



## On the calculation of x-ray scattering signals from pairwise radial distribution functions

**Dohn, Asmus Ougaard; Biasin, Elisa; Haldrup, Kristoffer; Nielsen, Martin Meedom; Henriksen, Niels Engholm; Møller, Klaus Braagaard**

*Published in:*  
Journal of Physics B: Atomic, Molecular and Optical Physics

*Link to article, DOI:*  
[10.1088/0953-4075/48/24/244010](https://doi.org/10.1088/0953-4075/48/24/244010)

*Publication date:*  
2015

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Dohn, A. O., Biasin, E., Haldrup, K., Nielsen, M. M., Henriksen, N. E., & Møller, K. B. (2015). On the calculation of x-ray scattering signals from pairwise radial distribution functions. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 48(24), [244010]. <https://doi.org/10.1088/0953-4075/48/24/244010>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# On the Calculation of X-ray Scattering Signals from Pairwise Radial Distribution Functions

Asmus O. Dohn<sup>\*</sup>, Elisa Biasin<sup>\*\*</sup>, Kristoffer Haldrup<sup>\*\*</sup>, Martin M.  
Nielsen<sup>\*\*</sup>, Niels E. Henriksen<sup>\*</sup>, and Klaus B. Møller<sup>\*</sup>

<sup>\*</sup>Department of Chemistry, Technical University of Denmark,  
Kemitorvet 207, 2800, Kgs. Lyngby

<sup>\*\*</sup>Department of Physics, Technical University of Denmark, Fysikvej  
307, 2800, Kgs. Lyngby

September 2, 2015

## Abstract

We derive a formulation for evaluating (time-resolved) X-ray scattering signals of solvated chemical systems, based on pairwise radial distribution functions, with the aim of this formulation to accompany molecular dynamics simulations. The derivation is described in detail to eliminate any possible ambiguities, and the result includes a modification to the atom-type formulation which to our knowledge is previously unaccounted for. The formulation is numerically implemented and validated.

## Introduction

This work is concerned with the derivation of a formulation for evaluating time-resolved X-ray scattering signals of chemical systems in solution phase, based on pairwise Radial Distribution Functions (RDFs). We aim to provide a full formalism, with the complete set of details behind its derivation, to go hand-in hand with Molecular Dynamics (MD) simulations. These simulations are becoming more and more ubiquitous in - and essential for - the analysis of ultra-fast X-ray scattering experiments of solvated molecules[1–4]. The motivation for the work is threefold:

**The physical aspect:** Due to quantum and/or statistical ensemble effects, interatomic distances are inherently governed by probability distributions. The much used Debye formulation[5] of X-ray scattering, historically derived with crystalline systems in mind, assumes definite interatomic distances. Broadening can be introduced retrospectively, e.g. by averaging over scattering signals from a collection of distinct structures or using the Debye-Waller model. However, it seems appropriate to have a formulation of X-ray scattering based directly on the continuous nature of the actual probability distributions describing the physical situation.

**The practical aspect:** For any analysis, it is always desirable to cut computational costs and unnecessary complexity. Averaging over Debye signals scales with the number of atoms squared times the number of system configurations (frames) used. The fastest programs developed (mainly for crystallographic purposes) state that computing the scattering e.g. for a system of the equivalent size of 1000 frames from a 50x50x50 Å box would take days, and should be run on a cluster[6]. Programs have been developed[7] that can calculate RDFs for systems of similar sizes in hours if not minutes on a desktop computer, equipped with a GPU[8]. The method presented here could furthermore allow for direct manipulation of the RDFs, in a fitting-scenario. Either way, adopting the formulation introduced in this work can drastically improve the analysis workflow of X-ray scattering experiments in the solution phase.

**The unambiguity aspect:** The first work on this formulation is almost a century old[9], and many versions and modifications have been presented[1, 5, 10–23]. Much of the work has been based on determining the atomic distributions from scattering signals, not vice versa, and not always with explanations for modifications made, compared to previous publications. If some of these models are lifted straight from their papers and implemented numerically, substantial errors in the resulting signal can occur. We therefore believe it is necessary to present a fully-detailed derivation of the formalism that explains the choices and background for each step, to take advantage of the improvements in the physical and practical aspects that this formulation has to offer.

## Derivation of the Formulation

For molecular X-ray scattering restricted to a single electronic state, the elastic scattering signal in units of the Thomson cross-section (a.k.a. electronic units)  $S(\mathbf{q})$ , where  $\mathbf{q}$  is the scattering vector, is a function of the nuclear probability distribution  $\rho^{(N)}(\mathbf{R}, t)$  of the  $N$ -atom system with coordinates  $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2 \dots \mathbf{R}_N)$ , and the molecular form factor  $F(\mathbf{R}, \mathbf{q})$ [24]:

$$S(\mathbf{q}) = \int_0^\infty I(t) \int_{VN} \rho^{(N)}(\mathbf{R}, t) |F(\mathbf{R}, \mathbf{q})|^2 d^N \mathbf{R} dt \quad (1)$$

where  $I(t)$  is the intensity function of the X-ray probe pulse, and  $V$  is the volume of the sample over which the irradiating X-ray beam is coherent. If the total irradiated sample volume is larger than the coherence lengths of the beam, the total scattering signal from the sample is the sum of individual scattering signals given by equation 1, arising from sub-volumes over which the beam can be considered coherent[5, 25]. The molecular form factor is in principle the expectation value of the scattering operator on the all electron wavefunction, which can be shown to give a Fourier transform of the electronic density  $\rho_e(\mathbf{r}; \mathbf{R})$  [24]:

$$F(\mathbf{R}, \mathbf{q}) = \int_V \rho_e(\mathbf{r}; \mathbf{R}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \quad (2)$$

