JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN: 1600-5767

journals.iucr.org/j

Efficient computation of the scattering intensity from systems of nonspherical particles

Rasmus A. X. Persson and Johan Bergenholtz

J. Appl. Cryst. (2016). **49**, 1524–1531

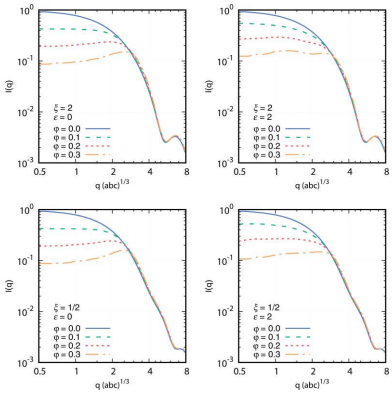
JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Received 3 May 2016
Accepted 14 July 2016

Edited by Th. Proffen, Oak Ridge National Laboratory, USA

Keywords: small-angle scattering; form factor amplitude; numerical algorithms; Monte Carlo.



Efficient computation of the scattering intensity from systems of nonspherical particles

Rasmus A. X. Persson[a]* and Johan Bergenholtz[a,b]

[a]Department of Chemistry and Molecular Biology, University of Gothenburg, SE-412 96 Gothenburg, Sweden, and [b]Division of Physical Chemistry, Center of Chemistry and Chemical Engineering, Lund University, SE-221 00 Lund, Sweden. *Correspondence e-mail: rasmus.a.persson@gmail.com

The analysis of the angle dependence of the elastic scattering of radiation from a sample is an efficient and non-invasive technique that is used in fundamental science, in medicine and in technical quality control in industry. Precise information on the shape, size, polydispersity and interactions of a colloidal sample is readily obtained provided an underlying scattering model, *i.e.* form and structure factors, can be computed for the sample. Here, a numerical method that can efficiently compute the form factor amplitude (and thus the scattering intensity) of nonspherical scatterers through an importance sampling algorithm of the Fourier integral of the scattering density is presented. Using the precomputed form factor amplitudes, the calculation of the scattering intensity at any particle concentration then scales linearly with the particle number and linearly with the number of q points for its evaluation. This is illustrated by an example calculation of the scattering by concentrated suspensions of ellipsoidal Janus particles and the numerical accuracy for the computed form factor amplitudes is compared with analytical benchmarks.

1. Introduction

Scattering experiments constitute a versatile analysis technique with important applications in the biomedical field (Wax & Backman, 2010), the food industry (Lu, 2016) and the chemical industry (Xu, 2000; Diebold, 2014). However, unlike direct techniques such as light and electron microscopy, the analysis of scattering experiments is typically oblique, being generally based on the fit of a calculated scattering intensity to the experimental one, in which the physical parameters of the calculation model are varied until optimal agreement is reached. In crystallography, this procedure yields the unit cell atomic arrangement, and in the study of isotropic media, typically the average size, shape and/or polydispersity of the sample are obtained (Pedersen, 1997). Moreover, at higher concentrations, second virial coefficients (Zimm, 1948) or in principle even complete particle–particle distribution functions (and thus potentials of mean force) can be deduced (Zimm, 1945). Needless to say, the accuracy of such techniques depends crucially on the underlying models to interpret the data. The problem is that the standard models for interpreting the scattered intensity typically assume spherical scatterers and may be misleading when applied to nonspherical ones. This is hence a great impediment to the use of scattering techniques to characterize these systems.

For a system of N monodisperse spherical scatterers located at positions $\mathbf{r}_1, \ldots, \mathbf{r}_N$ in a volume V, one may write the scattered intensity as a function of the angle-dependent scattering vector (defined in §2) as (Chen, 1986; Johnson & Gabriel, 1995; Weyerich *et al.*, 1999; Hansen, 2011)

$$I(\mathbf{q}) \propto \frac{N}{V} P(\mathbf{q}) S(\mathbf{q}), \tag{1}$$

where $P(\mathbf{q})$ is called the form factor (this function depends only on the particle form, *i.e.* size in the case of spheres) and $S(\mathbf{q}) = \langle N^{-1} | \sum_j \exp(i\mathbf{q}\cdot\mathbf{r}_j)|^2 \rangle$ is the structure factor (this function depends only on the interparticle arrangement). Because of the spherical symmetry, the structure factor may also be given the succinct representation (Debye, 1915)

$$S(q) = \left\langle \frac{1}{N} \sum_{jk}^{N} \frac{\sin(qr_{jk})}{qr_{jk}} \right\rangle, \tag{2}$$

where r_{jk} is the scalar distance between scatterers j and k. From this equation, one sees that $S(q) \to 1$ when $N/V \to 0$ (corresponding to increasing values of r_{jk}) but also that $S(q) \to 1$ as $q \to \infty$, both of which are in fact more general results than just for spheres. The form factor is thus always the dominant contribution to the scattered intensity at high enough values of q, although the scattered intensity is generally weak in this q regime.

For polydisperse or nonspherical scatterers (unless they all share the same alignment, as in a perfect crystal), this factorization is no longer generally valid. However, in the dilute limit, an effective form factor, $P_{\rm eff}(\mathbf{q})$, is unambiguously defined as the averaged scattered intensity of the individual scatterers, and if $P_{\rm eff}(\mathbf{q})$ is taken to be concentration independent, then an effective structure factor, for concentrated samples, can also be defined from

$$I(\mathbf{q}) \propto \frac{N}{V} P_{\rm eff}(\mathbf{q}) S_{\rm eff}(\mathbf{q}). \tag{3}$$

When averaged over time scales longer than that of the diffusive motion of the sample scatterers (*e.g.* isotropic fluids, not crystals or liquid crystals), the effective form and structure factor will both depend only on $q = |\mathbf{q}|$ rather than $\mathbf{q}$ because of the angular averaging. Unlike the case of monodisperse spheres, $S_{\rm eff}(q)$ is not a function only of the particle arrangement. Care must therefore be taken when interpreting $S_{\rm eff}(q)$ as $S(q)$.

