if we are to understand the true shape and position of the principle peak in $g_{\mathrm{OO}}(r)$, does not depend on whether EPSR is fully successful or not.

Finally it must always be borne in mind that the fundamental assumption here is that the quantum differences in structure between heavy and light water are sufficiently small to not interfere with the process of obtaining an accurate fit to the data. Since we know this assumption is ultimately untenable, since the structures are different to greater or lesser extent [23, 27, 28], we hope to report on a test of this assumption in the near future.

It should be noted that whereas joint use of x-ray and neutron data with EPSR is a relatively new feature, it has been incorporated into RMC simulations since near the beginning [29]. In comparing the present work with an earlier RMC study [30], which also performed a similar joint x-ray and neutron diffraction structure refinement on water diffraction data, the conclusions are similar, in that a range of models was found, consistent with the diffraction data, and that the existing diffraction data are currently not accurate enough to give an unambiguous view of water structure. However, because in EPSR the molecular structure is strongly constrained by the reference potential throughout the simulation, rather than by using coordination constraints as in [30], and because hydrogen bonding between water molecules is enforced via the partial atomic charges on the simulated water molecules, the range of structures that can be sampled by EPSR is much narrower than is found with RMC. Since it is almost universally agreed that hydrogen bonding between water molecules is a dominant feature of the structure of water and ice, the assumption of this bonding by means of a model interaction potential at the outset of the structure refinement process does not seem to be unreasonable assumption to make. This matter has been discussed at greater length in a recent article [31].

5. Conclusion

It has been demonstrated that joint structure refinement of neutron and x-ray diffraction data is certainly possible using empirical potential structure refinement, provided the x-ray data are placed on a scale that permits direct comparison with the neutron data and simulation. In the case of water the best agreement with the x-ray data is obtained by assuming that some fraction of the charge on the hydrogen atom is transferred to the oxygen atom. The overall fits using a simple point charge reference potential for the oxygen and hydrogen atoms are satisfactory after introducing a soft core repulsive potential, but even then seem to yield too small a near-neighbour OO distance. The OH and HH RDFs are not strongly affected by whether the x-ray data are included in the structure refinement, but the OO RDF is clearly strongly determined by those data. A significant finding is that the extant x-ray and neutron diffraction

data do not have sufficient high-Q resolution and accuracy for the true shape and position of the first peak in the OO RDF to be determined reliably. Equally, different neutron datasets obtained under different conditions give slightly different estimates for the radial distribution functions.

In spite of these caveats, however, it is clear that for the purposes of structure refinement of liquids and glassy materials data x-ray and neutron diffraction data are highly complementary, and should be incorporated wherever possible in structure refinement programs if a more complete structure analysis is to be obtained. The uncertainties surrounding the correct electron form factors and data normalization to use are more than offset by the contrasting information that the different datasets provide. The lack of reliable high-Q data in the case of water is of concern, however, as it does seem to affect the precision to which the underlying shape and positions of peaks in the radial distribution functions can be determined.

Acknowledgments

I am grateful to Chris Benmore, Teresa Head-Gordon and Greg Hura, for supplying numerical versions of their respective x-ray datasets.

References

[1] Warren B E 1969 *X-ray Diffraction* (Reading, MA: Addison-Wesley)
[2] Petkov V 1989 *J. Appl. Crystallogr.* **22** 387
[3] Keen D 2001 *J. Appl. Crystallogr.* **34** 172
[4] Hansen J P and McDonald I R 1986 *Theory of Simple Liquids* 2nd edn (London: Academic)
[5] Faber T E and Ziman J M 1965 *Phil. Mag.* **11** 153
[6] Placzek G 1952 *Phys. Rev.* **86** 377
[7] Waasmaier D and Kirfel A 1995 *Acta Crystallogr.* A **51** 416–31 These are listed in the DABAX library. More information on DABAX can be found at http://www.esrf.fr/computing/expg/subgroups/theory/DABAX/dabax.html
[8] Narten A H and Levy H A 1971 *J. Chem. Phys.* **55** 2263
[9] Sorensen J M, Hura G, Glaeser R M and Head-Gordon T 2000 *J. Chem. Phys.* **113** 9149
[10] Wang J, Tripathi A N and Smith V H 1994 *J. Chem. Phys.* **101** 4842
[11] Hajdu F, Lengyel S and Pálinkás G 1976 *J. Appl. Crystallogr.* **9** 134–8
[12] Neuefeind J, Benmore C J, Tomberli B and Egelstaff P A 2002 *J. Phys.: Condens. Matter* **14** L429–33
[13] Badyal Y S, Saboungi M-L, Price D L, Haeffner D R and Soper A K 2000 *J. Chem. Phys.* **112** 9206–8
[14] Sears V F 1992 *Neutron News* **3** 26
[15] Soper A K, Howells W S and Hannon A C 1989 ATLAS—analysis of time-of-flight diffraction data from liquid and amorphous samples *Rutherford Appleton Laboratory* Report RAL-89-046
[16] Hura G, Sorensen J M, Glaeser R M and Head-Gordon T 2000 *J. Chem. Phys.* **113** 9140
[17] Urquidi J, Benmore C J, Neuefeind J and Tomberli B 2003 *J. Appl. Crystallogr.* **36** 368
[18] Soper A K 2001 *Mol. Phys.* **99** 1503–16
[19] Soper A K 2005 *Phys. Rev.* B **72** 104204
[20] Soper A K 2005 *J. Phys.: Condens. Matter* **17** S3273–82
[21] Berendsen H J C, Grigera J R and Straatsma T P 1987 *J. Phys. Chem.* **91** 6269
[22] Soper A K 2000 *Chem. Phys.* **258** 121–37
[23] Hart R T, Benmore C J, Neuefeind J and Kohara S 2005 *Phys. Rev. Lett.* **94** 047801
[24] Hura G, Russo D, Glaeser R M, Head-Gordon T, Krack M and Parrinello M 2003 *Phys. Chem. Chem. Phys.* **5** 1981
[25] Soper A K, Bruni F and Ricci M A 1997 *J. Chem. Phys.* **106** 247
[26] Soper A K and Phillips M G 1986 *Chem. Phys.* **107** 47–60
[27] Kuharski R A and Rossky P J 1985 *J. Chem. Phys.* **82** 5164
[28] Guillot B and Guissani Y 1998 *J. Chem. Phys.* **108** 10162–74
[29] Keen D A and McGreevy R L 1990 *Nature* **344** 423–5
[30] Pusztai L 1999 *Phys. Rev.* B **60** 11851–4
[31] McLain S E, Imberti S, Soper A K, Botti A, Bruni F and Ricci M A 2006 *Phys. Rev.* B **74** 094201