Complex SANS/SAXS Data Evaluation, Simulation and Interpretation with Regard to Statistical Inference

vorgelegt von
Diplom-Ingenieur
Michael Muthig
aus Berlin

Von der Fakultät II - Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akadamischen Grades
Doktor der Ingenieurwissenschaften
Dr.-Ing.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Thorsten Ressler
Gutachter: Prof. Dr. Michael Gradzielski
Gutachter: Dr. Julian Oberdisse

Tag der wissenschaftlichen Aussprache: 23. Oktober 2014

Berlin 2015

D 83

Zusammenfassung

Die vorliegende Arbeit gibt einen grundlegenden Diskurs zu unterschiedlichen Methoden zur Auswertung von Kleinwinkelstreudaten (SAS Daten), insbesondere hier zum ersten Mal in einem präzise statistisch definierten Rahmen.

Die Ermittlung von Strukturinformationen via Anpassungen von physikalischen und Freiform-Modellen wird in zwei unterschiedlichen statistischen Inferenz-Ansätzen diskutiert, einem Bayes'schen und einem frequentistischen Ansatz. Somit basiert eine Auswertung und Interpretation der SAS Daten auf gut fundierten Theorien, die zudem aufzeigen wie optimale Schlussfolgerungen zu ziehen sind. Die Diskussion zeigt, dass es wichtig ist genug Informationen über das streuende System (a priori Strukturinformation und Information in den Streudaten) zur Verfügung zu haben, um zugrundeliegende Strukturen zu ermitteln. Beispielhaft wird eine simultane Anpassung eines physikalischen Modells an Kontrastvariationsdaten eines Inter-Polyelektrolyt-Komplexes diskutiert. Darüber hinaus werden die statistischen Inferenz-Ansätze auf die Indirekte Fourier Transformation (IFT) angewendet, um objektiv eine Freiform-Lösung zu erhalten, und moderne Methoden des Maschinellen Lernens (RVM, SVR, LASSO) werden benutzt, um eine stabilere Lösung zu ermitteln.

Ein neues, selbst entwickeltes Programm (SASET) wird präsentiert, das es erlaubt effizient umfangreiche und gekoppelte 1-dimensionale Datensätze/-serien auszuwerten; somit ermöglicht das Programm viele Informationen in den Auswertungsprozess einfließen zu lassen. Auch können 2-dimensionale Datensätze/-serien effizient ausgewertet werden. Außerdem wird dargelegt wie Strukturinformationen aus 2-dimensionalen anisotropen Streudaten gewonnen werden können.

Die Bildung von Vesikeln wird durch die Differentialgleichung von Smoluchowski simuliert. Experimentelle SAS Daten eines Vesikel-bildenden Systems werden simultan angepasst, somit viel Information verwendet, um wenige Parameter zuverlässig zu ermitteln.

Streuintensitäten von komplex, hierarchisch strukturierten kolloidalen Systemen werden mittels Monte Carlo (MC) Simulationen analysiert. Die MC Simulationen zeigen welche strukturellen Informationen in SAS Daten enthalten sind.

Abstract

This work presents an in-depth discussion of different methods used for Small-Angle Scattering (SAS) data analysis, and especially for the first time within a precisely defined statistical framework.

Inferring structural information from SAS data via physical model fitting and free-form model fitting is discussed within two different statistical inference approaches, namely Bayesian and frequentist statistics, which put the analysis and interpretation of SAS data on well founded theories, hence showing how to draw optimal inferences. The discussion shows the importance of having enough of scattering system information (a priori knowledge about the system and information contained in the experimental SAS data) available in order to infer structural information. As an example, simultaneous physical model fitting is performed on contrast variation data of an InterPolyElectrolyte Complex (IPEC) system. Moreover, statistical inference is applied to the Indirect Fourier Transform (IFT) in order to get objectively a free-form solution, and additionally, modern machine learning methods (RVM, SVR, LASSO) are employed to determine a more robust solution.

A new and homemade program, called SASET, is presented that easily allows to efficiently evaluate comprehensive and coupled 1-dimensional SAS data sets/series, hence allowing to include a lot of information in the evaluation process. 2-dimensional data sets/series can also

be evaluated efficiently. Moreover, a discussion about inferring structural information from anisotropic 2-dimensional scattering data is given.

Vesicle formation is simulated by the von Smoluchowski differential equation. Experimental SAS data of a vesicle building system are simultaneously fitted by adjusting parameters of the kernel of the differential equation, and therefore a lot of information is used to determine reliably a few parameters.

The scattering intensity of complex, hierarchically structured colloidal systems is analyzed by Monte Carlo (MC) simulations. The MC simulations show which structural information are contained in the SAS data.

Acknowledgement

First and foremost I want to thank my advisor, Prof. Dr. Michael Gradzielski, who offered me a fantastic opportunity to become an interdisciplinary researcher. He always supported me in many respects, gave me important advices and valuable inspiration that could only come from an open-minded, and very positive person like him.

I am very thankful to Dr. Julian Oberdisse for reviewing my thesis.

