Acta Cryst A FOUNDATIONS ADVANCES
ISSN 2053-2733

Site-occupancy factors in the Debye scattering equation. A theoretical discussion on significance and correctness

Fabio Ferri,[a]* Maria Chiara Bossuto,[a] Pietro Anzini,[a] Antonio Cervellino,[b] Antonietta Guagliardi,[c] Federica Bertolotti[a] and Norberto Masciocchi[a]*

[a]Dipartimento di Scienza e Alta Tecnologia & To.Sca.Lab, Università degli Studi dell'Insubria, via Valleggio 11, Como, 22100, Italy, [b]Swiss Light Source, Paul Scherrer Institute, Villigen PSI, 5232, Switzerland, and [c]Istituto di Cristallografia (IC) & To.Sca.Lab, Consiglio Nazionale delle Ricerche (CNR), via Valleggio 11, Como, 22100, Italy. *Correspondence e-mail: fabio.ferri@uninsubria.it, norberto.masciocchi@uninsubria.it

Received 27 June 2023
Accepted 26 September 2023

Edited by L. Palatinus, Czech Academy of Sciences, Czech Republic

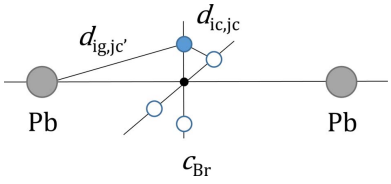
Keywords: Debye scattering equation; site-occupancy factors; defective nanocrystals.

The Debye scattering equation (DSE) [Debye (1915). *Ann. Phys.* **351**, 809–823] is widely used for analyzing total scattering data of nanocrystalline materials in reciprocal space. In its modified form (MDSE) [Cervellino *et al.* (2010). *J. Appl. Cryst.* **43**, 1543–1547], it includes contributions from uncorrelated thermal agitation terms and, for defective crystalline nanoparticles (NPs), average site-occupancy factors (s.o.f.'s). The s.o.f.'s were introduced heuristically and no theoretical demonstration was provided. This paper presents in detail such a demonstration, corrects a glitch present in the original MDSE, and discusses the s.o.f.'s physical significance. Three new MDSE expressions are given that refer to distinct defective NP ensembles characterized by: (i) vacant sites with uncorrelated constant site-occupancy probability; (ii) vacant sites with a fixed number of randomly distributed atoms; (iii) self-excluding (disordered) positional sites. For all these cases, beneficial aspects and shortcomings of introducing s.o.f.'s as free refinable parameters are demonstrated. The theoretical analysis is supported by numerical simulations performed by comparing the corrected MDSE profiles and the ones based on atomistic modeling of a large number of NPs, satisfying the structural conditions described in (i)–(iii).

1. Introduction

Powder diffraction (PD) is a powerful tool that has been used for decades in many fields of science and technology (chemistry, materials science, metallurgy, geology, forensic science, cultural heritage) and is nowadays commonplace in both academic and industrial sectors. Thanks to the availability of high-quality data and the development of advanced computational tools, PD is widely used for studying not only polycrystalline materials with large crystalline domains and sharp Bragg peaks, but also defective and/or nano-sized materials where both (broadened) Bragg peaks and diffuse scattering present in between the peaks carry relevant information about the material structure and defectiveness, and size and shape of nanocrystals. The analysis of the total (peaks + diffuse) elastic scattering data is customarily carried out in the reciprocal space by using the Debye scattering equation (DSE) or in the real space by recovering the pair distribution function (PDF) via a Fourier transform of the scattering data. Recent reviews on the two complementary techniques have been carried out by Scardi & Gelisio (2016), Billinge (2019) and Cervellino *et al.* (2016).

During the last decade, the DSE approach was further refined by introducing the effects of atomic thermal motion



and partial site-occupancy factors (s.o.f.'s). As a result, a modified version of the DSE (MDSE) was proposed and implemented in the open-source suite of programs known as *DEBUSSY* (Cervellino *et al.*, 2010, 2015). The introduction of s.o.f.'s in the MDSE allowed us to compute (exactly) the average scattered intensity $I(Q)$ of an ensemble of defective nanoparticles (NPs) with randomly distributed vacancies in a very efficient way. This task would have been much more time consuming (and occasionally less accurate) if the computation had been carried out by averaging the $I(Q)$'s of many defective NPs with the vacant sites omitted. Thus, the MDSE allows the usage of s.o.f.'s as free adjustable parameters with an efficient fitting procedure, which is fast and highly accurate from a statistical point of view. The MDSE has been intensively adopted for the characterization of many nano-sized and defective species, including metals, oxides, ionic and organometallic species, up to colloids (Bertolotti *et al.*, 2018).

However, in spite of its wide and successful use, the MDSE was never validated from a theoretical point of view regarding the atomic s.o.f. parameters, which were introduced heuristically, and no theoretical demonstration was ever provided. In this paper we detail such a demonstration, correct a glitch present in the original MDSE, and discuss the s.o.f. physical significance, providing three new MDSE expressions that refer to NP defectiveness characterized by vacant sites with: (i) uncorrelated constant site-occupancy probability, or (ii) a fixed number of randomly distributed atoms, or (iii) self-excluding disordered positional sites (split-atom model). For all these cases, we provide both theoretical and numerical evidence of the effectiveness of introducing s.o.f.'s as free refinable parameters by comparing the new MDSE profiles and the ones based on atomistic modeling of a large number of NPs, satisfying the structural conditions described in (i)–(iii). Based on the theoretical approach here presented, we further correct the original MDSE, showing that the s.o.f.'s appearing in the self-term summation [$I_{\rm self}(Q)$, see below] of the corrected MDSE should not be squared. This minor change only affects the smooth continuous behavior of $I_{\rm self}(Q)$ and does not modify the $I_{\rm dist}(Q)$ term of the original MDSE, where the structural NP information is encoded.

2. Theoretical background: the DSE

The DSE proposed by Debye (1915) describes the elastic scattering intensity distribution of randomly oriented monodisperse (*i.e.* of equal size, morphology and structure) non-interacting NPs, each composed of n atoms whose fixed positions in an arbitrary reference frame are known (regardless of any periodicity and order). The DSE reads

$$I(Q) = \sum_{i=1}^{n} \left|f_i(Q)\right|^2 + \sum_{i\neq j=1}^{n} f_i(Q) f_j^*(Q)\,\mathrm{sinc}\left(Qd_{ij}\right) = \sum_{i=1}^{n} \left|f_i(Q)\right|^2 + 2\sum_{i>j=1}^{n} \mathrm{Re}\left\{f_i(Q) f_j^*(Q)\right\}\mathrm{sinc}\left(Qd_{ij}\right), \quad (1)$$

where $\mathrm{sinc}(x) = \sin(x)/x$, $Q = (4\pi/\lambda)\sin\theta$ is the magnitude of the scattering vector, θ is half of the scattering (*i.e.* deflection) angle, λ is the radiation wavelength, d_{ij} is the Euclidean distance between atoms i and j, and $f_i(Q)$ is the X-ray atomic form factor of the ith atom. Note that the atomic form factors are complex functions whereas, as shown by the second line of equation (1), the overall intensity is a real quantity. In the case of X-ray radiation, $f(Q) = f^0(Q) + f' + if''$, where $f^0(Q)$ is the (real) elastic scattering term, and f' and f'' are the real and imaginary parts, respectively, of the anomalous scattering contribution due to atomic electron binding, which is Q independent but varies with radiation energy E. The behaviors of elemental f^0, f' and f'' as a function of Q and E can be found in the work of Cullen *et al.* (1989). As a historical remark, we point out that in the original DSE form, the form factors were considered real because at that time (1915) X-ray anomalous scattering effects were not known [first shown by Mark & Szilard (1925)].