Almost always, the assumption is made that the scattering can be described as scattering from independent atoms, with spherical electronic densities. This is called the Independent Atom Model (IAM), which effectively turns the molecular scattering factor into a sum of atomic form factors  $f_j(q)$ :

$$F_{\text{IAM}}(\mathbf{R}, \mathbf{q}) = \sum_j^N f_j(q) e^{i\mathbf{q}\cdot\mathbf{R}_j} \quad (3)$$

Even though this approximation ignores chemical bonding, it is in most cases very accurate for resolving molecular geometries[26], with few exceptions[18], see also[27]. While the full electronic distribution is directly available via *ab initio* simulations, using it directly with an adequate numerical precision within a multidimensional structural-fitting strategy, as employed e.g. by the experimental section of our group[28, 29], is unnecessarily cumbersome when working with systems where the IAM is sufficient.

The signal in the IAM is:

$$\begin{aligned} S(\mathbf{q}) &= \int_{V^N} \rho^{(N)}(\mathbf{R}) |F(\mathbf{R}, \mathbf{q})|^2 d^N \mathbf{R} \\ &= \int_{V^N} \rho^{(N)}(\mathbf{R}) \sum_j f_j(q) e^{-i\mathbf{q} \cdot \mathbf{R}_j} \sum_k f_k(q) e^{i\mathbf{q} \cdot \mathbf{R}_k} d^N \mathbf{R} \end{aligned} \quad (4)$$

where  $\rho^{(N)}(\mathbf{R})$  is the average nuclear probability during the duration of the X-ray probe pulse. We first separate the sums into sums of  $j = k$  and  $k \neq j$ :

$$S(\mathbf{q}) = \int_{V^N} \rho^{(N)}(\mathbf{R}) \left( \sum_j f_j(q)^2 + \sum_j \sum_{k \neq j} f_j(q) f_k(q) e^{-i\mathbf{q} \cdot (\mathbf{R}_j - \mathbf{R}_k)} \right) d^N \mathbf{R} \quad (5)$$

Since the density is normalised, after integration the first sum is simply  $\sum_j f_j(q)^2$ . We now focus on the second sum. In our previous work, it has been favourable to transform the coordinates to an internal basis[24, 30, 31], while here, we simply note that the exponential in each term is only dependent on the positions of two particles for each term in the sum. Hence, for each pair  $j, k$ , a pairwise density distribution function is constructed by integrating all the other dependencies out:

$$\begin{aligned} \rho_{j,k}^{(2)}(\mathbf{R}_j, \mathbf{R}_k) &= \\ &= \int_{V^{N-2}} \rho^{(N)}(\mathbf{R}) d\mathbf{R}_1 d\mathbf{R}_2 \cdots d\mathbf{R}_{j-1} d\mathbf{R}_{j+1} \cdots d\mathbf{R}_{k-1} d\mathbf{R}_{k+1} \cdots d\mathbf{R}_N \end{aligned} \quad (6)$$

We then substitute the integration variables  $\mathbf{r} = \mathbf{R}_j - \mathbf{R}_k$ , and  $\mathbf{r}' = (\mathbf{R}_j + \mathbf{R}_k)/2$ , and pull the distribution  $\rho_{j,k}^{(2)}(\mathbf{r}, \mathbf{r}')$  through the sum in the integral of equation 5:

$$\begin{aligned} \sum_j \sum_{k \neq j} f_j(q) f_k(q) \int_{V^2} \rho_{j,k}^{(2)}(\mathbf{r}, \mathbf{r}') e^{-i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} d\mathbf{r}' &= \\ \sum_j \sum_{k \neq j} f_j(q) f_k(q) \int_V \rho_{j,k}^{(1)}(\mathbf{r}) e^{-i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} \end{aligned} \quad (7)$$

where  $\rho_{j,k}^{(1)}(\mathbf{r}) = \int_V \rho_{j,k}^{(2)}(\mathbf{r}, \mathbf{r}') d\mathbf{r}'$ , which leaves us with sum of integrals over each atomic pair and its corresponding probability density. Dropping

the pair density-superscript on the probability distribution for brevity, we write up the full equation:

$$S(\mathbf{q}) = \sum_j f_j(q)^2 + \sum_j \sum_{k \neq j} f_j(q) f_k(q) \int_V \rho_{j,k}(\mathbf{r}) e^{-i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} \quad (8)$$

Now we make the isotropic assumption, meaning that there is equal probability of finding the entire molecule in any orientation. This must mean that there is also equal probability of finding *each* of the intermolecular vectors in any orientation, since a rotation of the molecule must mean a rotation of all of its pairwise vectors. In other words, we assume that  $\rho_{j,k}(\mathbf{r}) = \rho_{j,k}(r)$ . For ultrafast studies, the isotropic assumption might not always hold, but we note that all the available structural information can be extracted from the isotropic contribution to the total scattering[24]. Options for including angular dependence in scattered intensity-equations of isolated molecules have been derived elsewhere[24, 31, 32]. Evaluating the integral in the isotropic case leads to[5, 12]:

$$\int_V \rho_{j,k}(r) e^{-i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} = 4\pi \int_0^R \rho_{j,k}(r) \frac{\sin(qr)}{qr} r^2 dr \quad (9)$$

where  $R$  is the radius of the (on average spherically symmetric) coherence volume in the sample. Including the first sum in equation 8 we get:

$$S(q) = \sum_j f_j(q)^2 + \sum_j \sum_{k \neq j} f_j(q) f_k(q) 4\pi \int_0^R \rho_{j,k}(r) \frac{\sin(qr)}{qr} r^2 dr \quad (10)$$

The double sum in equation 10 is over all atomic pairs in the system.