The problem with nonspherical scatterers is that analytical expressions for the structure and form factors are typically not available (Svergun & Koch, 2003). For systems of monodisperse hard spheres, the most frequently used expression for the structure factor (Wertheim, 1963; Thiele, 1963; Ashcroft & Lekner, 1966) is based on the Percus–Yevick theory (Percus & Yevick, 1958) for hard spheres, later extended to the polydisperse case by Vrij (1979) and Blum & Stell (1979). Baravian *et al.* (2010) introduced an approximate correction to the spherical structure factor for nonspherical particles on the basis of geometrical considerations that worked well except at small q. Also, Hansen has considered the problem of nonspherical symmetry (Hansen, 2011, 2012), as well as that of polydispersity (Hansen, 2013), in approximate (although partly analytical) analyses similar in spirit to that of Baravian *et al.* (2010).

Whereas analytical results are typically not possible in the general case, accurate numerical results for the scattering intensity and/or the structure factor of hard ellipsoids at high volume fractions have been obtained from computer simulations by Sjöberg (1999), Donev *et al.* (2004) and Bezrukov & Stoyan (2006), and from integral equation theory by Perera *et al.* (1987) and Letz & Latz (1999). Weyerich *et al.* (1999) studied the validity of approximations for the effective structure factor of random orientations of rods, limiting themselves to the case of low q. Nevertheless, even when provided with an interaction potential, the numerical approach solves the problem of obtaining $I(q)$ only if the so-called form factor amplitudes (defined in §2) are available.

It is the purpose of this paper to present a general numerical method to compute the form factor amplitude (valid at any concentration) and thus also $P_{\rm eff}(q)$ (valid for dilute systems or at high q) efficiently. Provided with an interaction potential, one may then – as we shall see – obtain $S_{\rm eff}(q)$ from $I(q)$ for nondilute systems if the form factor amplitudes are precomputed. Notable among other numerical approaches to the problem of calculating the scattered intensity are the direct fast Fourier transform method of Schmidt-Rohr (2007) and the indirect methods where spatial correlation functions are first sampled and then Fourier transformed (Hansen, 1990; Henderson, 1996; Heller, 2006). A general comparison of the efficacy of the direct Fourier and correlation methods has been presented by Olds & Duxbury (2014). We view our method as complementary. Unlike any fast Fourier transform approach, the integration domain for the form factor amplitude is – as we shall see – focused to the regions where contributions are likely to be the greatest, which is a strategy that for obvious reasons is conducive to a highly efficient computation. Moreover, the single sum in equation (4) is linear in algorithmic complexity with respect to the particle number, which is better than the (weakly) superlinear complexity with respect to projected particle scattering densities of the fast Fourier algorithms.

2. Theory

The physical model that we use for calculating the scattering intensity assumes a sample-incident ray with wavevector $\mathbf{k}_{\rm i}$ which generates elastically scattered radiation with the wavevector $\mathbf{k}_{\rm s}$ (we ignore all inelastic scattering: the so-called Raman/Compton scattering for photons/neutrons). Introducing the so-called scattering vector $\mathbf{q} = \mathbf{k}_{\rm s} - \mathbf{k}_{\rm i}$, the scattered intensity, relative to the incident radiation intensity, is given by (Frenkel *et al.*, 1986; Hess *et al.*, 1989; Sjöberg, 1999)

$$I(\mathbf{q}) \propto \frac{1}{V} \left\langle \left| \sum_{j=1}^{N} f_j(\mathbf{q}, \boldsymbol{\omega}_j) \cos(\mathbf{q}\cdot\mathbf{r}_j) \right|^2 + \left| \sum_{j=1}^{N} f_j(\mathbf{q}, \boldsymbol{\omega}_j) \sin(\mathbf{q}\cdot\mathbf{r}_j) \right|^2 \right\rangle, \tag{4}$$

where $\mathbf{r}_j$ is the position of the jth scatterer in the sample, $\boldsymbol{\omega}_j$ is its orientation and $f_j(\mathbf{q}, \boldsymbol{\omega}_j)$ is the form factor amplitude – a measure of the intrinsic propensity to scatter radiation elastically with scattering vector $\mathbf{q}$ – of this scatterer. The sum extends in principle over all N scatterers impinged on by the radiation beam (they are contained in the volume V), but is

typically limited for reasons of computational tractability when the sample is not translationally symmetric.

The form factor amplitude has a nonnegligible dependence on $\mathbf{q}$ as long as the wavelength of the radiation is not much greater than the spatial extension of the scattering particle. For proteins, even scattering in the visible spectrum typically satisfies this criterion, and for other bioparticles studied by light scattering such as virus capsids, this is definitely true. The form factor amplitudes are directly related to the experimentally scattered intensity at high dilution or high q values. Under these conditions, it suffices to calculate $P_{\rm eff}(\mathbf{q})$ directly as the unweighted orientational average,

$$P_{\rm eff}(\mathbf{q}) = \left\langle \frac{1}{N}\sum_{j=1}^{N} |f_j(\mathbf{q})|^2 \right\rangle. \tag{5}$$

At higher concentrations, one may then calculate $S_{\rm eff}(\mathbf{q})$ by dividing equation (4) by $P_{\rm eff}(\mathbf{q})$.

2.1. Statement of the problem

To use equation (4), we need the $\mathbf{q}$ dependence of the form factor amplitude of each particle. Under the assumption of monodispersity,[1] this reduces to the computation of a single form factor amplitude common to all particles.[2] Nevertheless, depending on the orientation $\boldsymbol{\omega}$ of each particle with respect to the incident radiation, the form factor amplitude differs. Thus, the problem reduces to computing the function $f(\mathbf{q}, \boldsymbol{\omega})$.