The first term of equation (1), which corresponds to the *self-scattered* intensity $I_{\rm self}(Q)$ $(i = j)$, is given by the sum of the intensities scattered from all the atoms making up the NP; the second term, which is the *distinct-scattered* intensity $I_{\rm dist}(Q)$ $(i \neq j)$, takes into account the interference between all pairs of distinct atoms within the NP. Therefore, $I_{\rm dist}(Q)$ depends on the relative arrangement of the atoms and provides information on the structural features of the NP.

3. The modified Debye scattering equation

There are two main limitations of equation (1), which have prompted the development of a modified Debye scattering equation (MDSE).

The first one is the assumption of fixed atomic positions, thus neglecting the effects of unavoidable thermal vibrations. As done in conventional crystallography, the latter are usually taken into account by multiplying the form factors present in the distinct term of equation (1) by the Debye–Waller (DW) thermal factors (Warren, 1990) associated, in a simplified picture, to each atomic species s present in the NP. The atomic sites are then understood as the spatial averages of the vibrating atomic positions (equilibrium positions). In the case of isotropic and independent atomic vibrations, the DW factors associated to the sth species are given by $T_s(Q) = \exp[-(B_s/16\pi^2)Q^2]$, where the thermal factor B_s is related to the mean-square displacement $\langle u_s^2\rangle$ of all the atoms belonging to the s species about their equilibrium position by the relation $B_s = 8\pi^2\langle u_s^2\rangle$.

The second limitation of equation (1) is the difficulty of properly dealing with defective NP ensembles characterized by crystalline order but partial s.o.f.'s, where not all the sites that are present in the corresponding non-defective NPs are occupied by atoms. Thus, in spite of the assumption of identical NPs, the set of actual distances in equation (1) are not the same for all the NPs of the ensemble because, for each defective NP, the atoms are randomly distributed among the available sites. The observed profile $I(Q)$ of such a system corresponds to the ensemble average of the intensities $I_k(Q)$ scattered by (many) defective, randomly oriented NPs. If we

indicate with n_k the actual number of atoms inside the kth NP of the ensemble, we can use equation (1) and write

$$I_k(Q) = \sum_{i=1}^{n_k} \left|f_i^k\right|^2 + \sum_{i\neq j=1}^{n_k} f_i^k f_j^{k^*} T_i^k T_j^k \,\mathrm{sinc}\left(Qd_{ij}^k\right), \qquad (2)$$

where we have introduced the DW thermal factors and, to help readability, we have omitted the Q dependence of the form factors and thermal factors.

Let us assume that all the NPs of the ensemble are characterized by the same available sites labeled with the indexes i and j ($i, j = 1, \ldots, M$), and that each site i is assigned to a single atomic species through a function $s(i)$. If we define a binary variable ω_i^k, that assumes the values $\omega_i^k = 1$ for occupied sites and $\omega_i^k = 0$ for vacant sites, we can re-write equation (2) as

$$I_k(Q) = \sum_{i=1}^{M} \left(\omega_i^k\right)^2 \left|f_i\right|^2 + \sum_{i\neq j=1}^{M} \omega_i^k \omega_j^k f_i f_j^* T_i T_j \,\mathrm{sinc}\left(Qd_{ij}\right). \qquad (3)$$

Note that, though f_i and T_i depend only on the species, to keep the notation simple, we have indicated only their (implicit) dependence on i and j, meaning that $f_i \equiv f_{s(i)}$ and $T_i \equiv T_{s(i)}$.

Since $I_k(Q)$ describes the intensity profile of the defective k-NP, the variables ω_i^k are intended as a single realization of the stochastic variables ω_i that assume the values $\omega_i = 1$ with probability o_s [which depends only on the species $s(i)$] and $\omega_i = 0$ with probability $1 - o_s$. Note that for non-defective NPs (all sites occupied) the variables ω_i become deterministic with constant values $\omega_i = 1$ and correspondingly $o_s = 1$. The average of $I_k(Q)$ over the entire defective NP ensemble reads

$$\begin{aligned} I(Q) &\equiv \langle I_k(Q)\rangle \\ &= \sum_{i=1}^{M} \langle\omega_i^2\rangle \left|f_i\right|^2 + \sum_{i\neq j=1}^{M} \langle\omega_i\omega_j\rangle f_i f_j^* T_i T_j \,\mathrm{sinc}\left(Qd_{ij}\right), \end{aligned} \qquad (4)$$

where $\langle\cdot\rangle$ indicates the ensemble average. Equation (4) shows that $I(Q)$ depends on the second moment $\langle\omega_i^2\rangle$ and on the second-order correlation $\langle\omega_i\omega_j\rangle$, which, in turn, are related to the type of defectiveness associated to the NP ensemble. In the following, starting from equation (4), we will work out the analytical expression for $I(Q)$ for three specific cases of NP defectiveness.

3.1. Defective NPs with uncorrelated constant occupancy probabilities

Let us consider a defective NP ensemble where all the s.o.f.'s of a given atomic species s are characterized by the same *constant occupancy probability* o_s. Thus, the probability that any site i assigned to the species $s(i)$ is occupied is equal to o_s, whereas $(1 - o_s)$ is the probability that the site is vacant. Let us further suppose there is no spatial correlation between the occupancies of the different sites. As a consequence of these assumptions, the number of atoms n_s belonging to the s species present among the M_s sites available for that species is not constant, but is a stochastic variable described by a binomial distribution

$$B(o_s, M_s)(n_s) = \binom{M_s}{n_s} o_s^{n_s} (1 - o_s)^{M_s - n_s}, \qquad (5)$$

where $\binom{M_s}{n_s} = M_s/[n_s!(M_s - n_s)!]$ is the binomial coefficient. For such a binomial distribution, the average number of atoms is $\langle n_s\rangle = o_s M_s$ and the variance is $\sigma_{n_s}^2 = \langle n_s\rangle(1 - o_s)$.

Under these hypotheses and by using the property that, for a binary variable ω_i any moment $\langle\omega_i^m\rangle = \langle\omega_i\rangle = o_i$ ($m > 0$), the average quantities inside equation (4) read

$$\begin{cases} \langle\omega_i^2\rangle = o_i & (6a) \\ \langle\omega_i\omega_j\rangle = \langle\omega_i\rangle\langle\omega_j\rangle = o_i o_j, & (6b) \end{cases}$$

where o_i represents the occupancy probability of the ith site, which depends only on the atomic species $s(i)$, and we have adopted the simplifying notation $o_i \equiv o_{s(i)}$.

Upon substitution of equations (6) into equation (4), we get

$$I(Q) = \sum_{i}^{M} o_i \left|f_i\right|^2 + \sum_{i\neq j}^{M} o_i o_j f_i f_j^* T_i T_j \,\mathrm{sinc}\left(Qd_{ij}\right), \qquad (7)$$

where $M = \sum_{s=1}^{S} M_s$, S being the number of different species. Equation (7) represents a modified version of the original DSE (MDSE), in which both the DW thermal factors and the s.o.f.'s have been included. Clearly, when all the species are characterized by unitary occupancies ($M_s = n_s$), the total number of available sites is equal to the total number of atoms ($M = n$) and aside from the DW factors, equation (7) reduces to equation (1).

It should be pointed out that a different MDSE version was proposed (and implemented in the software package *DEBUSSY*) by Cervellino *et al.* (2010). That version of the MDSE contains a glitch in the $I_{\mathrm{self}}(Q)$ term, where the s.o.f.'s are (erroneously) squared. However, this error only affects the continuous, monotonic behavior of $I_{\mathrm{self}}(Q)$, without modifying the $I_{\mathrm{dist}}(Q)$ term, where all the structural features are encoded. Therefore, the presence of this glitch introduces only negligible effects on the NP structural characterization and, as will be shown in Section 5.1, it does not significantly affect any of the previously published results, where the total scattering data were interpreted according to the incorrect MDSE equation. A quantitative comparison between the original MDSE and our corrected version [equation (7)] is reported in Sections 4.1 and 4.2.