With distribution functions, one can collect correlations between atoms by grouping atoms into sets of 'types', for a suitable definition of 'type'. With X-ray scattering in mind, the largest possible sets can be made by grouping all atoms exhibiting identical scattering behaviour under the same type. Within the IAM this corresponds to equating type and element/ion<sup>1</sup>. However, we are free to further divide species of the same element/ion into different types, e.g. by distinguishing whether an atom belongs to the solute or the solvent, as shall become relevant later. With this definition of atom types, we need to redistribute the probability terms as follows: If  $N_l$  and  $N_m$  are the numbers of all atoms of type  $l$  and type  $m$ , respectively, we collect all the probability distributions  $\rho_{j,k}(r)$  where atom  $j$  is of type  $l$ , and atom  $k$  is of type  $m$  into one distribution,  $\rho_{l,m}(r)$ . Since we also need to construct the probability distributions between different atoms in the same

<sup>1</sup>since there are tabulated values for form factors for the same element with different charges[33].

type, we need to make sure that the same atom ( $j = k$ ) is not counted twice in these cases of  $l = m$ :

$$\rho_{l,m}(r) = \frac{1}{N_l(N_m - \delta_{l,m})} \sum_{j \in l} \sum_{\substack{k \in m \\ k \neq j}} \rho_{j,k}(r) \quad (11)$$

where  $\delta_{l,m}$  is the Kronecker delta. As mentioned,  $l$  can be equal to  $m$ . In these cases,  $k \neq j$  in the summation, and to avoid pairing atoms with themselves, the normalization factor is  $\frac{1}{N_l(N_m-1)}$ .

Since the form factors in equation 10 are not dependent on  $r$ , they can be included in the integral and the sum can be rewritten with the definition of atom types:

$$\begin{aligned} \sum_{j \neq k} f_j(q) f_k(q) \rho_{j,k}(r) &= \sum_l \sum_m \sum_{j \in l} \sum_{\substack{k \in m \\ k \neq j}} f_j(q) f_k(q) \rho_{j,k}(r) \\ &\text{where } f_j = f_l \text{ if } j \in l \text{ so:} \\ &= \sum_l \sum_m f_l(q) f_m(q) \sum_{j \in l} \sum_{\substack{k \in m \\ k \neq j}} \rho_{j,k}(r) \\ &\text{and using the definition in equation 11:} \\ &= \sum_l \sum_m N_l(N_m - \delta_{l,m}) f_l(q) f_m(q) \rho_{l,m}(r) \quad (12) \end{aligned}$$

which is then substituted into equation 10. Using the contracted notation for the double sum over  $l$  and  $m$ :

$$S(q) = \sum_l N_l f_l(q)^2 + \sum_{l,m} f_l(q) f_m(q) N_l(N_m - \delta_{l,m}) 4\pi \int_0^R \rho_{l,m}(r) \frac{\sin(qr)}{qr} r^2 dr \quad (13)$$

where we also rewrote the sum of form factors squared for each atom, to the atom-type notation, which is  $\sum_{j \in l} f_j(q)^2$ . To finally express the scattering in terms of the (pairwise) radial distribution functions  $g_{l,m}(r)$ , easily obtainable from MD simulations, we recall that the probability densities and number densities are proportional, so we can use an equivalent analogy to the standard definition[10]:  $g_{l,m}(r) = \frac{\rho_{l,m}(r)}{\rho_{0,l,m}}$ , where  $\rho_{0,l,m} = 1/V$  is the isotropic probability density, with  $V$  being the coherence volume. Thus:

$$S(q) = \sum_l N_l f_l(q)^2 + \sum_{l,m} f_l(q) f_m(q) \frac{N_l(N_m - \delta_{l,m})}{V} 4\pi \int_0^R r^2 g_{l,m}(r) \frac{\sin(qr)}{qr} dr \quad (14)$$

which is the main result of this work. Illustrations of  $g_{l,m}(r)$  functions can be found in figure 1. The  $l = m$  terms in the double sum seem to be left out of some presentations of the formalism[1, 22], which would result in the neglect of all contributions of scattering from different atoms of the same atom type. In other works[19],  $l$  can be equal to  $m$  in the sum, but the  $-\delta_{l,m}$  term is left out. If  $N_m$  is very large, this is a fine approximation, but the approximation will *not* hold for systems containing a few particularly intense scatterers such as Pt, Ir, I, etc. On a diatomic system of the same atom type, neglecting  $-\delta_{l,m}$  will result in a twice as large  $l = m$ -term contribution to the double sum, compared to the correct result. How the full scattering signal is affected is exemplified in figure 2.

For practical applications of equation 14 a few more steps are needed, but first, we show that equation 14 is a generalisation of the Debye-equation. If we choose the types such that each atom is its own type, the RDF for an isolated pair of atoms  $j$  and  $k$  at distance  $r_{j,k}$  must be a delta function, and  $N_j = N_k = 1$ :

$$4\pi r^2 \rho_{j,k}(r) dr = \frac{4\pi r^2 g_{j,k}(r)}{V} dr = \delta(r - r_{j,k}) dr \quad (15)$$

which we can insert into equation 14:

$$S(q) = \sum_j N_j f_k(q)^2 + \sum_{j,k} f_j(q) f_k(q) N_j (N_k - \delta_{j,k}) \int_0^R \delta(r - r_{j,k}) \frac{\sin(qr)}{qr} dr$$

leading to the Debye-equation:

$$S(q) = \sum_j \sum_k f_j(q) f_k(q) \frac{\sin(qr_{j,k})}{qr_{j,k}} \quad (16)$$

where we have collected the two sums back into one term, since  $\frac{\sin(qr_{j,k})}{qr_{j,k}} \rightarrow 1$  for  $r_{j,k} \rightarrow 0$ , i.e. for  $j = k$ .