Given an arbitrary continuous distribution $\rho(\mathbf{r}; \boldsymbol{\omega})$ of scattering propensity (for instance, the electron distribution for X-ray scattering, or the distribution of nuclei for neutron scattering) defining the (scattering) shape of the particle, the form factor amplitude can be formally obtained as the three-dimensional Fourier transform (Guinier, 1939),

$$\begin{aligned} f(\mathbf{q}, \boldsymbol{\omega}) &= \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r}; \boldsymbol{\omega}) \exp(i\mathbf{q}\cdot\mathbf{r}) \\ &= \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r}; \boldsymbol{\omega}) [\cos(\mathbf{q}\cdot\mathbf{r}) + i\sin(\mathbf{q}\cdot\mathbf{r})]. \end{aligned} \tag{6}$$

2.2. Numerical algorithm

Up to a normalization factor, we compute the integral in equation (6) by the Metropolis importance sampling technique (Metropolis *et al.*, 1953). Applied to the present problem, the method is realized as follows. A Markov chain is constructed in $\mathbb{R}^3$ with the probability of going from state $\mathbf{r}$ to state $\mathbf{r}'$ proportional to $\rho(\mathbf{r}'; \boldsymbol{\omega})$ and denoted by $P(\mathbf{r}' \mid \mathbf{r})$. The probability is normalized such that

$$\int_{v(\mathbf{r})} P(\mathbf{r}' \mid \mathbf{r})\, \mathrm{d}\mathbf{r}' = 1. \tag{7}$$

where $v(\mathbf{r})$ is a volume element of arbitrary finite extension containing $\mathbf{r}$. If the detailed balance condition (see Appendix *A*) is fulfilled, the averages of $\sin(\mathbf{q}\cdot\mathbf{r})$ and $\cos(\mathbf{q}\cdot\mathbf{r})$ along the Markov chain approach

$$\langle \sin(\mathbf{q}\cdot\mathbf{r})\rangle_{\omega} \to (1/\Omega)\int \rho(\mathbf{r}; \boldsymbol{\omega}) \sin(\mathbf{q}\cdot\mathbf{r})\, \mathrm{d}\mathbf{r}, \tag{8}$$

$$\langle \cos(\mathbf{q}\cdot\mathbf{r})\rangle_{\omega} \to (1/\Omega)\int \rho(\mathbf{r}; \boldsymbol{\omega}) \cos(\mathbf{q}\cdot\mathbf{r})\, \mathrm{d}\mathbf{r}, \tag{9}$$

where (independently of $\boldsymbol{\omega}$)

$$\Omega = \int \rho(\mathbf{r}; \boldsymbol{\omega})\, \mathrm{d}\mathbf{r} \tag{10}$$

as the Markov chain is extended to infinity. The constant Ω fixes the scale of the scattering intensity and is the unknown normalization factor alluded to earlier. If $\partial\rho(\mathbf{r}; \boldsymbol{\omega})/\partial\mathbf{r} = 0$ for all $\mathbf{r}$ such that $\rho(\mathbf{r}) \neq 0$ and ρ is always nonnegative, Ω is directly proportional to the volume of the particle (for a spatially varying ρ, the volume is not well defined). Note that for a centrosymmetric particle the form factor amplitude is real valued, and $\langle \sin(\mathbf{q}\cdot\mathbf{r})\rangle_{\omega} \to 0$.

The detailed balance condition is easily fulfilled by a simple random walk in $\mathbb{R}^3$ with an acceptance probability proportional to $\rho(\mathbf{r}; \boldsymbol{\omega})$, if the following constraints on $v(\mathbf{r})$ are met. The extension of the volume element is independent of $\mathbf{r}$, and $\mathbf{r}$ is centered in $v(\mathbf{r})$, in the sense that

$$\int_{-a}^{0} v[\mathbf{p}(t)]\, \mathrm{d}t = \int_{0}^{a} v[\mathbf{p}(t)]\, \mathrm{d}t, \tag{11}$$

where $\mathbf{p}(t)$ is an arbitrary straight line between points $\mathbf{p}(-a)$ and $\mathbf{p}(a)$ on the surface of $v(\mathbf{r})$ and $\mathbf{p}(t)$ satisfies $\mathbf{p}(0) = \mathbf{r}$. In this case, generating an unbiased random vector $\mathbf{r}'$ on any point along the line $\mathbf{p}(t)$ in $v(\mathbf{r})$ and accepting it with probability $\min[1, \rho(\mathbf{r}'; \boldsymbol{\omega})/\rho(\mathbf{r}; \boldsymbol{\omega})]$ (and otherwise counting $\mathbf{r}$ once more) is sufficient to obtain the average form factor amplitude in equation (6).

For selected values of $\mathbf{q}$ and $\boldsymbol{\omega}$, the sine and cosine (imaginary and real) parts of $f(\mathbf{q}, \boldsymbol{\omega})$ are calculated and stored in computer memory for later use in equation (4). The time complexity of the algorithm is such that it scales linearly with the number of q points selected. Linear interpolation ensures that $f(\mathbf{q}, \boldsymbol{\omega})$ is a continuous function of $\mathbf{q}$ and $\boldsymbol{\omega}$. The volume $v(\mathbf{r})$ is chosen as a small cube centered on $\mathbf{r}$ in the laboratory frame. Formally, there are no restrictions on its size, but it does affect the efficacy of the algorithm. In the calculations, the size of this cube was adjusted to achieve an acceptance probability close to 50% during trials without sampling.

2.3. General limitations

Two, closely related, limitations are apparent from the algorithm itself.

First, the algorithm assumes that $\rho(\mathbf{r}; \boldsymbol{\omega})$ is positive definite. For photon and neutron scattering *in vacuo*, this is not a problem. However, for scattering in solution, $\rho(\mathbf{r}; \boldsymbol{\omega})$ should be interpreted as the scattering contrast density between the solvent and the particle, in which case $\rho(\mathbf{r}; \boldsymbol{\omega})$ could conceivably attain negative values in certain domains.[3] While these domains are, however, unlikely to be large, it does present a

[1] This is the natural assumption in many cases when studying biological polymers such as proteins.

[2] If a distribution of particle shapes and sizes exists, the method may be applied, at arbitrary resolution, to shapes/sizes drawn from this distribution.