I feel particularly obliged to the DFG for the financial support of this work (DFG GR1030/11-1, DFG OR99/6-1).

Cordial thanks go to Dr. Sylvain Prévost, who patiently discussed many scattering techniques, evaluations and software issues with me, thus helping me a lot with proceeding and enhancing my work. In addition to it, I feel fortunate for his amiable and helpful attitude and relaxing time outside work.

I am very grateful to Dr. Katharina Bressel and Dr. Raphaël Michel for their enthusiastic and fruitful collaboration, and many thanks go to my student assistant Ana Kosareva, who has done some indispensable software programming for me.

Special thanks go to Dr. René Strassnick, who created a pleasant atmosphere and a green oasis in our common office.

Notably, I would like to thank my wonderful friends and colleagues in the department and at the TUB.

Finally, I would like to appreciate and thank my mum Ursel, brother Manuel and girlfriend Henriette for their support and love. I would like to dedicate this work to my mum, who gave me the opportunity and support to develop my interests and skills. Without her, this work would not have been possible.

Contents

1. Introduction

Small-Angle Scattering (SAS) of neutrons (SANS) or X-rays (SAXS) is a well established method to probe and analyze nanomaterials (e.g., nanoparticles or colloidal soft-matter). Nanomaterials are in between the size of atoms and bulk materials, hence the properties of nanomaterials can be totally different from the ones known from bulk materials. Therefore nanomaterials are interesting for a number of new applications, e.g., within nanomedicine, cosmetics, paints, electronics, etc., but with their usage in everyday products it is also very important to understand and analyze their nanotoxicity. By means of SAS sample systems are analyzed in situ, which, e.g., allows to study dynamics of a nanomaterial system or to probe it in an external field (electric or magnetic field, flow field, etc.). Often SAS is the only method to perform these analyses, and hence it is one of the most important tools for such research topics. Another important feature of SAS is that the scattering intensity represents the scattering from the illuminated system volume. Therefore, structural parameters that are inferred represent averages from a huge number of scattering objects, i.e., they are statistically meaningful.

However, determining structural information from the scattering intensity is an inverse problem, which is ill-posed due to the loss of phase information, finite resolution, smearing effects, noise, etc., and therefore a priori knowledge is required to find plausible solutions. A priori knowledge could be that, e.g., the shape or size of the scattering objects is known, experimental boundary conditions like volume fractions, densities, etc. are known, but also in case of a dynamic system that it changes continuously or that a system responds continuously in dependence of an applied external field. The inference process itself can be improved by performing the method of contrast variation. I.e., a number of different scattering patterns can be created by changing the scattering length density of one or more components in order to emphasize certain structures of the particles in the scattering intensity, thus to infer these structures more robustly — hence, via contrast variation the available information content in the data is increased.

This thesis is about evaluations and simulations of complex SANS/SAXS data. The inference procedure is mainly discussed in a statistical framework, either a Bayesian or a frequentist one, in order to have a well established theory at hand to perform plausible reasoning and to draw optimal inferences. Considering the SAS evaluation within a statistical framework is a logical consequence of the fact that the data are uncertain and system information is incomplete. The two statistical inference approaches are introduced in section 2 together with an introduction of the theory of SAS.

In order to easily and efficiently evaluate comprehensive 1- and 2-dimensional SAS data sets or data set series, which might result from dynamic experiments and/or from using the method of contrast variation, many new and important features have been added to the scattering evaluation software SASET, which has been initially implemented by the author in his diploma thesis. One of the new and noteworthy features of SASET is that the software allows to effortlessly perform model fitting to coupled 1-dimensional scattering series (e.g., series of different contrasts or different experimental set-ups), hence allowing to conveniently increase the information content of a SAS evaluation, and therefore to make the analysis more robust. In general, the analysis of comprehensive data sets/series was the bottleneck in the evaluation process so far, but can now be comfortably and efficiently done with SASET. Moreover, in order to speed up the evaluation process, parallel methods have been implemented. SASET is presented in section 3 together with an example evaluation of a contrast variation data set of an InterPolyElectrolyte Complex (IPEC) system, having the goal to infer the IPEC structure.

In section 4 complex colloidal systems, consisting of particles decorating vesicles and/or internalized in them, are simulated in order to study their scattering behavior. In these simulations the method of Monte-Carlo sampling is applied, which is a standard method within physical chemistry to simulate particle configurations, but also within statistical inference, where it is used to create samples from a probability distribution, and to calculate averages of some desired quantities with respect to this distribution.

The subject of spontaneous vesicle formation from mixing different amphiphiles is studied in section 5. In this study it is considered that the vesicles are built from relatively large disks that close to form vesicles. The disks themselves are built from smaller disks that coalesce. It is considered that the whole coalescence and vesiculation process is controlled by a few initial parameters that build harsh constraints. The analysis shows which parameters have an impact on the scattering intensity. Moreover an experimentally measured scattering data series of a spontaneously forming vesicle system is fitted by the considered model, i.e., the whole intensity series is simultaneously fitted by a model having only a few parameters. Hence, a lot of information is used to determine reliably a few parameters.