3.2. Defective NPs with a constant number of atoms

Let us consider a defective NP ensemble where each NP is characterized, species by species, by a *constant number* n_s of atoms randomly distributed among the M_s available sites ($n_s \leq M_s$). Thus, for all the sites i assigned to the species s, $\langle\omega_i\rangle = (n_s/M_s)$, but differently from Section 3.1, the assumption of independent occupancies remains valid only for sites belonging to different species. Since the average $\langle\omega_i\omega_j\rangle$ is equal to the probability of having $\omega_i\omega_j = 1$ (which occurs only when $\omega_i = \omega_j = 1$), for sites of the same species we have

$$\langle \omega_i \omega_j \rangle = \frac{\binom{M_s - 2}{n_s - 2}}{\binom{M_s}{n_s}} \quad (i \neq j) \text{ and } [s(i) = s(j)], \tag{8}$$

where the binomial coefficient in the denominator represents the number of combinations that can be realized by placing n_s atoms within M_s sites, and the one in the numerator is the number of combinations left available after the ith and jth sites have been occupied by two atoms, which occurs only when $\omega_i = \omega_j = 1$. By developing the two binomial coefficients of equation (8), and adopting the same simplified notation used in equations (6) [namely $n_i \equiv n_{s(i)}$ and $M_i \equiv M_{s(i)}$], we end up with the result

$$\begin{cases} \langle \omega_i^2 \rangle = \frac{n_i}{M_i} & (9a) \\ \langle \omega_i \omega_j \rangle = \langle \omega_i \rangle \langle \omega_j \rangle = \frac{n_i n_j}{M_i M_j} \quad (i \neq j) \text{ and } [s(i) \neq s(j)] & (9b) \\ \langle \omega_i \omega_j \rangle = \left(\frac{n_i}{M_i}\right)\left(\frac{n_j - 1}{M_j - 1}\right) \quad (i \neq j) \text{ and } [s(i) = s(j)]. & (9c) \end{cases}$$

Provided that n_i is equal to the average number of atoms of a defective NP ensemble with constant occupancy probability ($n_i = \langle n_i \rangle$), equations (9) are quite similar to equations (6). The only slight difference is in the right-hand term of equation (9c) which is (marginally) smaller than in equation (6b), implying that the constraint of constant number of atoms introduces a slightly negative (constant) correlation among the occupancies of the same atomic species. Indeed, since $[(n_i - 1)/(M_i - 1)] < (n_i/M_i)$, the covariance $\mathrm{Cov}(\omega_i, \omega_j) = \langle \omega_i \omega_j \rangle - \langle \omega_i \rangle^2$ associated to equation (9c) is slightly negative. By inserting equations (9) into equation (4), we get

$$I(Q) = \sum_{i=1}^{M} \frac{n_i}{M_i} |f_i|^2 + \sum_{\substack{i \neq j \\ s(i) \neq s(j)}}^{M} \left(\frac{n_i}{M_i}\right)\left(\frac{n_j}{M_j}\right) f_i f_j^* T_i T_j \operatorname{sinc}\left(Q d_{ij}\right) + \sum_{\substack{i \neq j \\ s(i) = s(j)}}^{M} \left(\frac{n_i}{M_i}\right)\left(\frac{n_j - 1}{M_j - 1}\right) f_i f_j^* T_i T_j \operatorname{sinc}\left(Q d_{ij}\right), \tag{10}$$

where n_i and M_i represent the number of atoms and sites of the $s(i)$ species, respectively, and M is the total number of sites. Under realistic conditions, *i.e.* for nano-sized NPs with n_i, $M_i \gg 1$, equation (10) is almost identical to equation (7) because $(n_i/M_i) = o_i$ and $(n_i/M_i)(n_j - 1/M_j - 1) \cong o_i o_j$. We anticipate here that the difference between the two equations (and therefore between the two types of defectiveness) shows up only in the small-angle X-ray scattering (SAXS) region where the influence of the number of fluctuations present in equation (7) introduces an extra contribution, which is absent in equation (10). A quantitative comparison between these two equations is deferred to Section 5.2.

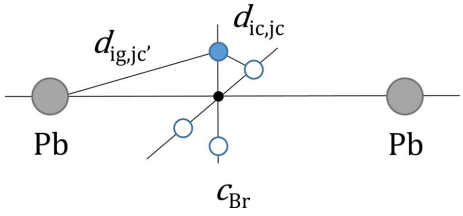


Figure 1
Schematic detail of a locally disordered APbBr$_3$ perovskite cluster where the Br atom, instead of occupying the highly symmetric position midway between the two Pb atoms (small black circle), is randomly located at one of the four split locations belonging to the c_{Br} cluster, that lie on the plane normal to the Pb···Pb vector.

3.3. Defective NPs with self-excluding positional sites (split-atom model)

In this example we report a special case of correlated occupancies, where the constituent sites belonging to a given atomic species s can be split into different subsets (or clusters), each of them characterized by a given number C_s of (usually symmetry-related) crystallographic site positions. Since all the distances between the split site positions of a given cluster are typically smaller than a physically meaningful minimal distance (for example, the atom size), only one atom can be randomly placed within the C_s self-excluding crystallographic sites of that species ('split-atom' model).

As an example of where a subset of sites can host one atom only (of a single species), we discuss here the so-called 'split-cubic' model (Mashiyama *et al.*, 1998) that was proposed for disordered lead halide perovskites (APbX_3, A = a monovalent cation, X = Cl, Br, I). Here, considering the case A = Cs, X = Br, the Br atom joining two Pb atoms occupies one of the four different positions that, with respect to its ideal location (midway along the Pb···Pb vector, taken as the [100] edge of the unit cell aligned with x, with a Pb—Br—Pb angle of 180°), are shifted up–down–left–right in the (002) plane by some 0.3–0.4 Å in the $\pm y$ or $\pm z$ directions. This situation, clearly visible in Fig. 2 of Mashiyama *et al.* (1998), is sketched in Fig. 1 where three of the four Br sites are (*must be*) vacant (white circles), whereas only one site (blue circle) contains (*must contain*) a Br atom. Other examples of self-excluding positional disorder can be found in the literature, either with smaller [$C_s = 2$, as in hexagonal hydroxyapatite (Leeuw, 2001)] or larger [$C_s = 26$, as in another Pb-containing perovskite (Cervellino *et al.*, 2011)] multiplicities.

In our example, $C_{Br} = 4$, whereas $C_{Pb} = C_{Cs} = 1$, meaning that the Pb and Cs sites are not split. Thus, the site occupancies are uncorrelated only when different clusters are considered. Conversely, the occupancies are fully (anti)correlated for atoms within the same cluster (in the example of Fig. 1, the Br cluster, c_{Br}), because only one site of the cluster is occupied.

As in the previous cases, let us indicate with o_i the occupancy probability of the ith site assigned to the $s(i)$ species. Thus, if we assume that all the NPs of the ensemble have a constant number of atoms and there is no spatial correlation between the occupancies of sites belonging to different clusters, we can immediately write

$$\begin{cases} \langle \omega_i^2 \rangle = o_i & (11a) \\ \langle \omega_i \omega_j \rangle = 0 \quad (i \neq j)_{\text{intra-cluster}} & (11b) \\ \langle \omega_i \omega_j \rangle = \langle \omega_i \rangle \langle \omega_j \rangle = o_i o_j \quad (i \neq j)_{\text{inter-cluster}}, & (11c) \end{cases}$$

where $o_i = 1/C_{s(i)}$. Note that the number of atoms of the unsplit species is equal to the number of corresponding sites ($n_{Cs} = M_{Cs}$ and $n_{Pb} = M_{Pb}$), whereas for the split species

$n_{Br} = M_{Br}/4$, implying that $o_{Cs} = o_{Pb} = 1$ and $o_{Br} = 1/4$. Although equations (11) refer to NPs with a constant number of atoms (as in Section 3.2) and unitary occupancies for unsplit species, its generalization to partial uncorrelated occupancies is rather straightforward.