[3] Strictly speaking, this is an approximation valid for low solvent scattering density (in which case negative domains do not arise) and/or large wavelengths. In principle, one may avoid this problem by treating the system as a mixture and defining form factor amplitudes also for the solvent molecules.

formal limitation for the algorithm that precludes, for instance, numerical studies corresponding to experiments employing contrast variation.

Second, and related to the first limitation, is the case of a disjunct $\rho(\mathbf{r};\boldsymbol{\omega})$ distribution, in the sense that there are at least two points $\mathbf{r}'$ and $\mathbf{r}''$ that satisfy $\rho(\mathbf{r}';\boldsymbol{\omega}), \rho(\mathbf{r}'';\boldsymbol{\omega}) > 0$ that cannot, however, be connected through the random walk since one or the other is separated on all sides by regions of $\rho(\mathbf{r};\boldsymbol{\omega})$ that are greater than the volume v and strictly zero. Even if this problem can be formally solved by increasing the size of v, it does present a practical problem for the convergence of the Markov chain.

We expect these two formal problems to be of minor practical hindrance in the application to actual scattering problems, as particles are unlikely to be composed of completely disjunct domains in their scattering density and can almost always be assured to scatter radiation more intensely than the solvent through a more appropriate choice of the latter.

3. Results and discussion

To validate the method numerically, we will first compute the form factor amplitude by the importance sampling algorithm and compare with analytical results. Analytical results are only available in a handful of high-symmetry cases. In addition to the spherical case, we have computed form factor amplitudes also for ellipsoids of revolution, both prolate and oblate. Both the case of a homogeneous scattering density distribution and that of a Gaussian one will be investigated. Finally, we show through a simple case study of dense systems of interacting ellipsoids that the approach of storing precomputed form factor amplitudes at a finite resolution is sufficient to yield good scattering curves also for strongly interacting systems.

3.1. Numerical estimates of the form factor amplitude

Although in general one must store two six-dimensional arrays for $f(\mathbf{q},\boldsymbol{\omega})$ (three dimensions for $\mathbf{q}$, three for $\boldsymbol{\omega}$, both once each for the imaginary and real parts of the form factor amplitude), in high-symmetry cases, one needs only a one-dimensional array. This is the case for the ellipsoid of revolution generally and spherical particles in particular. Incidentally, it is also for the high-symmetry cases that analytical results to compare with are available and let us judge the accuracy of the algorithm. However, we do not let the algorithm take advantage of this extra symmetry: symmetry-equivalent points are not averaged for improved statistics and thus contribute more than once to the average unsigned error. Also, the imaginary part of the form factor amplitude is sampled and included in the error, even if it should vanish by symmetry. In this way, the error provides a fair assessment of the algorithmic accuracy also in the general, low-symmetry case.

Finally, note that the algorithm is – in the language of computer science – 'embarrassingly parallel' as the integral in equation (6) can be independently evaluated in parallel by any number of random walkers and then averaged for improved convergence with effectively no upper limit on the number of Markov chains employed.

3.1.1. Homogeneous ellipsoids of revolution. Consider an ellipsoid of revolution with semiaxes a, b and c. For simplicity, we assume that $b = c$ and define the axial ratio $\xi = a/b$ which is greater than unity for the prolate and less than unity for the oblate. For homogeneous ellipsoids of revolution, the exact expression for the form factor amplitude of the particles is (Guinier, 1939; Saitô & Ikeda, 1951)

$$f(\mathbf{q},\boldsymbol{\omega}) \equiv f(qK) = 3\frac{\sin(qK) - qK\cos(qK)}{(qK)^3} \qquad (12)$$

with $K = [a^2\cos^2(\varphi_a) + b^2\cos^2(\varphi_b) + c^2\cos^2(\varphi_c)]^{1/2}$. Here φ_a, φ_b and φ_c are the angles between the wavevector $\mathbf{q}$ and the principal a, b and c axes of the ellipsoid, respectively. Note that this function is one dimensional in the combined variable qK. Consequently, the actual choice of the orientation $\boldsymbol{\omega}$ used in the numerical evaluation becomes arbitrary after projection into the qK coordinate.

Plotted in Fig. 1 is the mean unsigned relative error, averaged over 100 independent Markov chains of the indicated length, in the q range $0 < qK < 1$ for different lengths of the Markov chain compared with the exact result for the form factor amplitude of an ellipsoid at different axial ratios. It is noteworthy that already for Markov chain lengths of 1000 states the mean unsigned relative error is less than one percent. For longer Markov chains, the form factor amplitudes obtained become virtually indistinguishable from the exact expression.

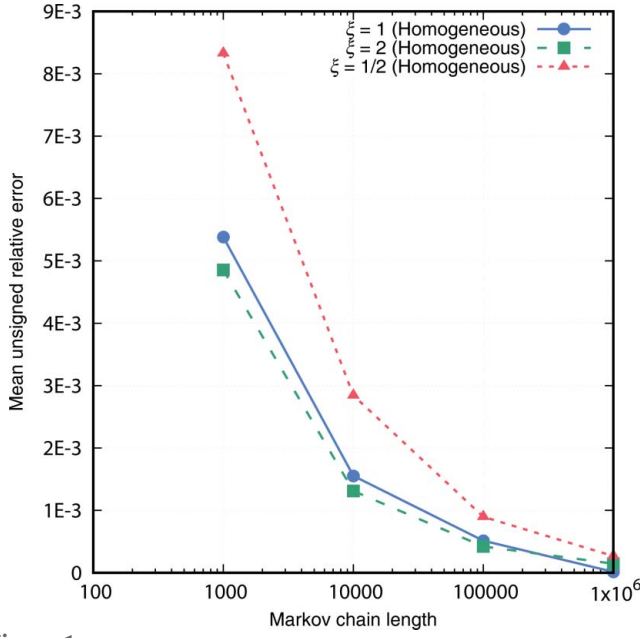


Figure 1
Mean unsigned relative error defined as $\langle||f_{\rm calc}(q)|^2/|f_{\rm exact}(q)|^2 - 1|\rangle$ for $0 < q < (abc)^{-1/3}$ as a function of the length of the Markov chain for the numerically computed form factor amplitude $f_{\rm calc}(q)$ with respect to the exact analytical expression $f_{\rm exact}(q)$ for inhomogeneous Gaussian distributions.