The evaluation of 2-dimensional anisotropic scattering images, often resulting from the application of an external field (shear, magnetic or electric field, etc.) on the scattering system, is a demanding task because of the huge amount of data. However, determining only the anisotropy within an image can be efficiently done if the analysis is restricted to certain regimes in an image. In section 6 different methods of quantifying the anisotropy in an image are studied and compared with each other.

In section 7, the Indirect Fourier Transform (IFT) is discussed within the two statistical inference approaches, the Bayesian one and the frequentist one. The IFT determines the pair distance distribution function from the scattering intensity by performing a regularized inverse Fourier transform. While the IFT was already discussed in a Bayesian framework elsewhere, the method is discussed here in a frequentist one. Moreover modern machine learning approaches, which can be seen as special approaches within statistical inference, are applied with the goal to get a more robust solution.

Hence, this work presents all possible analyses that can be performed on SAS data, from physical model fitting to model-free analysis, demonstrating how these methods can be efficiently applied and why a well-defined statistical framework is necessary for a proper data interpretation resulting in robust and reliable scientific findings.

The last section finishes with an overall conclusion and outlook.

2. Theory

This section introduces and discusses the main principles of small-angle scattering and of statistical inference, which are required for this work.

2.1. Small-Angle Scattering

In this chapter the principles of Small-Angle Scattering (SAS) are explained. For a rigorous discussion the reader is referred to [52, 58, 144, 94, 41, 133].

In SAS a primal plane wave of neutrons (SANS) or photons (SAXS) strikes a sample, cf. Fig. 1. Then, the wave interacts within the sample with the nuclei of the atoms in case of a wave of neutrons, and mainly with the electrons in case of a wave of photons[1]. Hence, the interaction strength in any point of space can be described by the so called Scattering Length Density (SLD), and the scattering mechanism is according to the Huygens-Fresnel principle that states: Each point in space that is covered by the primal wave is the source of a secondary, spherical wave, whereby the amplitude of the secondary wave in its source point is proportional to the SLD in the source point. Moreover, the amplitude of any receiving point in space is given by the superposition of all secondary waves.

If the far-field condition is assumed, i.e., if the receiving point is relatively far away from the sample, the superposition of the secondary waves in this point is essentially given by a Fourier Transform (FT) of the SLD[2]:

$$A(\mathbf{q}) = \int_V \mathrm{SLD}(\mathbf{r}) \exp\left(-i\mathbf{q}^T\mathbf{r}\right)\, \mathrm{d}\mathbf{r}. \tag{1}$$

V is the sample volume and $\mathbf{q}$ is the scattering vector in the reciprocal space, which is defined as

$$\mathbf{q} := \mathbf{k_R} - \mathbf{k_0}, \tag{2}$$

where $\mathbf{k_0}$ is the wave vector of the incident wave and $\mathbf{k_R}$ is the wave vector in the direction $\mathbf{R}$. The magnitude of $\mathbf{q}$ is given by

$$q = \frac{4\pi}{\lambda} \sin\left(\frac{\vartheta}{2}\right), \tag{3}$$

where ϑ is the angle enclosed between $\mathbf{k_R}$ and $\mathbf{k_0}$, i.e., $\vartheta = \sphericalangle(\mathbf{k_R}, \mathbf{k_0})$. Finally, the intensity measured by a detector is given by[3]

$$I(\mathbf{q}) = A(\mathbf{q}) \cdot A^*(\mathbf{q}). \tag{4}$$

Thus, if the intensity is measured any information about the phase of the complex amplitude is lost. In case of a real amplitude the sign of the amplitude is lost.

If the definition of the amplitude, Eq. (1), is substituted into Eq. (4), the intensity can be written as the FT of the autocorrelation $\gamma(\mathbf{r})$ of the SLD [133, §1.5.2]:

$$I(\mathbf{q}) = \int_V \gamma(\mathbf{r}) \exp\left(-i\mathbf{q}^T\mathbf{r}\right)\, \mathrm{d}\mathbf{r}, \tag{5}$$

where

$$\gamma(\mathbf{r}) = \int_V \mathrm{SLD}(\mathbf{u}) \cdot \mathrm{SLD}(\mathbf{u}+\mathbf{r})\, \mathrm{d}\mathbf{u}. \tag{6}$$

[1] The interaction of the photons with the protons is negligible, mainly because a proton is much heavier than an electron.

[2] The symbol ^{T} in the formula means the transpose of a vector, and $i^2 = -1$.

[3] The symbol * in the formula means conjugate complex.