Upon substitution of equations (11) into equation (4), we obtain

$$I(Q) = \sum_{i=1}^{M} o_i \left|f_i(Q)\right|^2 + \sum_{(i\neq j)_{\rm inter}}^{M} o_i o_j f_i f_j^* T_i T_j \,\mathrm{sinc}\left(Q d_{ij}\right), \quad (12)$$

where the double sum of the distinct term is restricted only to the inter-cluster distances. Note that, except for the restricted sum of the distinct term, equation (12) is formally identical to equation (7), thus extending the application of the (correct) MDSE from defective nanocrystals with vacant sites *only*, to those with self-excluding disordered positional sites.

Another way of recasting equation (12) is by using the Heaviside Θ function, defined as $\Theta(x) = 1$ for $x > 0$, $\Theta(x) = 0$ for $x \leq 0$. This version is computationally efficient and is the one actually implemented (apart from the correction to the self-scattering term) in the *DEBUSSY* software suite,

$$I(Q) = \sum_{i=1}^{M} o_i \left|f_i(Q)\right|^2 + \sum_{i\neq j}^{M} o_i o_j f_i f_j^* T_i T_j \,\Theta\left(d_{ij} - d_{i,j}^{\rm min}\right) \mathrm{sinc}\left(Q d_{ij}\right), \quad (13)$$

where $d_{i,j}^{\rm min}$ is the minimal distance allowed between atoms of species $s(i)$ and $s(j)$ and, as above, we have adopted the simplifying notation $d_{i,j}^{\rm min} \equiv d_{s(i),s(j)}^{\rm min}$. DSE computation through equation (13) is faster than using equation (12) because it does not need a pre-classification of atoms into clusters, just a look-up table for the possible minimal bond distances for each atomic species pair. Then all pairs with distances shorter than $d_{i,j}^{\rm min}$ are simply ignored.

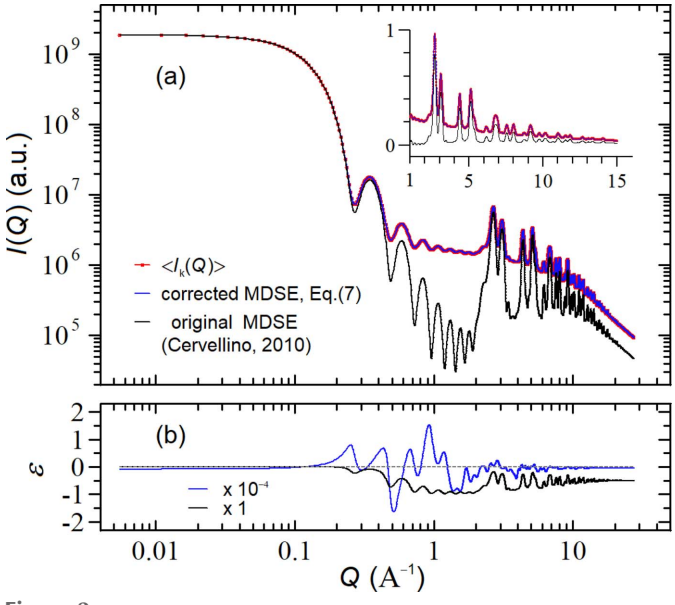


Figure 2
(*a*) Comparison between the $I(Q)$ profiles computed with the corrected MDSE [equation (7), blue curve] and the original MDSE [(Cervellino *et al.*, 2010), black curve] against the expected $\langle I_k(Q)\rangle$ obtained by averaging the intensity profiles of many defective NCs in the case of a f.c.c. Au NC (a = 4.08 Å) made of 6 × 6 × 6 unit cells (side L = 2.45 nm) with an average occupancy factor $o = 0.5$; in the inset the same (rescaled) data are shown on a (reduced range) linear plot. (*b*) Relative deviations ε between the corrected and the original MDSE against the expected $\langle I_k(Q)\rangle$ profiles, $\varepsilon = [I(Q) - \langle I_k(Q)\rangle]/\langle I_k(Q)\rangle$. The factors appearing in the legend indicate the absolute scales of the y-axis values.

4. Numerical simulations

To crosscheck the validity of the analysis outlined in the previous section, we carried out several numerical simulations corresponding to single- and multi-species defective NP ensembles selected as case studies. In the following, we report three examples of such numerical simulations. Owing to computational issues intrinsic to the DSE calculation from scratch, our analysis has been limited to ultra-small NPs, without any lack of generality.

4.1. Case 1: elemental f.c.c. Au nanocrystals with constant vacant site probability

For the first example, we selected an ideal monoatomic nanocrystal (NC), namely a f.c.c. (face-centered cubic) Au NC (a = 4.080 Å), in the form of a cube with six unit cells per edge, characterized by M = 1099 sites, side L = 2.45 nm and equivalent diameter (of the sphere of equal volume) $d_{\rm eqv}$ = 3.04 nm. Then, we generated many defective NCs by filling the M sites with n_k atoms that were randomly spread among the sites. The numbers n_k were extracted stochastically from a binomial distribution $B(o, M)(n)$ with a given occupancy probability o. To emphasize the differences between the original and corrected MDSE, we intentionally selected a fairly small (and unrealistic) occupancy probability, *i.e.* $o = 0.5$. Then, for each defective NC, we computed $I_k(Q)$ by using equation (2) (with B = 0.5 Å^2 or $\sqrt{\langle u^2\rangle}$ = 7.96 × 10^{-2} Å) and averaged such a profile over a large number $N_{\rm ave} \sim 2 \times 10^6$ of NCs until the average profile $\langle I_k(Q)\rangle$ was determined with very high accuracy $R_{\rm acc} = \sqrt{(1/P)\sum_{i=1}^{P}(\sigma_i/\langle I\rangle_i)^2} = 2.6 \times 10^{-5}$, where the sum runs over all the Q_i points ($i = 1, \ldots, P$) of the intensity profile and σ_i is the standard deviation associated to $\langle I\rangle_i$.

Fig. 2(*a*) compares $\langle I_k(Q)\rangle$ (red symbols) with the profiles $I(Q)$ computed by using the original MDSE [(Cervellino *et al.*, 2010), black curve] and the corrected MDSE [equation (7), blue curve]. The same (rescaled) profiles are shown in the inset of Fig. 2(*a*) on a linear plot. As is evident, the original MDSE is not capable of reproducing $\langle I_k(Q)\rangle$ over most of the Q range. Conversely, the match between the $I(Q)_{\rm corr}$ profile calculated by using the corrected MDSE and $\langle I_k(Q)\rangle$ is almost perfect [not distinguishable in Fig. 2(*a*)], proving the correctness of equation (7). The small relative deviations ε between these two curves $\{\varepsilon = [I(Q)_{\rm corr} - \langle I_k(Q)\rangle]/\langle I_k(Q)\rangle\}$, shown in Fig. 2(*b*), can be altogether quantified by the (dis)agreement parameter

$$R = \sqrt{\frac{1}{P}\sum_{i=1}^{P}\left[\frac{I(Q_i)_{\rm corr} - \langle I_k(Q_i)\rangle}{\langle I_k(Q_i)\rangle}\right]^2}, \qquad (14)$$

which, in the case of the blue and red curves of Fig. 2(*a*), is $R = 1.7 \times 10^{-5}$, a figure that is consistent with the accuracy associated to $\langle I_k(Q)\rangle$.