For all of the finite Markov chain lengths studied, the accuracy of the numerical integration is the lowest for the oblate ellipsoid, although its marginal improvement is also the strongest on increasing the chain length. For all but the longest set of chains sampled, the accuracy is higher for the prolate than for the sphere, although the rate of convergence with respect to the number of Markov states is higher for the sphere. In any case, for 10^6 Markov steps, the accuracy for either shape is so high (relative unsigned error less than one-fifth of one part per thousand) that it is virtually indistinguishable from the analytical result.

3.1.2. Gaussian body. Arguably, the homogeneous body not only is probably a crude approximation for many real particles but also represents the least efficient use of the importance sampling in the algorithm. Because of the constant ρ, the importance sampling is nullified and the algorithm reduces almost to traditional Monte Carlo integration. As another test case, we therefore consider the general inhomogeneous Gaussian body defined by

$$\rho(\mathbf{r}) = \pi^{-3/2} \exp(-\alpha^{-2}x^2 - \beta^{-2}y^2 - \gamma^{-2}z^2), \qquad (13)$$

where $\mathbf{r} = (x, y, z)$. Its form factor amplitude is analytically given by

$$f(\mathbf{q}) = (abc)^{-1/2} \exp[-\tfrac{1}{4}(\alpha^2 q_x^2 + \beta^2 q_y^2 + \gamma^2 q_z^2)], \qquad (14)$$

where $\mathbf{q} = (q_x, q_y, q_z)$.

The error – averaged without sign in the range $0 < \alpha^2 q_x^2 + \beta^2 q_y^2 + \gamma^2 q_z^2 < 1$ – with respect to this analytical expression of the numerical integration is provided graphically in Fig. 2. Since the radius of the Gaussian is not well defined, the comparison with the homogeneous ellipsoids is not perfect as qK is scaled by this quantity. For instance, for the spherical case, plotting the values of Figs. 1 and 2 on the same abscissa implicitly corresponds to taking the 'homogeneous equivalence' radius of the Gaussian from the condition $\exp(-\mathbf{r}^2) = \exp(-1) \simeq 0.37$, which is a very small value for the 'equivalent' radius as the scattering-equivalent volume contained within this radius is less than half of that of the homogeneous sphere of the same radius. Of course, the condition that both the Gaussian and the homogeneous sphere share the same value of Ω leads to an infinite radius for the Gaussian, and thus the equivalent q range is reduced to zero and the average error in this zero-length range vanishes automatically.

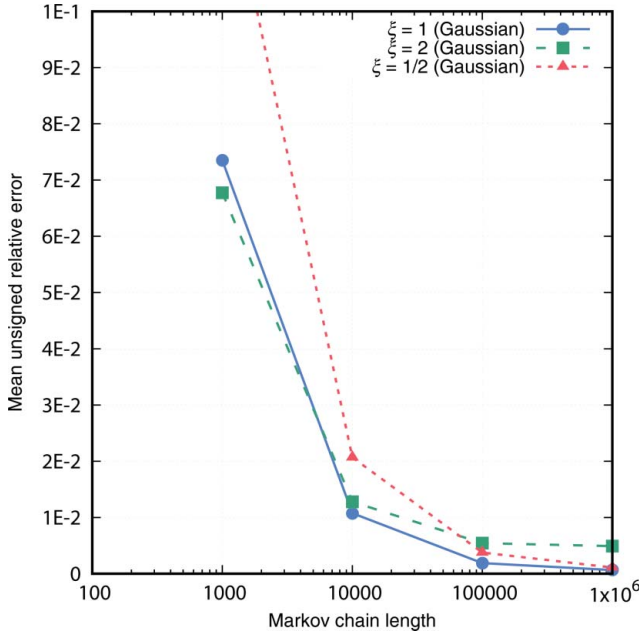


Figure 2
Mean unsigned relative error defined as $\langle||f_{\rm calc}(q)|^2/|f_{\rm exact}(q)|^2 - 1|\rangle$ for $0 < \alpha^2 q_x^2 + \beta^2 q_y^2 + \gamma^2 q_z^2 < 1$ as a function of the length of the Markov chain for the numerically computed form factor amplitude $f_{\rm calc}(q)$ with respect to the exact analytical expression $f_{\rm exact}(q)$ for different aspect ratios $\xi = \alpha/\beta = \alpha/\gamma$ for three-dimensional Gaussians.

3.2. Scattering signal from concentrated systems of Janus ellipsoids

Having established the feasibility of the approach, let us investigate some typical scattering signals and their dependence on the underlying particle interactions, for the particular case of the homogeneous ellipsoid. This is intended as a proof of concept.

While our focus on ellipsoidal particles with Janus interactions – particles that are non-centrosymmetric in their potential energy of interaction, having one side which attracts and one side which repels – is merely for illustrative purposes, the choice is judicious. In nature, such strong directional interaction is quite commonly encountered in proteins which tend to dimerize (Marianayagam *et al.*, 2004). Virus capsids are also made up of self-assembling proteins with strong directionality in their interactions (Almendral, 2013; San Martín, 2013). Moreover, synthetic Janus particles are also very prevalent, as they exhibit many actual and possible industrial applications (Walther & Müller, 2013), and consequently their self-assembly behavior has attracted interest through both computer simulation studies (Sciortino *et al.*, 2009, 2010; Liu *et al.*, 2011) and experiments (Chen *et al.*, 2011). Here, we use them as an illustrative model example of an analysis of the scattering intensity.