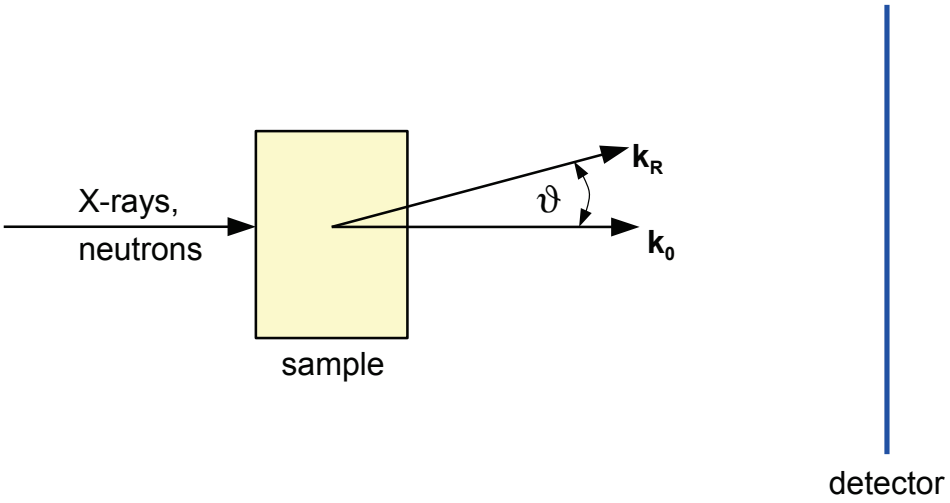


Figure 1: The scattering experiment.

The autocorrelation of the SLD also has an inherent loss of information, such that the SLD cannot uniquely be reconstructed from the autocorrelation. The relationships between the SLD and the intensity are completely summarized in Fig. 2, and hence the following three approaches can be conceived to determine the SLD profile from measured intensities:

1. (*i*) Determine the autocorrelation function from the intensity. (*ii*) Decorrelate the autocorrelation.

2. (*i*) De-square the intensity in order to retrieve the amplitude. (*ii*) Calculate the inverse Fourier transform of the amplitude.

3. (*i*) Build a (physical) model of the SLD and calculate the intensity. (*ii*) Change the model parameters, such that the calculated intensity fits well the measured intensity.

The practical implementation of approach 1 has been mainly developed by Glatter [49, 48, 51], and approach 2 has been mainly developed by Svergun [148, §5]. Approach 3 is probably used since more than a 100 years, but the inventor is unknown.

From the preceding discussion and Fig. 2 it follows that even if the intensity is given analytically over the whole reciprocal space the solution is in general non-unique. Such a problem is called an ill-posed (inverse) problem and is the topic of chapter 7.1. Ill-posed problems often arise if one tries to determine causes from effects, i.e., if one tries to solve an inverse problem. On the other hand calculating the effects from the causes, i.e., calculating the intensity from a given SLD distribution, is called a direct problem and is well posed. For simple models (as, e.g., some spherical objects, see text below and section 3.3.1) the direct solution can often be given as a closed-form expression or it often only needs deterministic numerical integrations. In general, however, more complicated models require Monte-Carlo (MC) simulations (e.g., decorated vesicles or vesicles containing internalized particles as used in section 4, see also [13]) or molecular dynamic simulations (e.g., [80]). Several 'simple' models are given in the paper of Pedersen [123] as well as in the documentations of the program SASfit (Joachim Kohlbrecher, PSI, Villigen, Switzerland, and Ingo Bressler, BAM, Berlin, Germany) and of the program of S. Kline [86] based on IGOR. In this work, many different models have been implemented and are documented in the homemade program SASET or they are discussed within this thesis, e.g., models of complex clusters based on MC simulations — cf. chapter 4.

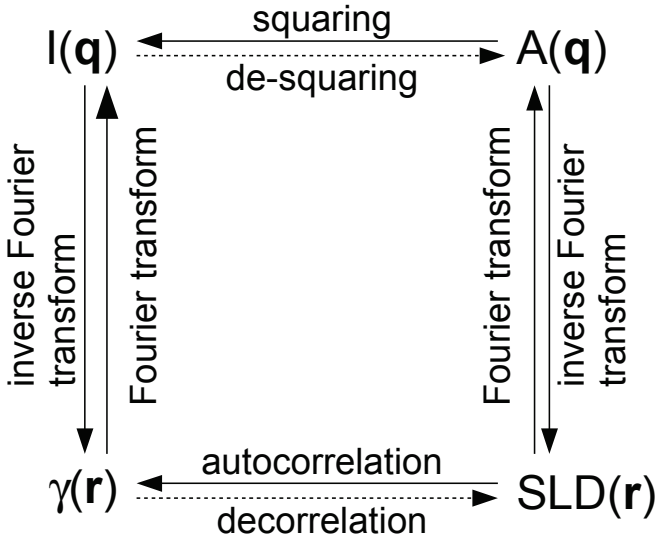


Figure 2: Relation between the SLD and the intensity, cf. [133, §1.5.4]. A dotted arrow indicates an ambiguous solution.

Because calculating the intensity of a given physical model is well-posed, scientific inference is often done in accordance with approach 3 from above, and hence yielding a relatively robust inference procedure. In detail, e.g., if the sample only consists of homogeneous spherical particles of the same size, and the size is looked for, the scattering intensity is calculated from a homogeneous sphere[4] SLD model given a certain sphere radius that is a freely adjustable parameter. The parameter can then be adjusted in a 2-step iterative procedure that is: (*i*) Calculate the intensity for a given radius and (*ii*) change the radius and iterate as long as the intensity is not fitted 'well' by the model or another termination criterion is not fulfilled (e.g., the maximum number of iterations is not reached). Often gradient-search methods can be used to efficiently adjust the model parameters — see, e.g., [14, §8].