Interestingly, Fig. 2(*a*) shows that at large Q's there is a factor ∼2 between the corrected and original MDSE. This occurs because, for $Q \to \infty$, the main contribution to $I(Q)$ comes from $I_{\rm self}(Q)$ and the ratio $[I_{\rm self}(Q)]_{\rm original}/[I_{\rm self}(Q)]_{\rm corrected} = o = 0.5$. Conversely, at small Q's, the original and corrected MDSE are almost superimposed, consistent with the fact that, for $Q \to 0$, the main contribution to $I(Q)$ comes from the $I_{\rm dist}(Q)$ term, which is the same in the original and corrected MDSEs. As a final comment, we point out that the Bragg peaks of the original MDSE appear to be much more pronounced than those of the corrected one due to the (erroneously) reduced $I_{\rm self}(Q)$ contribution present in the original MDSE.

As anticipated, the presented f.c.c. Au phase with 50% of randomly positioned vacancies is clearly a non-physical system, prone to collapse into a (twice as) denser phase. However, the same data analysis approach used here can be employed to analyze randomly mixed alloys when the scattering contrast of the constituent elements is rather high. This is valid for example for any f.c.c. Ni_xV_{1-x} alloy [x = 0.75 (Koester & Gmoehling, 1960); x = 0.60 (Pearson & Hume-Rothery, 1952)], if studied by neutron radiation. In this case, the coherent scattering lengths for natural abundance Ni and V nuclei are 10.4 and −0.38 fm, respectively (Sears, 1992). More practically, random nano-alloys of Au_xAg_{1-x}, largely studied for their catalytic and surface plasmonic states (Newmai *et al.*, 2022; Kozioł *et al.*, 2021; Coviello *et al.*, 2022), can also be studied by similar data analysis protocols, but this analysis is beyond the purpose of this article.

4.2. Case 2: non-stoichiometric cubic PbS nanocrystals with constant vacant site probability

In the second example, we selected a biatomic NC, namely a PbS NC (a = 5.924 Å) made of 5 × 5 × 5 unit cells, which corresponds to a cubic crystal characterized by M = 1000 sites ($M_{\rm Pb}$ = 500, $M_{\rm S}$ = 500), side L = 2.67 nm and equivalent diameter $d_{\rm eqv}$ = 3.67 nm. Then, we followed the same procedure described for the Au NC, but in this case we selected less extreme (Bertolotti *et al.*, 2016) occupancies, $o_{\rm Pb} = 1$ and $o_{\rm S} = 0.9$, and obtained $\langle I_k(Q)\rangle$ with $R_{\rm acc} = 9.8 \times 10^{-6}$ ($N_{\rm ave} = 1.4 \times 10^5$). Analogously to what was done in the previous subsection, Fig. 3(*a*) compares the simulated $\langle I_k(Q)\rangle$ (red symbols) with the $I(Q)$ computed by using the original MDSE [(Cervellino *et al.*, 2010), black curve] and with the corrected MDSE [equation (7), blue curve]. As one can easily appreciate, the three curves are almost perfectly superimposed, both on the main plot (log–log scale) and on the inset (linear scale), showing that, under these working conditions, the effects of the corrected MDSE are rather marginal. Nevertheless, the difference is evident in the residual plots of Fig. 3(*b*) near Q = 1 Å^{-1}, where $\varepsilon_{\rm MDSE_{orig}}$ and $\varepsilon_{\rm MDSE_{corr}}$ differ by more than three orders of magnitude. Correspondingly $R_{\rm MDSE_{corr}} = 8.2 \times 10^{-6}$, whereas $R_{\rm MDSE_{orig}} = 8.3 \times 10^{-3}$.

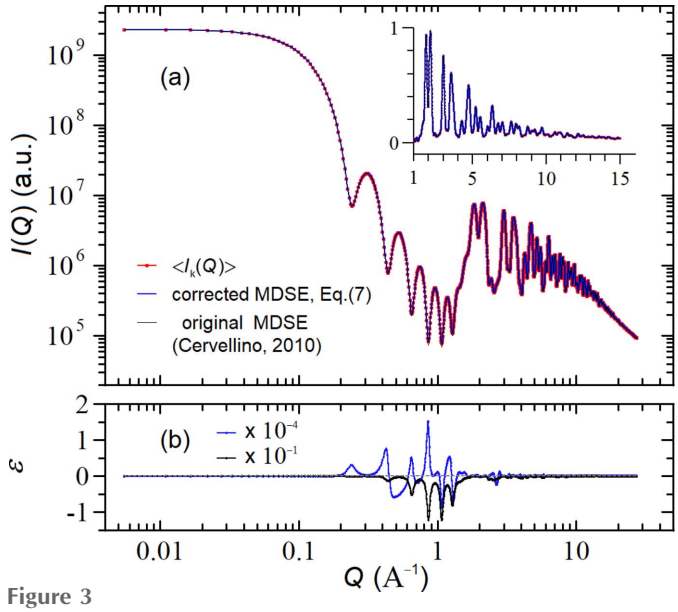


Figure 3
(*a*) Comparison between the $I(Q)$ profiles computed with the corrected MDSE [equation (7), blue curve] and the original MDSE [(Cervellino *et al.*, 2010), black curve] against the expected $\langle I_k(Q)\rangle$ obtained by averaging the intensity profiles of many defective NCs in the case of a PbS NC (a = 5.924 Å) made of 5 × 5 × 5 unit cells (L = 2.67 nm) with average occupancies $o_{\rm Pb} = 1$ and $o_{\rm S} = 0.9$; in the inset the same (rescaled) data are shown on a (reduced range) linear plot. (*b*) Relative deviations $\varepsilon_{\rm MDSE_{corr}}$ and $\varepsilon_{\rm MDSE_{orig}}$. The factors appearing in the legend indicate the absolute scales of the y-axis values.

4.3. Case 3: stoichiometric $CsPbBr_3$ perovskite with self-excluding positional sites (split-cubic model)

In the last example we considered a cubic $CsPbBr_3$ perovskite NC (a = 5.927 Å), made of 5 × 5 × 5 unit cells with a resulting side L = 2.67 nm where the sites corresponding to the Cs and Pb atoms are completely occupied ($o_{\rm Cs} = 1$, $M_{\rm Cs} = n_{\rm Cs}$ = 125 and $o_{\rm Pb} = 1$, $M_{\rm Pb} = n_{\rm Pb}$ = 125), whereas the Br atoms are randomly displaced crosswise by 0.5 Å normally from the center of the Pb···Pb vector (see Fig. 1), maintaining an average holoedric cubic symmetry with a distance $d_{i,j}^{\rm min}$ = 0.71 Å. Thus, the number of Br atoms is $n_{\rm Br} = 375$ and the corresponding self-excluding positional sites are $M_{\rm Br} = 375 \times 4 = 1500$, with occupancy $o_{\rm Br} = 1/4$. For each defective NC, we computed $I_k(Q)$ (for simplicity, with all atoms at rest, *i.e.* $B = 0$) and averaged such a profile until $\langle I_k(Q)\rangle$ was determined with $R_{\rm acc} \sim 1.5 \times 10^{-5}$ ($N_{\rm ave} \sim 1.5 \times 10^5$).