3.2.1. Interaction model. As a generic model for the interaction between the Janus particles, we employ a generalization of the Kern–Frenkel potential (Kern & Frenkel, 2003) to ellipsoids of revolution. In a local Cartesian coordinate system (x, y, z), a hard ellipsoid is defined by

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 + \left(\frac{z}{c}\right)^2 = 1, \qquad (15)$$

where a, b and c denote the semiaxes of the ellipsoid. Overlap of two such hard ellipsoids, as determined by the algorithm of Perram & Wertheim (1985), leads to an infinite potential energy. Centered on the hard ellipsoid and aligned with it, but with semiaxes μa, μb and μc ($\mu \geq 1$), is another ellipsoid of revolution. Overlap of this 'soft' ellipsoid with its homolog on a neighboring particle leads to an energy

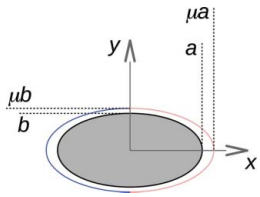


Figure 3
Schematic cross-section illustration of the modified Kern–Frenkel potential defining the particle interactions. Overlap of two inner ellipsoids (denoted by the grey region enclosed within the black solid line) leads to an infinite potential energy. Overlap of two outer ellipsoids (enclosed by the dark-blue and light-red solid lines) leads to an energy of $-\varepsilon$ if the point of contact is 'red' for both ellipsoids, $+\varepsilon$ if 'blue' for both ellipsoids and zero otherwise. The extents of the semiaxes a and b are marked on the x and y axes, respectively, by dashed lines. Likewise, the outer ellipsoid is defined by the semiaxes μa, μb and μc.

$$\begin{aligned} -\varepsilon < 0 \quad &\text{if} \quad \hat{\mathbf{e}}_i \cdot \hat{\mathbf{r}}_{ij} < 0 \wedge \hat{\mathbf{e}}_j \cdot \hat{\mathbf{r}}_{ij} > 0 \\ \varepsilon > 0 \quad &\text{if} \quad \hat{\mathbf{e}}_i \cdot \hat{\mathbf{r}}_{ij} > 0 \wedge \hat{\mathbf{e}}_j \cdot \hat{\mathbf{r}}_{ij} < 0 \\ 0 \quad &\text{otherwise} \end{aligned}$$

where $\hat{\mathbf{e}}_i$ is the unit vector pointing along the x axis of particle i and $\hat{\mathbf{r}}_{ij}$ is the interparticle unit vector. A schematic illustration of the interaction sites is given in Fig. 3. We arbitrarily set $\mu = 1.1$ in all our simulations. Only the inner ellipsoid scatters radiation. For conciseness, we restrict our study to the two cases $a > b = c$ (prolate ellipsoid) and $a < b = c$ (oblate ellipsoid), although the model is applicable also when $b \neq c$. Most importantly, nothing except computer power and patience limits the range of validity to small scattering angles in our treatment.

After sufficient equilibration, Metropolis Monte Carlo simulations in a cubic simulation cell with periodic boundary conditions and 512 particles were performed for 10^8 cycles at every reported state point. The thermodynamic temperature was set to $T = k_{\rm B}^{-1}$, $k_{\rm B}$ being the Boltzmann constant, in all simulations. Equation (4) was evaluated and averaged over snapshots from numerical simulations of collections of ellipsoids. The summation over all particles is of linear algorithmic complexity in the number of particles, *i.e.* $\mathcal{O}(N)$, and thus asymptotically faster than fast Fourier algorithms that scale as $\mathcal{O}(N_{\rm c} \ln N_{\rm c})$, where $N_{\rm c}$ is the number of grid cells.

3.2.2. Scattered intensity. In Fig. 4, we employ the precomputed form factors (at a homogeneous q resolution of $\Delta(qK) = 0.062832$; sampled for 10^6 Markov steps) to gauge the effect of changing the volume fraction φ, the Janus interaction strength ε and the axial ratio ξ on the scattered intensity as a function of q. Such curves correspond directly to that which is measured in the scattering experiment, and have previously been reported in part by Sjöberg (1999), although for a much more restricted range of q and only for the case $\varepsilon = 0$ in the present notation. For the q range that we consider here, the influence of the form factor of the particles is evident also at the highest volume fractions of $\varphi = 0.3$, with or without Janus interactions, and its exact curve is included in the figure for comparison purposes. Whereas the increasing volume fraction leads to a substantial suppression of the scattered intensity at low q even for fairly dilute concentrations of $\varphi = 0.1$, the inclusion of the Janus interactions – the strength of which was roughly chosen to be as large as possible yet without inducing crystallization – leads only to a very small difference in the scattered intensity with respect to that of the purely repulsive interaction.