In the following the model intensity of a homogeneous sphere with radius R is calculated in order to discuss some important principles. The sphere model as well as the technique to derive the scattering formula is used again in later chapters to build more complicated hierarchical systems. Consider a single homogeneous sphere with volume v_{sphere} within the sample volume V. According to Eq. (1) the amplitude is written as

$$\begin{aligned} A(\mathbf{q}) &= \int_{v_{\text{sphere}}} \text{SLD}_{\text{sphere}} \exp\left(-i\mathbf{q}^T\mathbf{r}\right)\, \mathrm{d}\mathbf{r} + \int_{V\setminus v_{\text{sphere}}} \text{SLD}_{\text{matrix}} \exp\left(-i\mathbf{q}^T\mathbf{r}\right)\, \mathrm{d}\mathbf{r} \\ &= \int_{v_{\text{sphere}}} (\text{SLD}_{\text{sphere}} - \text{SLD}_{\text{matrix}}) \exp\left(-i\mathbf{q}^T\mathbf{r}\right)\, \mathrm{d}\mathbf{r} + \int_{V} \text{SLD}_{\text{matrix}} \exp\left(-i\mathbf{q}^T\mathbf{r}\right)\, \mathrm{d}\mathbf{r} \\ &\overset{V\to\infty}{\to} \int_{v_{\text{sphere}}} \Delta\text{SLD} \exp\left(-i\mathbf{q}^T\mathbf{r}\right)\, \mathrm{d}\mathbf{r} + \text{SLD}_{\text{matrix}}\delta(\mathbf{q}), \end{aligned} \tag{7}$$

where $\text{SLD}_{\text{sphere}}$ is the SLD within the sphere and $\text{SLD}_{\text{matrix}}$ is the SLD of the matrix/solvent and $\Delta\text{SLD} := \text{SLD}_{\text{sphere}} - \text{SLD}_{\text{matrix}}$. $\delta(\mathbf{q})$ is the Dirac delta function. Thus, for $\mathbf{q} \neq \mathbf{0}$ the

[4]Note that the technical term sphere is ambiguous, since in mathematics a sphere is the surface of a ball, whereas here as well as in the SAS-literature it is meant to be a ball in the mathematical sense.

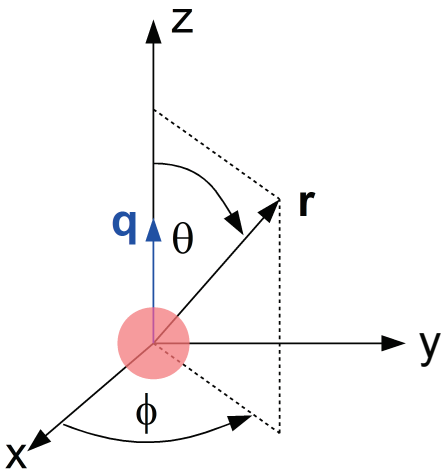


Figure 3: A homogeneous sphere within a Cartesian and spherical coordinate system and the scattering vector $\mathbf{q}$.

amplitude is given by

$$\begin{aligned} A_{\text{sphere}}(\mathbf{q}) &= \int_{v_{\text{sphere}}} \Delta\text{SLD} \cdot \exp\left(-i\mathbf{q}^T\mathbf{r}\right) \,\mathrm{d}\mathbf{r} \\ &= \Delta\text{SLD} \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\int_{r=0}^{R} \exp\left(-iqr\cos\theta\right) r^2 \sin\theta \,\mathrm{d}r \,\mathrm{d}\theta \,\mathrm{d}\phi \\ &= 4\pi\Delta\text{SLD} \int_{r=0}^{R} r^2 \frac{\sin qr}{qr} \,\mathrm{d}r \\ &= 4\pi\Delta\text{SLD} \frac{\sin(qR) - qR\cos(qR)}{q^3} \cdot \frac{3R^3}{3R^3} \\ &= v_{\text{sphere}}(R) \cdot \Delta\text{SLD} \cdot 3\frac{\sin(qR) - qR\cos(qR)}{(qR)^3} \\ &=: A_{\text{sphere}}(q, R), \end{aligned} \tag{8}$$

and is called amplitude form factor of a sphere[5], and

$$v_{\text{sphere}}(R) = \frac{4\pi}{3}R^3 \tag{9}$$

is the volume of the sphere. In Eq. (8) the scattering vector is taken to be aligned along the z-direction such that $\mathbf{q}^T\mathbf{r} = qr\cos\theta$, see Fig. 3, which is always possible since a particle of spherical symmetry is assumed.