Fig. 4(*a*) compares $\langle I_k(Q)\rangle$ (red symbols) with the profile $I(Q)$ computed by using equation (12) (blue curve) where $M = M_{\rm Cs} + M_{\rm Pb} + M_{\rm Br} = 1750$. The same (rescaled) profiles are shown in the inset of Fig. 4(*a*) on a (reduced range) linear plot. As in the previous example, the almost perfect match of the two curves [indistinguishable in Fig. 4(*a*) and the corre-

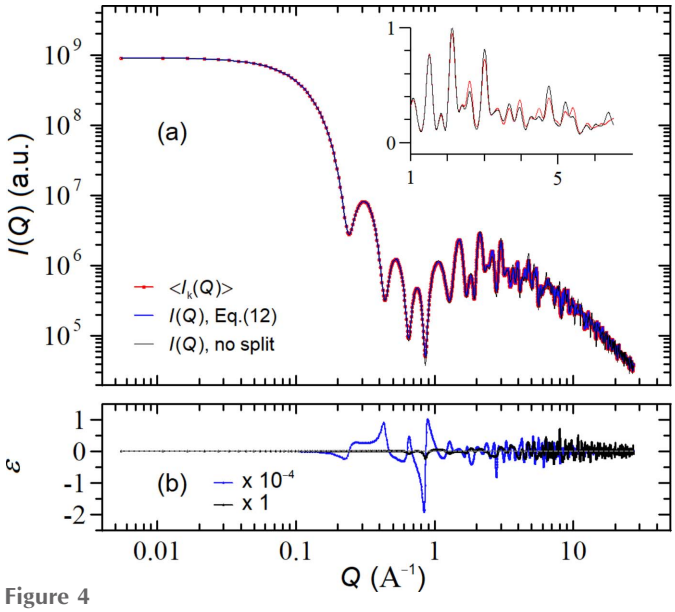


Figure 4
(*a*) Comparison between the expected $\langle I_k(Q)\rangle$ (red symbols) obtained by averaging the intensity profiles of many defective cubic $CsPbBr_3$ perovskite NCs where the Br atoms were randomly displaced crosswise (see text) and the $I(Q)$ computed via equation (12) (blue curve). For completeness, the $I(Q)$ relative to an ideal $CsPbBr_3$ NC (unsplit) is also reported (black curve); in the inset the same (rescaled) data are shown on a (reduced range) linear plot. (*b*) Relative deviations ε between $\langle I_k(Q)\rangle$ and the two $I(Q)$ curves of panel (*a*). The factors appearing in the legend indicate the absolute scales of the *y*-axis values.

sponding inset], with a small (dis)agreement parameter $R \sim 1.4 \times 10^{-5}$ and non-systematic deviations [shown in Fig. 4(*b*)], demonstrates the correctness of equation (12). In order to highlight the effects of the splitting in the Br positions, we also compare $\langle I_k(Q)\rangle$ with the $I(Q)$ relative to an ideal $CsPbBr_3$ NC, where the Br atom (small black circle in Fig. 1) is located midway along the two Pb atoms (black curve). Although the differences between $\langle I_k(Q)\rangle$ and the $I(Q)$'s relative to the two models are barely visible in Fig. 4(*a*) and in the corresponding inset, the match to $\langle I_k(Q)\rangle$ is much worse for the unsplit case [as shown in Fig. 4(*b*)], with $R \sim 1.4 \times 10^{-1}$. Worthy of note, it is exactly the difference in peak intensities [rather than in positions, see inset of Fig. 4(*a*)] appreciable only with high-quality (*e.g.* synchrotron) X-ray data that paved the way to the interpretation of the data in favor of the split-cubic model at the expense of the 100% ordered one (Hanusch *et al.*, 2014; Protesescu *et al.*, 2016, 2017; Lignos *et al.*, 2018).

5. Discussion

In this section we address two important issues related to: (i) types and extent of errors made in recovering the NP parameters if data analysis of defective NPs is carried out by using the original MDSE (Cervellino *et al.*, 2010) instead of the corrected one [equation (7)]; (ii) comparison of the $I(Q)$ profiles associated to NPs with constant site-occupancy probability versus constant number of atoms.

5.1. Error estimation when using the original MDSE instead of the corrected one

To evaluate and quantify the errors from previous data analyses performed by using the original MDSE instead of the corrected one, we focus on the first two examples of the previous section. To this purpose, we analyzed the $I(Q)$ profiles generated by using the corrected MDSE [the blue curves of Figs. 2(*a*) and 3(*a*), almost identical to the simulated data] and compared them with those computed by the original MDSE. This analysis was carried out over a Q range typical of wide-angle X-ray total scattering (WAXTS) data (1.4–27.2 $Å^{-1}$) by using the 'standard' *DEBUSSY* suite (Cervellino

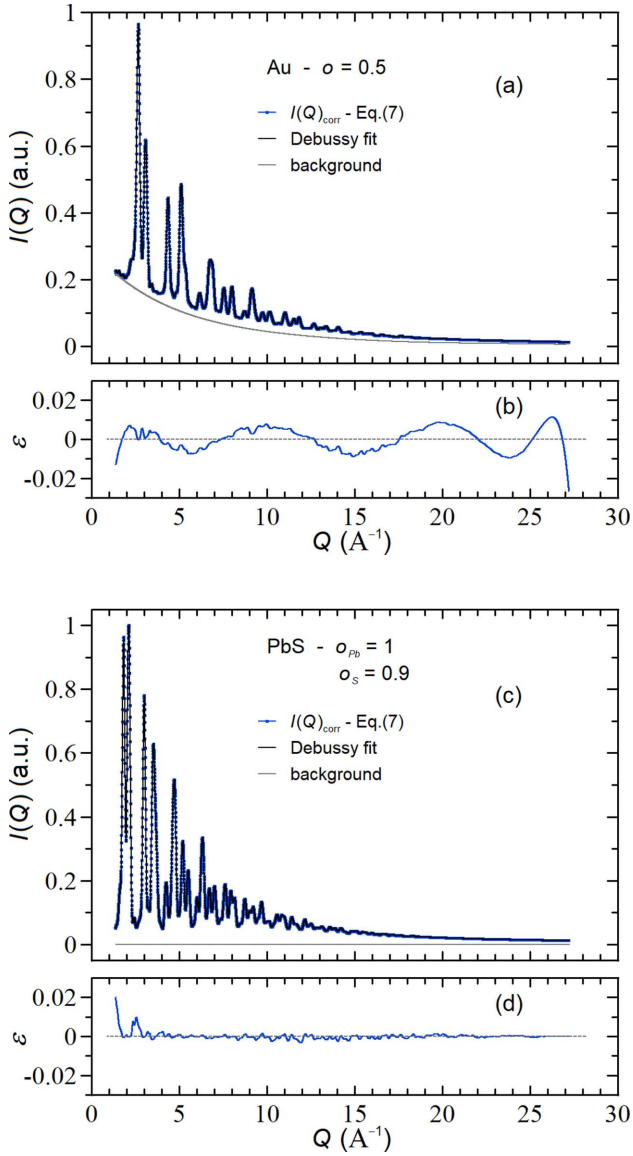


Figure 5
Comparison between the $I(Q)_{corr}$ profiles computed with the corrected MDSE [equation (7), blue curves] and the background-adjusted original MDSE (black curves) for defective Au (*a*) and PbS (*c*) NCs (see Sections 3.1 and 3.2). Panels (*b*) and (*d*) show the relative residuals between $I(Q)_{corr}$ profiles and the original MDSE-based fits. In the case of PbS, a zero background level was imposed [gray line of panel (*c*)].

et al., 2015). The results are summarized in Figs. 5(*a*)–5(*b*) and Figs. 5(*c*)–5(*d*) for the two examples described in Sections 4.1 and 4.2, respectively.

Fig. 5(*a*) shows that, in the case of the Au NC, the profile $I(Q)_{\rm corr}$ (blue curve) can be accurately reconstructed by using as fitting function (black curve) the profile given by the original MDSE plus a background profile (gray curve) optimized by using the Chebyshev polynomials with seven coefficients. The result is quite satisfactory as showcased in Fig. 5(*b*) where the relative residuals $\varepsilon = [I(Q)_{\rm corr} - {\rm fit}]/{\rm fit}$ are well balanced around zero and the overall (dis)agreement parameter is $R = 5.9 \times 10^{-3}$. Consistently with this result, the recovered NC crystallographic and size parameters match quite well the input ones, namely $\langle d_{\rm eqv}\rangle = 3.04$ nm, $\sigma_{d_{\rm eqv}} = 0$ nm and $B = 0.5$ Å^2.