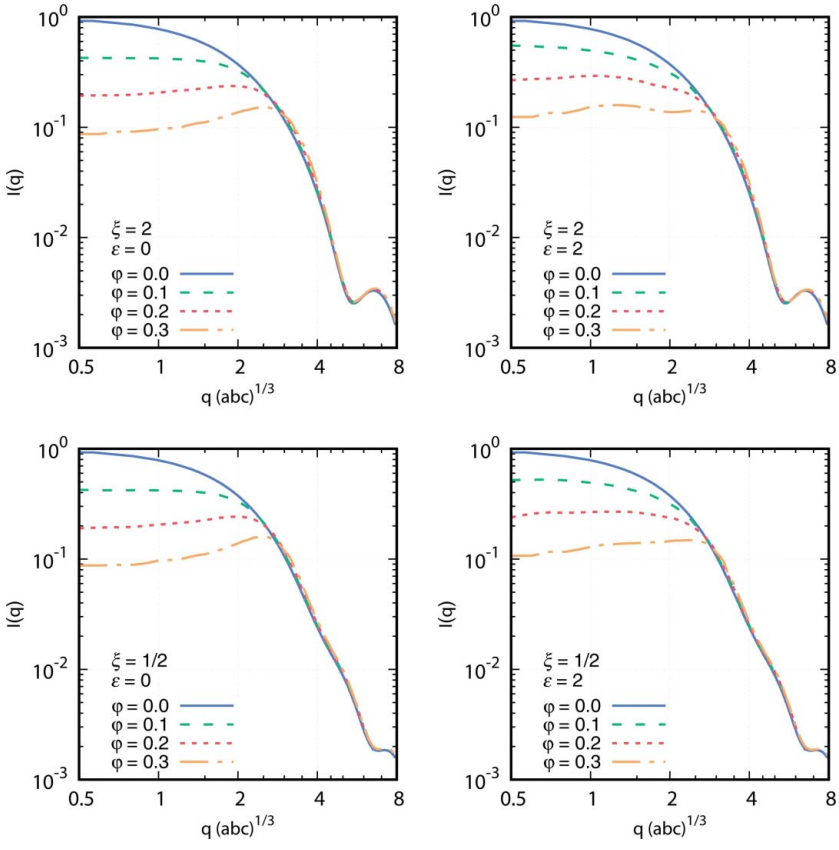


Figure 4
Normalized intensity of the scattered radiation as a function of q at different volume fractions φ, axial ratios ξ and Janus interaction strengths ε. The curve for $\varphi = 0.0$ (the form factor) is calculated using the analytical expression (12), the others using the numerical algorithm.

4. Conclusion

We have presented a numerical algorithm to compute the form factor amplitude of any continuous body. For homogeneous ellipsoids of revolution, the algorithm yields an average error of less than one-tenth of a percent for q values up to $(abc)^{-1/3}$, and for Gaussian bodies the accuracy is arguably even greater. The algorithm scales linearly with the number of q points chosen for the resolution. Given the speed of the algorithm on modern computers,

an adaptive search over arbitrary scattering distributions should be possible to allow fitting to experimentally measured form factors of dilute suspensions or to experimental scattering intensities of concentrated suspensions but for large q values.

Using the precomputed form factor amplitudes, we have calculated the scattering intensity, at both low and high q, of concentrated systems of Janus ellipsoids using a single sum over the particle positions. In total, the computation of the scattering signal of a single snapshot scales bilinearly with the number of particles and with the number of q points, which is asymptotically better than fast Fourier transform approaches.

APPENDIX A
Proof of equations (8) and (9)

The following proof has appeared in a slightly different form (Metropolis *et al.*, 1953) for the calculation of thermodynamic partition functions and is here adapted to the present case, for the benefit of the reader who is not familiar with Metropolis Monte Carlo simulations.

Since equations (8) and (9) share the same mathematical form, we may succinctly write

$$\langle A(\mathbf{q}\cdot\mathbf{r})\rangle_{\omega} \rightarrow (1/\Omega)\int \rho(\mathbf{r};\boldsymbol{\omega})\,A(\mathbf{q}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r} \qquad (16)$$

for both equations and let A represent either the sine or the cosine function. The quantity Ω is in any case independent of A.

First, we recognize that the right-hand side of equation (16) is the arithmetic average of $A(\mathbf{q}\cdot\mathbf{r})$ weighted by the function $\rho(\mathbf{r};\boldsymbol{\omega})$, whereas the left-hand side of said equation is the average along the Markov chain in $\mathbb{R}^3$. To prove that the left-hand side converges to the right-hand side, we must show that the Markov chain generates points in $\mathbb{R}^3$ distributed according to $\rho(\mathbf{r};\boldsymbol{\omega})$.

Now consider two spatial points $\mathbf{r}$ and $\mathbf{r}'$ that denote different states of the Markov chain (a state is understood to be defined completely by the spatial location of the random walker). Let the population of random walkers at $\mathbf{r}, \mathbf{r}'$ be $\mathcal{P}_n(\mathbf{r})$, $\mathcal{P}_n(\mathbf{r}')$ after n steps. A walker at position $\mathbf{r}$ will leave for position $\mathbf{r}'$ with a probability $P(\mathbf{r}'\mid\mathbf{r})$; likewise, a walker at position $\mathbf{r}'$ will leave for position $\mathbf{r}$ with a probability $P(\mathbf{r}\mid\mathbf{r}')$. The net flux of walkers between $\mathbf{r}$ and $\mathbf{r}'$ should be zero when the distribution of random walkers is equal to $\rho(\mathbf{r};\boldsymbol{\omega})$, hence the equilibrium conditions

$$\mathcal{P}_n(\mathbf{r})\,P(\mathbf{r}'\mid\mathbf{r}) - \mathcal{P}_n(\mathbf{r}')\,P(\mathbf{r}\mid\mathbf{r}') \rightarrow 0 \qquad (17)$$

and

$$\mathcal{P}_n(\mathbf{r}),\,\mathcal{P}_n(\mathbf{r}') \rightarrow \rho(\mathbf{r};\boldsymbol{\omega}),\,\rho(\mathbf{r}';\boldsymbol{\omega}) \qquad (18)$$

as $n\rightarrow\infty$. In other words,

$$\frac{\rho(\mathbf{r};\boldsymbol{\omega})}{\rho(\mathbf{r}';\boldsymbol{\omega})} = \frac{\mathcal{P}_\infty(\mathbf{r})}{\mathcal{P}_\infty(\mathbf{r}')} = \frac{P(\mathbf{r}\mid\mathbf{r}')}{P(\mathbf{r}'\mid\mathbf{r})}. \qquad (19)$$

Let us now factor the transition probability into separate trial, T, and acceptance, Y, probabilities: $P(\cdot) = T(\cdot)Y(\cdot)$. The trial probability is the probability of attempting a move from one state to another, whereas the acceptance probability is the probability of accepting such a move. Invoking the detailed balance condition $T(\mathbf{r}\mid\mathbf{r}') = T(\mathbf{r}'\mid\mathbf{r})$, this transforms equation (19) into

$$\frac{\rho(\mathbf{r};\boldsymbol{\omega})}{\rho(\mathbf{r}';\boldsymbol{\omega})} = \frac{\mathcal{P}_\infty(\mathbf{r})}{\mathcal{P}_\infty(\mathbf{r}')} = \frac{Y(\mathbf{r}\mid\mathbf{r}')}{Y(\mathbf{r}'\mid\mathbf{r})}. \qquad (20)$$

Hence, if we set

$$Y(\mathbf{r}\mid\mathbf{r}') = \begin{cases} \rho(\mathbf{r};\boldsymbol{\omega})/\rho(\mathbf{r}';\boldsymbol{\omega}), & \rho(\mathbf{r};\omega) < \rho(\mathbf{r}';\omega) \\ 1, & \rho(\mathbf{r};\omega) \geq \rho(\mathbf{r}';\omega) \end{cases} \qquad (21)$$

then equation (19) is satisfied and equations (8) and (9) proved.