Returning to the applications of equation 14, one can conveniently rewrite the integral in the equation by adding and subtracting the distribution in the constant-density limit at long distances,  $g_{0,l,m}$  (see discussion below) such that:

$$\begin{aligned} \int_0^R g_{l,m}(r) \frac{\sin(qr)}{qr} r^2 dr &= \\ \int_0^R [g_{l,m}(r) - g_{0,l,m}] \frac{\sin(qr)}{qr} r^2 dr + g_{0,l,m} \int_0^R \frac{\sin(qr)}{qr} r^2 dr & \quad (17) \end{aligned}$$

For large values of  $R$ , the last term has been argued to only contribute at  $q \rightarrow 0$ , and therefore often excluded[10, 11, 13–15], since this part of the



$q$ -range is often covered in the experimental setup by the beamstop. This claim deserves a bit of elaboration. Evaluating the integral gives

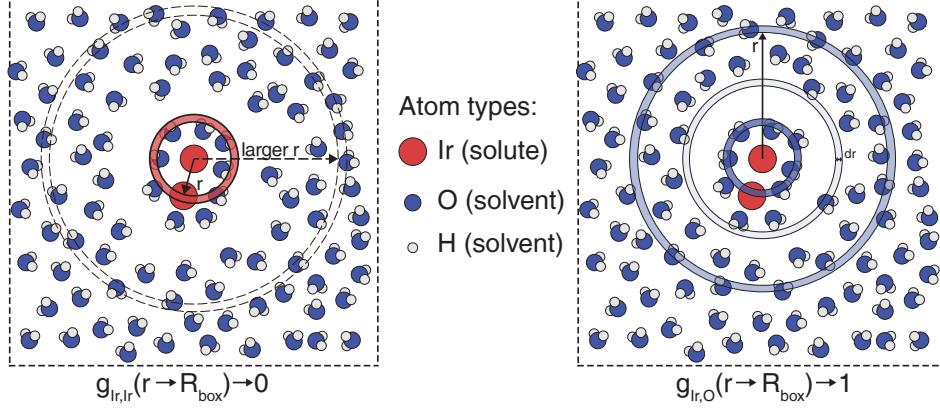
$$\int_0^R \frac{\sin(qr)}{qr} r^2 dr = \frac{\sin(qR)}{q^3} - \frac{R \cos(qR)}{q^2} \quad (18)$$

Since  $R \gg q^{-1}$ , for  $q$  being on the order of  $1 \text{ \AA}^{-1}$ , the resulting absolute value of the integral will not be small, compared to the first term of equation 17, in the experimentally interesting region. However, the resulting value is rapidly oscillating in  $q$  with a period of  $2\pi/R$ . In traditional X-ray experiments, the coherence length  $R$  of the beam is typically much shorter (approx  $0.1\text{-}1 \mu\text{m}$  [34]) than the extent of the irradiated sample (on the order of, say,  $10\text{-}100 \mu\text{m}$ ). This implies that the rapidly oscillating integrals, eqn. 18, cancel out in the total scattering emitted from the irradiated sample [5, 16, 25]. We note that it is indeed possible on modern X-ray sources to construct experiments where the probed sample volume is tailored to match the coherence length of the beam [25, 35–39], but this will not be considered further here. For the time-resolved scattering community, the discussion of this particular term in equation 18 can be avoided altogether by evaluating ‘difference-distributions’ and calculating the difference scattering directly from those, as we shall see later.

In conclusion, when the total experimental signal is a sum of scattering signals  $S(q)$  from single coherence volumes, we can omit the last term in equation 17:

$$S(q) = \sum_l N_l f_l(q)^2 + \sum_{l,m} f_l(q) f_m(q) \frac{N_l(N_m - \delta_{l,m})}{V} 4\pi \int_0^R r^2 [g_{l,m}(r) - g_{0,l,m}] \frac{\sin(qr)}{qr} dr \quad (19)$$

which can be implemented numerically. Often [1, 19, 22] (but not always [14, 15]),  $g_{0,l,m}$  is simply written as 1, since  $g_{l,m}(r)$  is normalised w.r.t the macroscopic density. In the case of pure liquids, and even for dilute samples with macroscopic volumes, this is true. However, in MD simulations of solutions with finite box dimensions, not all pairs of atom types have radial distributions that reach the convergence limit of 1 within the simulation dimensions. Considering the actual computation of the scattering signal from an MD simulation, in order to subtract the correct  $g_{0,l,m}$ -value, we can make use of the previously mentioned freedom in defining atom types to define subgroups based on whether the atoms belong to the solute or the solvent (see fig. 1). For solvent-solvent RDFs, where both atom types  $l$  and  $m$  belong to the solvent subgroup,  $g_{l,m}(r) \rightarrow 1$  typically within  $10\text{-}20 \text{ \AA}$ . The same is the case for solute-solvent pairs, as long as the solute itself is not too large. The MD box should be large enough for the convergence to take place. The solute-solute distance in a dilute sample is so long that the



**Figure 1:** Illustration of the calculation of various pairwise RDFs, where the atom types are divided into subgroups based on whether they belong to the solute or the solvent. The solute here is an artificial Ir-diatomic, a simplified example based on real, previously studied systems, [40, 41]. On the left illustration where  $l = m = \text{Ir}$ , the  $g_{\text{Ir,Ir}}(r)$  will go to 0 for  $r \rightarrow R_{\text{box}}$ . For the solute-solvent pair on the right illustration, the simulation box is large enough that  $g_{\text{Ir,O}}(r)$  will converge to 1, past the solvent-shell quasi-structure. In the region just after the first solvation shell,  $g_{\text{Ir,O}}(r) < 1$

intersolute scattering can be neglected. In an MD simulation typically only one solute molecule is included, and therefore, the computed  $g_{l,m}(r)$  for  $l, m$  types in the solute will go to zero.