Similarly, for the PbS NC, Fig. 5(*c*) shows that the profile $I(Q)_{\rm corr}$ (blue curve) can be accurately reconstructed (black curve) by using the original MDSE profile only. Indeed, in this (much more realistic) case, the original and corrected MDSE are so close [see Fig. 3(*a*)] that no background contribution (gray curve) has been added, as shown by the residual plot of Fig. 5(*d*), with $R = 1.6 \times 10^{-3}$. NC crystallographic and size parameters are fully recovered also in this case, *i.e.* $\langle d_{\rm eqv}\rangle = 3.67$ nm, $\sigma_{d_{\rm eqv}} = 0$ nm, $o_{\rm S} = 0.90$, $B_{\rm Pb} = 0.5$ Å^2 and $B_{\rm S} = 0.5$ Å^2.

Finally, we would like to comment on the fact that the use of a polynomial background is sufficient for compensating the difference between the $I(Q)$'s computed via the corrected (o) and the original (o^2) MDSE. This is not just a mere coincidence, but it works because this difference is a curve with a very smooth Q dependence that is proportional to the average NP atomic form factors, which are known to be accurately approximated by high-order polynomial functions (Freeman & Smith, 1958).

5.2. Constant site-occupancy probability versus constant number of atoms

As already reported in Sections 3.1 and 3.2, the $I(Q)$ profiles associated to NPs with constant site-occupancy probability and constant number of atoms are described by equations (7) and (10), respectively. These two types of defectiveness are typically associated to NP systems characterized by quite different physical–chemical conditions: (i) the constant site-occupancy probability is linked to a thermodynamically equilibrated system, where NPs of different stoichiometries can be formed. This can occur, *e.g.*, in nano-alloys (Andreazza *et al.*, 2015; Front & Mottet, 2021) and, in the organic chemistry fields, in ideal solid solutions like the ivermectin drug (Shubin *et al.*, 2021). (ii) The fixed (randomly distributed) number of atoms is the common approach in computational chemistry, where stochastic, uncorrelated configurations are studied, and ranked in terms of energetic criteria. This is the case of binary or ternary iron oxides, in the form of maghemite (γ-Fe_2O_3) or tripuhyite ($FeSbO_4$). In both systems, ideal periodicity is broken by the presence of vacant sites in the former (Grau-Crespo *et al.*, 2010) or partially correlated short-range ordering in the latter (Grau-Crespo *et al.*, 2004). In these solids, a constant number of metal ions (for a definite size), granting NC electroneutrality, is required.

Fig. 6 compares the theoretical profiles $I(Q)_{\rm corr}$ given by equations (7) and (10) relative to a f.c.c. Au NC ($a = 4.080$ Å) in the form of a cube with five unit cells per edge, characterized by $M = 666$ sites. In the case of constant occupancy probability, we set $o = 0.5$ so that $\langle n\rangle = Mo = 333$, whereas for a (randomly distributed) constant number of atoms, we set $n' = 333$. As one can easily notice, the two curves are indistinguishable on the log–log plot of Fig. 6(*a*), but exhibit systematic deviations ($\sim 10^{-3}$) at small Q's, as shown in the relative residual plot of Fig. 6(*b*), where $\varepsilon = [I(Q)_{\rm Eq.(7)} - I(Q)_{\rm Eq.(10)}]/I(Q)_{\rm Eq.(10)}$. We can further notice that, while these deviations tend to zero for $Q \to \infty$ (because $n_s' = n_s$), they remain systematically positive for $Q \to 0$. Indeed, by studying the asymptotic behaviors of equations (7) and (10) for $Q \to 0$ we obtain

$$\begin{cases} I(0) = |f|^2\langle n^2\rangle & \{\text{constant occupancy probability}\} \quad (15a) \\ I(0) = |f|^2(n')^2 & \{\text{constant number of atoms}\}, \quad (15b) \end{cases}$$

where in equation (15*a*) we have used the property that the variance of a binomial distribution is $\sigma_n^2 = \langle n\rangle(1-o)$. Equations (15) show that, in the $Q \to 0$ limit, the intensity scales as the square of the average total number of atoms within the NP, consistent with the fact that at zero angle the scattering amplitudes of all the atoms sum up coherently, regardless of their spatial arrangements. However, there is an important difference between the two cases: while for equation (15*b*) the number of atoms is constant, for equation (15*a*) it is a stochastic variable and therefore $\langle n^2\rangle = \langle n\rangle^2 + \sigma_n^2 > \langle n\rangle^2$.

As a consequence, NP ensembles with constant occupancy probability exhibit a higher scattered intensity for $Q \to 0$, the

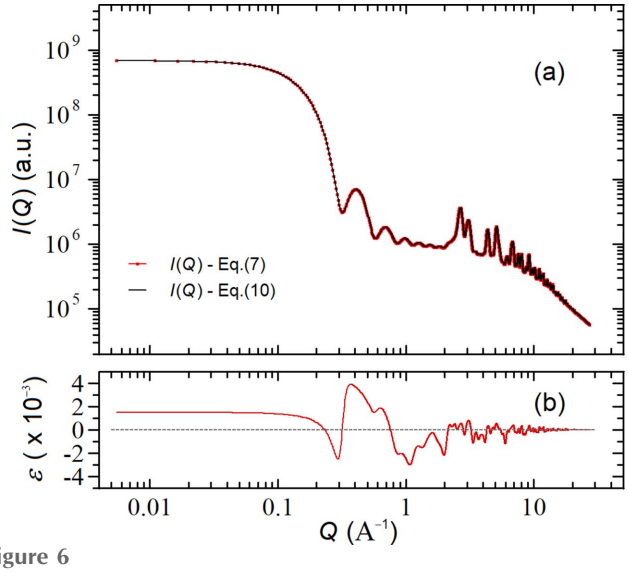


Figure 6
(*a*) Comparison between the $I(Q)$ profiles of defective f.c.c. Au NPs (a = 4.08 Å, M = 666) computed with equation (7) (constant occupancy probability $o_s = 0.5$, red curve) and equation (10) (constant number of atoms $n_s' = 333$, black curve) so that $n_s' = \langle n_s\rangle = Mo_s = 333$; (*b*) relative deviations ε between the two curves, where $\varepsilon = [I(Q)_{\rm Eq.(7)} - I(Q)_{\rm Eq.(10)}]/I(Q)_{\rm Eq.(10)}$.

extra contribution being due to the fluctuations of the number of atoms inside the NPs. Such an extra contribution is given by $\varepsilon(0) = (1 - o)/\langle n \rangle$ (valid for a binomial distribution) and, therefore, under most (realistic) conditions where $n \gg 1$, can be neglected [$\varepsilon(0) \sim 1.43 \times 10^{-3}$ for the curve of Fig. 6(*b*)]. We conclude that, unless ultra-small NPs with sizes of ∼1–2 nm are considered, the two types of defectiveness are indistinguishable in a standard WAXTS analysis where the Q range typically starts from $Q_{\rm min} \geq 0.5$ Å^{-1}.

6. Conclusions

In this paper we have revised (and corrected) the original modified Debye scattering equation (Cervellino *et al.*, 2010) by providing theoretical justification of using atomic s.o.f.'s as free adjustable parameters and presenting cases of application to a variety of defective NP ensembles. For all these cases, we thoroughly discussed the significance of the s.o.f.'s o_i's, provided the analytical expression for the $I(Q)$ profiles, and showed that the o_i's appearing in the $I_{\rm self}$ contribution of the original MDSE must not be squared. The revised versions of MDSE are suitable for describing:

(i) Defective NP ensembles with vacant sites with *uncorrelated constant site-occupancy probability* which describes systems at thermodynamic equilibrium where the number of atoms is not constant, but varies stochastically according to the binomial distribution described in Section 3. The expression for the $I(Q)$ profile is given by equation (7).