Acknowledgements

This work was supported by the Swedish Research Council (Vetenskapsrådet). JB gratefully acknowledges financial support from the Knut & Alice Wallenberg Foundation.

References

Almendral, J. M. (2013). *Structure and Physics of Viruses*, edited by M. G. Mateu, ch. 10. Heidelberg: Springer.
Ashcroft, N. W. & Lekner, J. (1966). *Phys. Rev.* **145**, 83–90.
Baravian, C., Michot, L. J., Paineau, E., Bihannic, I., Davidson, P., Impéror-Clerc, M., Belamie, E. & Levitz, P. (2010). *Europhys. Lett.* **90**, 36005.
Bezrukov, A. & Stoyan, D. (2006). *Part. Part. Systems Charact.* **23**, 388–398.
Blum, L. & Stell, G. (1979). *J. Chem. Phys.* **71**, 42–46.
Chen, Q., Whitmer, J. K., Jiang, S., Bae, S. C., Luijten, E. & Granick, S. (2011). *Science*, **331**, 199–202.
Chen, S.-H. (1986). *Annu. Rev. Phys. Chem.* **37**, 351–399.
Debye, P. (1915). *Ann. Phys.* **351**, 809–823.
Diebold, M. P. (2014). *Application of Light Scattering to Coatings.* Heidelberg: Springer.
Donev, A., Cisse, I., Sachs, D., Variano, E. A., Stillinger, F. H., Connelly, R., Torquato, S. & Chaikin, P. M. (2004). *Science*, **303**, 990–993.
Frenkel, D., Vos, R. J., de Kruif, C. G. & Vrij, A. (1986). *J. Chem. Phys.* **84**, 4625–4630.
Guinier, A. (1939). *Ann. Phys. (Paris)*, **12**, 161–237.
Hansen, S. (1990). *J. Appl. Cryst.* **23**, 344–346.
Hansen, S. (2011). *J. Appl. Cryst.* **44**, 265–271.
Hansen, S. (2012). *J. Appl. Cryst.* **45**, 381–388.
Hansen, S. (2013). *J. Appl. Cryst.* **46**, 1008–1016.
Heller, W. T. (2006). *J. Appl. Cryst.* **39**, 671–675.
Henderson, S. J. (1996). *Biophys. J.* **70**, 1618–1627.
Hess, O., Loose, W., Weider, T. & Hess, S. (1989). *Physica B*, **156**, 505–507.
Johnson, C. S. & Gabriel, D. A. (1995). *Laser Light Scattering.* New York: Dover.
Kern, N. & Frenkel, D. (2003). *J. Chem. Phys.* **118**, 9882–9889.
Letz, M. & Latz, A. (1999). *Phys. Rev. E*, **60**, 5865–5871.
Liu, Y., Li, W., Perez, T., Gunton, J. D. & Brett, G. (2011). *Langmuir*, **28**, 3–9.
Lu, R. (2016). Editor. *Light Scattering Technologoy for Food Property, Quality and Safety Assessment.* London: CRC Press.
Marianayagam, N. J., Sunde, M. & Matthews, J. M. (2004). *Trends Biochem. Sci.* **29**, 618–625.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087–1092.
Olds, D. P. & Duxbury, P. M. (2014). *J. Appl. Cryst.* **47**, 1077–1086.

Pedersen, J. S. (1997). *Adv. Colloid Interface Sci.* **70**, 171–210.
Percus, J. K. & Yevick, G. J. (1958). *Phys. Rev.* **110**, 1–13.
Perera, A., Kusalik, P. G. & Patey, G. N. (1987). *J. Chem. Phys.* **87**, 1295–1306.
Perram, J. W. & Wertheim, M. S. (1985). *J. Comput. Phys.* **58**, 409–416.
Saitô, N. & Ikeda, Y. (1951). *J. Phys. Soc. Jpn*, **6**, 305–308.
San Martín, C. (2013). *Structure and Physics of Viruses*, edited by M. G. Mateu, ch. 11. Heidelberg: Springer.
Schmidt-Rohr, K. (2007). *J. Appl. Cryst.* **40**, 16–25.
Sciortino, F., Giacometti, A. & Pastore, G. (2009). *Phys. Rev. Lett.* **103**, 237801.
Sciortino, F., Giacometti, A. & Pastore, G. (2010). *Phys. Chem. Chem. Phys.* **12**, 11869–11877.
Sjöberg, B. (1999). *J. Appl. Cryst.* **32**, 917–923.
Svergun, D. I. & Koch, M. H. J. (2003). *Rep. Progr. Phys.* **66**, 1735.
Thiele, E. (1963). *J. Chem. Phys.* **39**, 474–478.
Vrij, A. (1979). *J. Chem. Phys.* **71**, 3267–3270.
Walther, A. & Müller, A. H. E. (2013). *Chem. Rev.* **113**, 5194–5261.
Wax, A. & Backman, V. (2010). *Biomedical Applications of Light Scattering*. New York: McGraw-Hill.
Wertheim, M. S. (1963). *Phys. Rev. Lett.* **10**, 321–323.
Weyerich, B., Brunner-Popela, J. & Glatter, O. (1999). *J. Appl. Cryst.* **32**, 197–209.
Xu, R. (2000). *Particle Characterization: Light Scattering Methods*. Dordrecht: Kluwer.
Zimm, B. H. (1945). *J. Chem. Phys.* **13**, 141–145.
Zimm, B. H. (1948). *J. Chem. Phys.* **16**, 1093–1098.