By substituting Eq. (8) into Eq. (4) the intensity form factor of a sphere is obtained as

$$I_{\text{sphere}}(q) = A^2_{\text{sphere}}(q, R) = v^2_{\text{sphere}}(R) \cdot (\Delta\text{SLD})^2 \cdot 9\frac{\left(\sin(qR) - qR\cos(qR)\right)^2}{(qR)^6}. \tag{10}$$

From this formula it is obvious that the intensity does not change if the SLD of the sphere is exchanged with the one of the matrix — only the squared SLD difference is important. This

[5] The amplitude form factor is also defined for $\mathbf{q} = \mathbf{0}$, but for $\mathbf{q} = \mathbf{0}$ the (overall) amplitude consists of the amplitude form factor term plus an additional Dirac term, cf. Eq. (7). In the following, the addition 'form factor' is often not written explicitly, but from the context it should be clear that this meant. The same holds for the technical term 'intensity form factor'.

effect is known as the Babinet principle [133, §1.6] and exemplifies the phase/sign loss, as discussed above.

Taking the limit $q \to 0$ of Eq. (10) yields the forward intensity (form factor) of a sphere:

$$I_{\text{sphere}}(0) = A^2_{\text{sphere}}(0, R), \tag{11}$$

where

$$A_{\text{sphere}}(0, R) = v_{\text{sphere}}(R) \cdot \Delta\text{SLD} \tag{12}$$

is the forward amplitude (form factor) of a sphere[6]. The calculation of the forward intensity often gives an easy check if a model is correctly implemented, since the intensity of an implemented model has to approach the forward intensity calculation when $\mathbf{q} \to \mathbf{0}$. Therefore it was usually used to check the models that are implemented within this work.

The mathematical decomposition of the volume as done in Eq. (7) is used in this work to derive formulas of more complex systems, e.g., in section 4.3. If there are different scattering objects it is important to note that the amplitudes of these objects have to be added before the intensity is calculated. Furthermore, the amplitude of an object that is shifted out of the origin is given by the amplitude of the object at the origin times a complex phase lift, since there is the Fourier theorem:

$$\text{SLD}(\mathbf{r} - \mathbf{r}_0) \overset{\text{FT}}{\longleftrightarrow} A(\mathbf{q}) \exp\left(-i\mathbf{q}^T\mathbf{r}_0\right). \tag{13}$$

As an extension of the simple sphere amplitude/intensity form factor given above, a cluster of N_c spheres with different radii R_k is considered. In this case Eq. (8) is used together with Eq. (13) in order to calculate the overall scattering amplitude

$$\begin{aligned} A_{\text{cluster}}(\mathbf{q}) &= \sum_{k=1}^{N_c} A_{\text{sphere}}(q, R_k) \exp\left(-i\mathbf{q}^T\mathbf{r}_k\right) \\ &= \sum_{k=1}^{N_c} A_k(q) \exp\left(-i\mathbf{q}^T\mathbf{r}_k\right), \end{aligned} \tag{14}$$

where

$$A_k(q) := A_{\text{sphere}}(q, R_k), \tag{15}$$

and $\mathbf{r}_k$ is the position of the k-th sphere. Hence, the intensity is

$$\begin{aligned} I_{\text{cluster}}(\mathbf{q}) &= A_{\text{cluster}}(\mathbf{q}) \cdot A^*_{\text{cluster}}(\mathbf{q}) \\ &= \sum_{k=1}^{N_c}\sum_{l=1}^{N_c} A_k(q) A_l(q) \exp\left(-i\mathbf{q}^T\mathbf{r}_{kl}\right), \end{aligned} \tag{16}$$

where $\mathbf{r}_{kl} = \mathbf{r}_k - \mathbf{r}_l$. This expression can be further modified, assuming that the cluster is freely moving (but the relative positions of spheres to each other stay fixed) and that the scattering experiment captures a time average of the scattering intensity. Then, the average

[6] In general, the forward amplitude form factor of an object can be written as $A_{\text{object}}(0) = \int_{v_{\text{obj.}}} \Delta\text{SLD}(\mathbf{r})\, \mathrm{d}\mathbf{r}$, where $v_{\text{obj.}}$ is the volume of the considered object.

intensity is given by an orientational average (indicated by $\langle\cdot\rangle_\Omega$)

$$\begin{aligned}\langle I_{\text{cluster}}(\mathbf{q})\rangle_\Omega &= \langle A_{\text{cluster}}(\mathbf{q})A^*_{\text{cluster}}(\mathbf{q})\rangle_\Omega \\ &= \left\langle \sum_{k=1}^{N_c}\sum_{l=1}^{N_c} A_k(q)A_l(q)\exp\left(-i\mathbf{q}^T\mathbf{r}_{kl}\right)\right\rangle_\Omega \\ &= \sum_{k=1}^{N_c}\left\langle A_k^2(q)\right\rangle_\Omega + \left\langle \sum_{k=1}^{N_c}\sum_{l\neq k}^{N_c} A_k(q)A_l(q)\exp\left(-i\mathbf{q}^T\mathbf{r}_{kl}\right)\right\rangle_\Omega \\ &= \sum_{k=1}^{N_c} A_k^2(q) + 2\sum_{k=1}^{N_c-1}\sum_{l=k+1}^{N_c} A_k(q)A_l(q)\left\langle\exp\left(-i\mathbf{q}^T\mathbf{r}_{kl}\right)\right\rangle_\Omega.\end{aligned} \tag{17}$$