With these values for  $g_{0,l,m}$ , the difference  $g_{l,m}(r) - g_{0,l,m}$  will go to zero within the simulation box. We can therefore replace  $R$  of the sample by  $R_{\text{box}}$  of the box dimensions as the upper integration limit, as the length of the simulation box is typically much shorter than the coherence length of the X-ray beam. Furthermore, for clarity (and sometimes for convenience in the further analysis) we can split up the scattering signal into contributions from solvent-solvent terms  $S_v(q)$ , solute-solvent (cross) terms  $S_c(q)$ , and solute-solute terms  $S_u(q)$ :

$$S_v(q) = \sum_l^v N_l f_l(q)^2 + \sum_{l,m}^v f_l(q) f_m(q) \frac{N_l(N_m - \delta_{l,m})}{V} 4\pi \int_0^{R_{\text{box}}} r^2 [g_{l,m}(r) - 1] \frac{\sin(qr)}{qr} dr \quad (20a)$$

$$S_c(q) = \sum_l^c N_l f_l(q)^2 + \sum_{l,m}^c f_l(q) f_m(q) \frac{N_l(N_m - \delta_{l,m})}{V} 4\pi \int_0^{R_{\text{box}}} r^2 [g_{l,m}(r) - 1] \frac{\sin(qr)}{qr} dr \quad (20b)$$

$$S_u(q) = \sum_l^u N_l f_l(q)^2 + \sum_{l,m}^u f_l(q) f_m(q) \frac{N_l(N_m - \delta_{l,m})}{V} 4\pi \int_0^{R_{\text{box}}} r^2 g_{l,m}(r) \frac{\sin(qr)}{qr} dr \quad (20c)$$

where the  $v, c, u$ -notation above the sums indicate that the types included in the sums should belong to their respective groups.

In standard MD analysis programs, the RDFs are normalized using the simulation box volume  $V = V_{\text{box}}$ , in conjunction with the number of particles  $N_l$  of type  $l$  and  $N_m$  of type  $m$  in the simulation box. Hence, for numerical consistency, one should use the same values in the above equations. However, this will result in a scaling of the absolute signal by  $N_{\text{box}}/N_{\text{sample}}$ . Furthermore, since the solute concentration in the simulation box probably varies from the solute concentration in the sample, the three contributions  $S_u(q)$ ,  $S_c(q)$ , and  $S_v(q)$  must be scaled accordingly.

Two further issues arise when the numerical implementation is made:

1. Constraining bonds is an often used technique within MD. The numerical representation of constrained bond intramolecular  $g_{j,k}(r)$ 's with delta-distribution-like characteristics depends on the numerical precision in  $dr \approx \Delta r$ , the bin width for the sampled distributions.
2. The integral in equation 19 is truncated at  $R_{\text{box}}$ .

The second issue introduces spurious truncation oscillations in the calculated scattering signal, since the integral is essentially the Fourier transformation of  $g_{l,m}(r)$ . Many methods have been applied to this problem, often for the reverse version of obtaining  $g_{l,m}(r)$  functions from  $S(q)$ [12, 18, 21, 23, 42–45]. Some fit the tail of the data to an analytic function[42], while others apply a damping function to the Fourier transformation[21, 44], and others again have developed more involved methods[18, 23, 43]. We have found it adequate so far to simply employ a damping function  $\frac{\sin(\pi r/L)}{\pi r/L}$ [20, 21, 45] in the transformation:

$$S(q) = \sum_l N_l f_l(q)^2 + \sum_{l,m} f_l(q) f_m(q) \frac{N_l(N_m - \delta_{l,m})}{V} 4\pi \int_0^{R_{\text{box}}} r^2 [g_{l,m}(r) - g_{0,l,m}] \frac{\sin(qr)}{qr} \frac{\sin(\frac{\pi r}{L})}{\frac{\pi r}{L}} dr \quad (21)$$

Some authors[21] choose  $L$  to be half the size of the simulation box, while others[44] provide no physical justification for their chosen value.

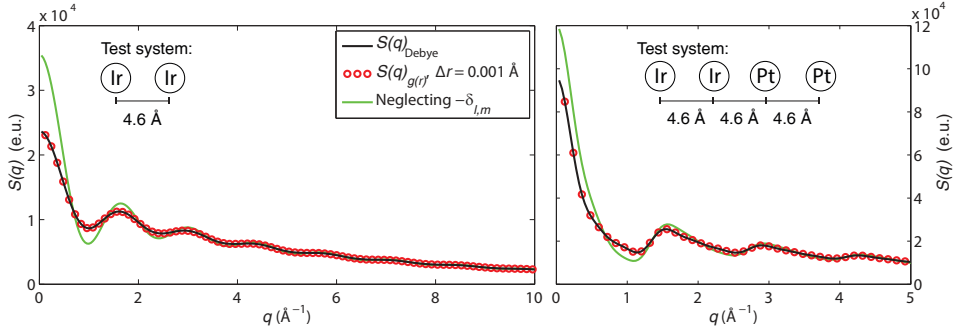
In its most basic description, time resolved X-ray pump-probe scattering experiments are carried out by using a 'pump' laser pulse to electronically excite the solute, instigating the dynamics, and then measuring the scattering,  $S_{\text{on}}(q)$  at a given time  $t$  after excitation, by an ultrashort X-ray 'probe' pulse. Then, the signal from the unpumped sample  $S_{\text{off}}(q)$  is subtracted to create  $\Delta S(q) = S_{\text{on}}(q) - S_{\text{off}}(q)$ , only containing contributions from the transient features of the studied geometries. Exploiting the fact that form factors are identical in the 'on' and 'off' state within the IAM, namely via 'difference-distributions', the integral in equation 14 can be made to con-

verge inside the simulation box in another way:

$$\Delta S(q) = \sum_{l,m} f_l(q) f_m(q) \frac{N_l(N_m - \delta_{l,m})}{V} 4\pi \int_0^{R_{\text{box}}} \Delta g_{l,m}(r) \frac{\sin(qr)}{qr} r^2 dr \quad (22)$$

where  $\Delta g_{l,m}(r) = g_{l,m,\text{on}}(r) - g_{l,m,\text{off}}(r)$ . Here, the discussion of equation 18 becomes moot, since  $g_{l,m,\text{on}}(r)$  and  $g_{l,m,\text{off}}(r)$  have the same limits at long distances.