(ii) Defective NP ensembles with a *constant number of randomly distributed atoms* that are imposed by stoichiometric constraints. In these systems, the number of atoms of each species is fixed and smaller than the number of available sites for that species. The expression for the $I(Q)$ profile is given by equation (10).

(iii) Defective NP ensembles associated to a special (but common) case of correlated occupancies, *i.e.* the ones with *self-excluding disordered positional sites* where the sites belonging to a given atomic species can be split in different clusters, each of them comprising only one atom (a 'split-atom' model). In this case, the $I(Q)$ profile is computed by restricting the $I_{\rm dist}$ contribution to distances belonging to different clusters and is given by equation (12).

As already pointed out in Section 3.1, the main difference between the original MDSE expression and the corrected one is the presence of (wrong) squared s.o.f.'s appearing in $I_{\rm self}(Q)$. This glitch reduces the continuous and smooth contribution of $I_{\rm self}(Q)$ to the overall scattering profile but does not affect the $I_{\rm dist}(Q)$ term from which the NP structural characterization is extracted. Therefore, the effect on the total intensity $I(Q) = I_{\rm self}(Q) + I_{\rm dist}(Q)$ is a reduction of the diffuse scattering and a corresponding spurious enhancement of the Bragg peaks. The relative discrepancy between the original and corrected MDSE profiles may be compensated by the usage of a smoothly changing polynomial function acting as a 'background' contribution. Our numerical simulations performed on highly defective Au NCs show that this background adjustment is very effective and allows the accurate recovery of the NC crystallographic parameters. More importantly, under realistic conditions, that is with almost unitary occupancy factors in non-elemental systems (less defective PbS NCs), such an additional background adjustment is not required at all and these (negligible) effects do not affect previously published results.

Acknowledgements

We thank A. Parola for the critical reading of the manuscript.

Funding information

MCB thanks the Italian MUR PhD program for the FSE-REACT-EU grant. FB acknowledges partial funding from Fondazione Cariplo (project 2020-4382). AG thanks project PRIN 2017L8WW48 for partial funding.

References

Andreazza, P., Pierron-Bohnes, V., Tournus, F., Andreazza-Vignolle, C. & Dupuis, V. (2015). *Surf. Sci. Rep.* **70**, 188–258.
Bertolotti, F., Dirin, D. N., Ibáñez, M., Krumeich, F., Cervellino, A., Frison, R., Voznyy, O., Sargent, E. H., Kovalenko, M. V., Guagliardi, A. & Masciocchi, N. (2016). *Nat. Mater.* **15**, 987–994.
Bertolotti, F., Moscheni, D., Guagliardi, A. & Masciocchi, N. (2018). *Eur. J. Inorg. Chem.* **2018**, 3789–3803.
Billinge, S. J. L. (2019). *Philos. Trans. R. Soc. A*, **377**, 20180413.
Cervellino, A., Frison, R., Bertolotti, F. & Guagliardi, A. (2015). *J. Appl. Cryst.* **48**, 2026–2032.
Cervellino, A., Frison, R., Masciocchi, N. & Guagliardi, A. (2016). *X-ray and Neutron Techniques for Nanomaterials Characterization*, edited by C. S. S. R. Kumar, pp. 545–608. Berlin: Springer.
Cervellino, A., Giannini, C. & Guagliardi, A. (2010). *J. Appl. Cryst.* **43**, 1543–1547.
Cervellino, A., Gvasaliya, S. N., Zaharko, O., Roessli, B., Rotaru, G. M., Cowley, R. A., Lushnikov, S. G., Shaplygina, T. A. & Fernandez-Diaz, M. T. (2011). *J. Appl. Cryst.* **44**, 603–609.
Coviello, V., Forrer, D. & Amendola, V. (2022). *ChemPhysChem*, **23**, e202200136.
Cullen, D. E., Chen, M. H., Hubbell, J. H., Perkins, S. T., Plechaty, E. F., Rathkopf, J. A. & Scofield, J. H. (1989). *Tables and Graphs of Photon-Interaction Cross Sections from 10 eV to 100 GeV Derived from the LLNL Evaluated Photon Data Library (EPDL)*. Lawrence Livermore National Laboratory, CA, USA.
Debye, P. (1915). *Ann. Phys.* **351**, 809–823.
Freeman, H. C. & Smith, J. E. W. L. (1958). *Acta Cryst.* **11**, 819–822.
Front, A. & Mottet, C. (2021). *Theor. Chem. Acc.* **141**, 2.
Grau-Crespo, R., Al-Baitai, A. Y., Saadoune, I. & De Leeuw, N. H. (2010). *J. Phys. Condens. Matter*, **22**, 255401.
Grau-Crespo, R., de Leeuw, N. H. & Catlow, R. (2004). *Chem. Mater.* **16**, 1954–1960.
Hanusch, F. C., Wiesenmayer, E., Mankel, E., Binek, A., Angloher, P., Fraunhofer, C., Giesbrecht, N., Feckl, J. M., Jaegermann, W., Johrendt, D., Bein, T. & Docampo, P. (2014). *J. Phys. Chem. Lett.* **5**, 2791–2795.
Koester, W. & Gmoehling, W. (1960). *Z. Metallkd.* **51**, 385–391.
Kozioł, R., Łapiński, M., Syty, P., Sadowski, W., Sienkiewicz, J. E., Nurek, B., Adrian Maraloiu, V. & Kościelska, B. (2021). *Appl. Surf. Sci.* **567**, 150802.
Leeuw, N. H. de (2001). *Chem. Commun.* pp. 1646–1647.
Lignos, I., Morad, V., Shynkarenko, Y., Bernasconi, C., Maceiczyk, R. M., Protesescu, L., Bertolotti, F., Kumar, S., Ochsenbein, S. T., Masciocchi, N., Guagliardi, A., Shih, C.-J., Bodnarchuk, M. I., deMello, A. J. & Kovalenko, M. V. (2018). *ACS Nano*, **12**, 5504–5517.

Mark, H. & Szilard, L. (1925). *Z. Phys.* **33**, 688–691.
Mashiyama, H., Kurihara, Y. & Azetsu, T. (1998). *J. Korean Phys. Soc.* **32**, 156.
Newmai, M. B., Verma, M., Dahiya, A. & Kumar, P. S. (2022). *J. Phys. Chem. Solids*, **161**, 110371.
Pearson, W. B. & Hume-Rothery, W. (1952). *J. Inst. Met.* **80**, 641–652.
Protesescu, L., Yakunin, S., Bodnarchuk, M. I., Bertolotti, F., Masciocchi, N., Guagliardi, A. & Kovalenko, M. V. (2016). *J. Am. Chem. Soc.* **138**, 14202–14205.
Protesescu, L., Yakunin, S., Kumar, S., Bär, J., Bertolotti, F., Masciocchi, N., Guagliardi, A., Grotevent, M., Shorubalko, I., Bodnarchuk, M. I., Shih, C.-J. & Kovalenko, M. V. (2017). *ACS Nano*, **11**, 3119–3134.
Scardi, P. & Gelisio, L. (2016). *Sci. Rep.* **6**, 22221.
Sears, V. F. (1992). *Neutron News*, **3**, 26–37.
Shubin, K., Bērziņš, A. & Belyakov, S. (2021). *Crystals*, **11**, 172.
Warren, B. E. (1990). *X-ray Diffraction*. New York: Dover Publications.