The average over the exponential function yields

$$\begin{aligned}\left\langle\exp\left(-i\mathbf{q}^T\mathbf{r}_{kl}\right)\right\rangle_\Omega &= \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\exp\left(-iqr_{kl}\cos(\theta)\right)\sin(\theta)\ \mathrm{d}\theta\ \mathrm{d}\phi \\ &= \frac{1}{2}\int_0^{\pi}\exp\left(-iqr_{kl}\cos(\theta)\right)\sin(\theta)\ \mathrm{d}\theta \\ &= \frac{1}{2}\frac{\exp\left(iqr_{kl}\right)-\exp\left(-iqr_{kl}\right)}{iqr_{kl}} \\ &= \frac{\sin(qr_{kl})}{qr_{kl}}\end{aligned} \tag{18}$$

where in the second line the substitution $u = -\cos(\theta)$ is used to perform the integration. The final result is

$$\begin{aligned}\langle I_{\text{cluster}}(\mathbf{q})\rangle_\Omega &= \sum_{k=1}^{N_c} A_k^2(q) + 2\sum_{k=1}^{N_c-1}\sum_{l=k+1}^{N_c} A_k(q)A_l(q)\frac{\sin(qr_{kl})}{qr_{kl}} \\ &= \sum_{k=1}^{N_c}\sum_{l=1}^{N_c} A_k(q)A_l(q)\frac{\sin(qr_{kl})}{qr_{kl}},\end{aligned} \tag{19}$$

and is known as Debye formula [35]. The second term on the r.h.s. of the first line defines a structure factor term of the interacting spheres that is negligible if the spheres do not interact with each other, i.e., if $|A_k(q)A_l(q)\sin(qr_{kl})| \ll qr_{kl} \quad \forall k,l$. Often this equation is written as

$$\langle I_{\text{cluster}}(\mathbf{q})\rangle_\Omega = P_{\text{cluster}}(q)S^{\text{eff.}}_{\text{cluster}}(q), \tag{20}$$

where

$$P_{\text{cluster}}(q) = \sum_{k=1}^{N_c} A_k^2(q) \tag{21}$$

defines an intensity form factor, and

$$S^{\text{eff.}}_{\text{cluster}}(q) = 1 + \frac{2}{\sum_{k=1}^{N_c} A_k^2(q)}\sum_{k=1}^{N_c-1}\sum_{l=k+1}^{N_c} A_k(q)A_l(q)\frac{\sin(qr_{kl})}{qr_{kl}} \tag{22}$$

describes an effective structure factor.

In general, the positions of the particles depend on an interaction potential of the particles, external forces (electric/magnetic field, shear forces), and the temperature of the system.

Usually, determining a plausible configuration of such a system is non-trivial, and analytical expressions of structure factors are approximations, and are only available for a few simple cases, e.g., for monodisperse or polydisperse hard spheres with a hard sphere structure factor (cf. section 3.3 together with appendices D.1 and D.2). In general, structure factors are only available for simple symmetries (mainly isotropic, i.e., for spherical potentials) and are approximations based on the choice of a closure relation (Percus-Yevick PY [124] (cf. appendix D.1,D.2), HyperNetted Chain HNC (e.g., in [135])) or a mixture of closure-relations to reduce thermodynamical inconsistencies (e.g., Rogers-Young [134]). In more complex cases only Monte Carlo (MC) methods or molecular dynamics (MD) simulations can determine the structure factor of a given system. In section 4 the MC method is used to determine structure factors of complex clusters consisting of vesicles that are decorated with particles and/or contain internalized particles.

The Debye formula Eq. (19) can be used to derive another scattering property. For this, consider a system of N/N_c identical clusters, and all spheres are assumed to be identical. N is the total number of particles. Then, the total scattering intensity of such a system is obtained by multiplying the Debye formula Eq. (19) with N/N_c, yielding

$$\langle I(\mathbf{q})\rangle_\Omega = \frac{N}{N_c} A^2_{\text{sphere}}(q,R) \sum_{k=1}^{N_c}\sum_{l=1}^{N_c} \frac{\sin(qr_{kl})}{qr_{kl}}, \tag{23}$$

and the forward intensity reads

$$\begin{aligned}\lim_{q\to 0}\langle I(\mathbf{q})\rangle_\Omega &= \frac{N}{N_c}\lim_{q\to 0}\left(A^2_{\text{sphere}}(q,R)\right)\cdot\lim_{q\to 0}\left(\sum_{k=1}^{N_c}\sum_{l=1}^{N_c}\frac{\sin(qr_{kl})}{qr_{kl}}\right)\\ &= \frac{N}{N_c}\cdot\left(v_{\text{sphere}}(R)\cdot\Delta\text{SLD}\right)^2\cdot N_c^2\\ &= N\cdot N_c\cdot\left(v_{\text{sphere}}(R)\cdot\Delta\text{SLD}\right)^2,\end{aligned} \tag{24}$$

where Eq. (11) has been used, and $\lim_{q\to 0}\sin(qr_{kl})/(qr_{kl}) = 1$. Thus showing that the forward intensity scales proportionally with the number of particles in a cluster N_c, though the total number of particles N stays constant.