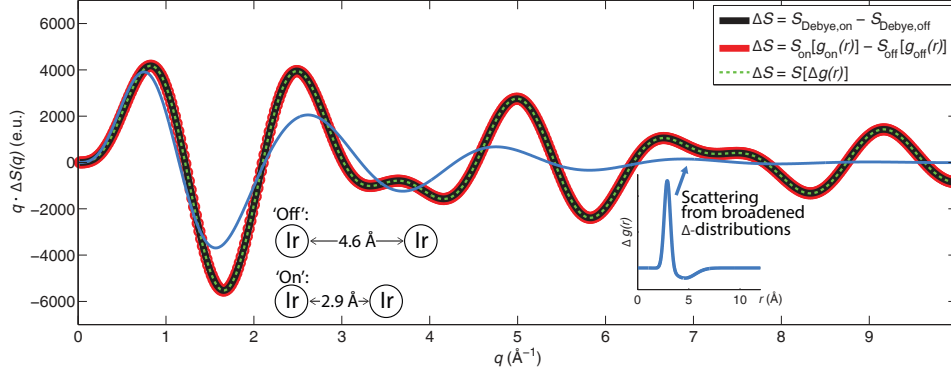
## Evaluating the Derivation Using a Numerical Implementation



**Figure 2:** Comparisons of numerical implementations of calculated X-ray scattering from two single geometries (i.e. no MD-based thermal averaging),  $S(q)$ , implemented in Matlab<sup>®</sup>. The scattered intensity (in electronic units) goes to the number of electrons squared in the sample, when  $q \rightarrow 0$ . The green curves show the resulting calculated scattering from neglecting the  $-\delta_{l,m}$  term, as discussed in the text.

Figure 2 shows a comparison of the numerical implementation of equation 21 made in the Matlab<sup>®</sup> programming language, with a numerical implementation of the Debye formula (equation 16), as previously implemented and used in our group[28, 29]. The simplified test systems are chosen to confirm the validity of the derivation, especially with regards to the counting of atoms of each type, since we in this aspect diverge from previous derivations. Test system 1 consists of a single frame of two Ir atoms 4.6 Å apart, while test system 2 contains two Ir atoms and two Pt atoms on a line, with the same nearest-neighbour spacing as in system 1. For the  $S(q)$  calculated via equation 20c,  $g_{l,m}(r)$  was evaluated in the VMD program[7], with  $\Delta r = 0.001$  Å in a square box of 100 Å side lengths, thus numerically approximating the delta-function when using  $V = V_{\text{box}}$ . For both test systems, the two different methods of calculating the scattering provide identical results, within the numerical accuracy. Furthermore, the validity of the implementation (and therefore the derivation) is supported by the scattered intensity value at  $q \rightarrow 0$ : In this limit, the scattering goes to  $\left| \sum_j f_j(q \rightarrow 0) \right|^2$ . Remembering that, in principle,  $f(q) = \int \rho(r) e^{iqr} dr$ ,

where  $\rho(r)$  is the atomic electronic density, then  $S(q \rightarrow 0) \rightarrow n^2$  where  $n$  is the number of electrons in the system<sup>2</sup>. Thus, it is confirmed that the two implementations produce the same scattered intensity for molecular systems.

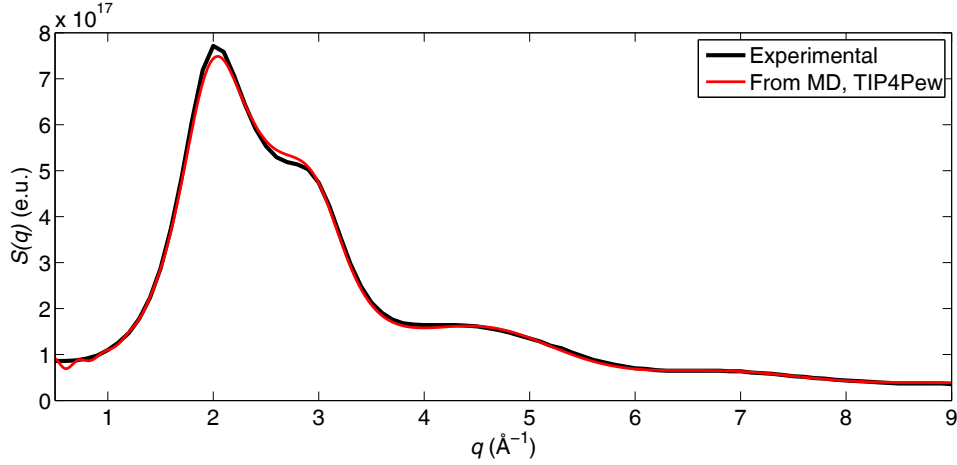


**Figure 3:** Comparison of numerically evaluating the difference-scattering  $\Delta S(q)$  via the Debye equation (equation 16), equation 20c, and equation 22. The example is a simulation of two Ir-Ir atoms 4.6 Å apart ('off') and 2.9 Å ('on'). The blue curve on the plot shows simulated scattering from broadened distributions, more akin to real experimental conditions. The  $\Delta g_{l,m}(r)$ -distribution is constructed by subtracting a gaussian distribution with  $\mu_{\text{off}} = 4.6$  and  $\sigma_{\text{off}} = 1$  from another gaussian distribution with  $\mu_{\text{on}} = 2.9$  and  $\sigma_{\text{off}} = 0.3$  (see inset).

Figure 3 shows that equation 22 successfully reproduces the Debye formulation result for the prototypical time-dependent scattering experiment where two atoms are at a shorter distance after laser excitation (assuming infinitely short pump and probe pulses). The blue curve represents scattering simulated using a broadened  $\Delta g_{l,m}(r)$ -distribution shown in the inset, using gaussian broadening factors based on our recent findings for a bimetallic Ir-system[40] (neglecting the observed anharmonicity of the underlying potential, since the illustration done here is only for explanatory purposes). We note that probabilistic distributions of atomic positions, inherent everywhere in nature, change the observable scattering signal.

The next benchmarking step involves using equation 20a to calculate the scattering of neat water at standard pressure and 300 K, as shown in figure 4. Due to the previously described impracticalities of using the Debye-equation (equation 16) to calculate scattering signals from MD simulations, we compare our calculated scattering of an MD simulation of neat water, using the TIP4P-eW potential[46], to experimentally obtained results[45, 47, 48]. This of course means that differences in the two signals can arise

<sup>2</sup>The calculated scattering intensity in figure 2, left at  $q = 0.02 \text{ \AA}^{-1}$  is  $2.37 \cdot 10^4 = (77 \cdot 2)^2$



**Figure 4:** Red curve: Simulated scattering from an NVT MD simulation of  $(50 \text{ \AA})^3$  pure water at 300 K, using the TIP4P-eW potential[46], calculated using  $N_{\text{O}}$  and  $N_{\text{H}}$  values for a  $46 \mu\text{m}$  sphere of neat water. Black curve: An experimentally obtained[47] scattered X-ray intensity profile from neat water, scaled in  $2.9 \text{ \AA}^{-1} < q < 6.0 \text{ \AA}^{-1}$  to the simulated signal.

from inconsistencies of the model potential with actual water. The simulation reproduces the experimental signal within the precision of the TIP4P-eW potential, and taking into account the already explored problems with using the IAM for water[48]. This concludes the benchmarking of the new numerical implementation, which confirms the derivation.

## Concluding Remarks

In this work, we have provided an in depth derivation of what can be described as a generalisation of the Debye-scattering equation (equation 16), as we demonstrated collapsing the probability distributions to just single distances reverted the derived formulation back into the Debye result. The pairwise RDF-based equations (equations 20a,20b, and 20c) for the X-ray scattering readily provides significant advantages: Both practical, such as being able to use already established tools of high efficiency to obtain the needed RDFs, and from a physical interpretation point of view, e.g. readily splitting up scattering-contributions from various parts of the studied system.

With this review of the background behind the present formulation and considering its successful reproduction of known results, we hope that it will provide the motivation needed for including the derived formalism in the ever expanding toolbox of the (time-resolved) X-ray solution scattering community. In particular, the advancement towards extracting from experiment quantum and ensemble effects on structure and dynamics. Fu-

ture theoretical developments will include further explorations on how to invert[31, 45, 49] the presented equations to obtain, directly from experimental signals, nuclear and electronic probability distributions.

## **Acknowledgements**

The authors thank R. Hartsock for discussions and feedback, and the reviewers for insightful comments. This work was funded by Lundbeck Foundation and the Danish Council for Independent Research.

## References

- [1] H. Ihee, M. Wulff, J. Kim, and S. Adachi, *International Reviews in Physical Chemistry* **29**, 453 (2010).
- [2] M. Haldrup, G. Vanko, W. Gawelda, A. Galler, G. Doumy, A. March, E. Kanter, A. Bordage, A. Dohn, T. Brandt van Driel, et al., *Journal of Physical Chemistry A* **116**, 9878 (2012).
- [3] K. Kim, J. KIM, J. Lee, and H. Ihee, *Structural Dynamics* **1**, 011301 (1986).
- [4] G. Vankó, A. Bordage, M. Pápai, K. Haldrup, P. Glatzel, A. March, G. Doumy, A. Britz, A. Galler, T. Asefa, et al., *Journal of Physical Chemistry C* **119**, 5888 (2015).
- [5] J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics* (Wiley, 2011), 2nd ed.
- [6] <https://code.google.com/p/debyer/>, accessed: June 1st, 2015.
- [7] W. Humphrey, A. Dalke, and K. Schulten, *Journal of Molecular Graphics* **14**, 33 (1996).
- [8] B. G. Levine, J. E. Stone, and A. Kohlmeyer, *Journal of Computational Physics* **230**, 3556 (2011).
- [9] F. Zernike and J. A. Prins, *Zeitschrift Für Physik* **41**, 184 (1927).
- [10] D. A. McQuarrie, *Statistical Mechanics* (Harper & Row, 1976).
- [11] N. S. Gingrich and B. E. Warren, *Physical Review* **46**, 248 (1934).
- [12] B. E. Warren, *X-Ray Diffraction* (Dover Publications, 1990).
- [13] N. S. Gingrich, *Reviews of Modern Physics* **15**, 90 (1943).
- [14] H. P. Klug and L. E. Alexander, *X-ray Diffraction Procedures: For Polycrystalline and Amorphous Materials* (Wiley, 1974), ISBN 978-0-471-49369-3, 2nd ed.
- [15] R. F. Kruh, *Chemical Review* **62**, 319 (1962).
- [16] A. H. Narten and H. A. Levy, *The Physics and Physical Chemistry of Water* **1**, 311 (1972).
- [17] J. M. Sorenson, G. Hura, R. M. Glaeser, and T. Head-Gordon, *Journal of Chemical Physics* **113**, 9149 (2000).
- [18] T. Head-Gordon and G. Hura, *Chemical Reviews* **102**, 2651 (2002).



- [19] T. K. Kim, M. Lorenc, J. H. Lee, M. L. Russo, J. Kim, M. Cammarata, Q. Kong, S. Noel, A. Plech, M. Wulff, et al., *Proceedings of the National Academy of Sciences* **103**, 9410 (2006).
- [20] Z. Lin and L. V. Zhigilei, *Physical Review B* p. 184113 (2006).
- [21] G. Gutiérrez and B. Johansson, *Physical Review B* **65**, 104202 (2002).
- [22] T. J. Penfold, I. Tavernelli, R. Abela, M. Chergui, and U. Rothlisberger, *New Journal of Physics* **14**, 113002 (2012).
- [23] J. Köfinger and G. Hummer, *Physical Review E* **87**, 052712 (2013).
- [24] K. B. Møller and N. E. Henriksen, *Structure and Bonding* **142**, 185 (2012).
- [25] F. Livet, *Acta Cryst. A* **63**, 87 (2006).
- [26] P. Coppens, *Annual Reviews of Physical Chemistry* **43**, 663 (1992).
- [27] T. Northey, N. Zotev, and A. Kirrander, *Journal of Chemical Theory and Computation* **10**, 4911 (2014).
- [28] K. Haldrup, M. Christensen, and M. M. Nielsen, *Acta Crystallographica A* **66**, 261 (2010).
- [29] K. Haldrup and M. Nielsen, *Measuring and understanding ultrafast phenomena using X-rays* (Springer Science+Business Media B.V., 2014), pp. 91–113, NATO Science for Peace and Security Series A: Chemistry and Biology, ISBN 978-94-017-8549-5.
- [30] N. E. Henriksen and K. B. Møller, *Journal of Physical Chemistry B* **112**, 558 (2008).
- [31] U. Lorenz, K. B. Møller, and N. E. Henriksen, *New Journal of Physics* **12**, 113022 (2010).
- [32] J. S. Baskin and A. H. Zewail, *Chemphyschem* **6**, 226 (2005).
- [33] D. T. Cromer and J. B. Mann, *Acta Crystallographica A* **24**, 321 (1968).
- [34] G. Grübel and F. Zontone, *Journal of Alloys and Compounds* **362**, 3 (2004).
- [35] S. Dierker, in *Light Scattering and Photon Correlation Spectroscopy*, edited by E. Pike and J. Abbiss (Springer Netherlands, 1997), vol. 40 of *NATO ASI Series*, pp. 65–78, ISBN 978-94-010-6355-5.
- [36] D. L. Abernathy, G. Grübel, S. Brauer, I. McNulty, G. B. Stephenson, S. G. J. Mochrie, A. R. Sandy, N. Mulders, and M. Sutton, *Journal of Synchrotron Radiation* **5**, 37 (1998).

- [37] G. Beutier, A. Marty, F. Livet, G. van der Laan, S. Stanescu, and P. Ben-  
cok, *Review of Scientific Instruments* **78**, 093901 (2007).
- [38] K. A. Nugent, *Advances in Physics* **59**, 1 (2009).
- [39] S. O. Hruszkewycz, M. Sutton, P. H. Fuoss, B. Adams, S. Rosenkranz,  
K. F. L. Jr., W. Roseker, D. Fritz, M. Cammarata, D. Zhu, et al., *Physical  
Review Letters* **109**, 185502 (2012).
- [40] A. O. Dohn, E. O. Jónsson, K. S. Kjær, T. B. van Driel, M. M. Nielsen,  
K. W. Jacobsen, N. E. Henriksen, and K. B. Møller, *Journal of Physical  
Chemistry Letters* **5**, 2414 (2014).
- [41] S. E. Canton, K. S. Kjær, G. Vankó, T. B. van Driel, S.-i. Adachi, A. Bor-  
dage, C. Bressler, P. Chabera, M. Christensen, A. O. Dohn, et al., *Nature  
Communications* **6**, 6359 (2015).
- [42] A. H. Narten, C. G. Venkatesh, and S. A. Rice, *Journal of Chemical  
Physics* **64**, 1106 (1976).
- [43] P. F. Peterson, E. S. Božin, T. Proffen, and S. J. L. Billinge, *Applied  
Crystallography* **36**, 53 (2003).
- [44] J. H. Lee, K. H. Kim, T. K. Kim, Y. Lee, and H. Ihee, *Journal of Chemical  
Physics* **125**, 174504 (2006).
- [45] L. B. Skinner, C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nils-  
son, and C. J. Benmore, *Journal of Chemical Physics* **138**, 074506 (2013).
- [46] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L.  
Hura, and T. Head-Gordon, *Journal of Chemical Physics* **120**, 9665  
(2004).
- [47] G. Hura, J. M. Sorenson, R. M. Glaeser, and T. Head-Gordon, *Journal  
of Chemical Physics* **113**, 9140 (2000).
- [48] G. L. Hura, D. Russo, M. Glaeser, T. Head-Gordon, M. Krack, and  
M. Parrinello, *Physical Chemistry, Chemical Physics* **5**, 1981 (2003).
- [49] Q. Kong, J. Kim, M. Lorenc, T. K. Kim, H. Ihee, and M. Wulff, *Journal  
of Physical Chemistry A* **2005**, 109 (10451-10458).