Last but not least it is mentioned that in a scattering experiment there is an illuminated scattering volume, which is taken into account by the subsequent transmission correction (besides other transmission corrections, see, e.g., [121, 146]). Accordingly, the modeled scattering intensity has to be normalized by the volume of the considered system , i.e.,

$$I^n(q) = \frac{I(q)}{V}. \tag{25}$$

E.g., the normalized scattering intensity of the collection of clusters described in Eq. (23) becomes

$$I^n(q) = nA^2_{\text{sphere}}(q,R)\frac{1}{N_c}\sum_{k=1}^{N_c}\sum_{l=1}^{N_c}\frac{\sin(qr_{kl})}{qr_{kl}}, \tag{26}$$

where

$$n = \frac{N}{V} \tag{27}$$

is the number density of particles.

2.2. Statistical Inference

A comprehensive and important part of this work is about regression, i.e., a function shall be fitted to some noisy data. E.g., (*i*) the fitted function may represent a physical model, whereby the determined parameters represent some physical quantities (parametric model), or (*ii*) the fitted function is a free-form solution, i.e., the determined parameters do not have a physical meaning, but the functional behavior is of interest. Because of the noise in the data and the finite data set size, i.e., there are uncertainties incorporated, methods from statistical inference[7] have to be applied in order to perform an optimal regression and in order to perform model selection.

There are mainly two different schools of statistical inference, the Bayesian and the frequentist one. These two approaches differ in their interpretation of the notion of probability. In the Bayesian school probability represents the degree of belief (degree of plausibility) of a person about something, while in the frequentist school probability represents the long-run relative frequency with which an event occurs [141, §1.4]. In order to exemplify these two approaches let us consider the following: there is an observed data set $\mathcal{D} = \{(x_n, y_n)\}_{n=1}^{N}$ and a regression model $f(x, \mathbf{w})$ having the adjustable parameter vector $\mathbf{w}$ containing M scalar parameters w_m. x_n is an abscissa value and y_n its appropriate ordinate value, cf. Fig. 4. Then, for regression — case (*iii*) in Fig. 4:

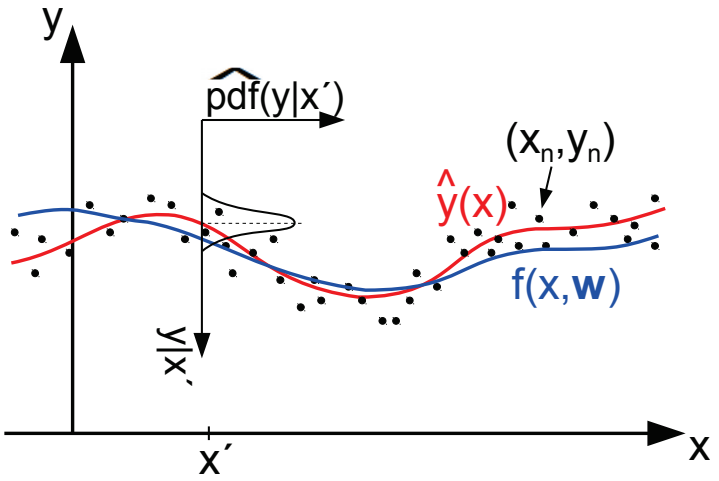


Figure 4: Inference example. It is considered that data (x, y) are generated from the (true) probability distribution $\widehat{\text{pdf}}(x, y)$. At position x' the data distribution with respect to y is described by the (true) conditional distribution $\widehat{\text{pdf}}(y|x')$, which is given by the (true) deterministic function $\hat{y}(x)$ and additional noise ϵ, i.e., $y|x' = \hat{y}(x') + \epsilon$. Then, inference goals can be (*i*) to determine the joint probability distribution $\widehat{\text{pdf}}(x, y)$, in this case the overall data generation process is inferred, (*ii*) to determine $\widehat{\text{pdf}}(y|x')$, in this case the generation of y given x' is inferred or (*iii*) to infer only the (true) deterministic function $\hat{y}(x)$, in this case a noise model is assumed, and then a function $f(x, \mathbf{w})$ is fitted to the random data set $\{(x_n, y_n)\}_{n=1}^{N}$ with the goal to have $f(x, \mathbf{w}) = \hat{y}(x)$, i.e., regression is performed. Note, in the text of this chapter a data set $(\mathbf{x}, \mathbf{y})$ is considered to be a sample from the distribution $\widehat{\text{pdf}}(\mathbf{x}, \mathbf{y})$, which describes the more general case that data in the set can be interdependent.

[7] Statistical inference (also called scientific inference) means the process of drawing (optimal) conclusions from uncertain data and hypotheses.

Frequentist inference: It is considered that there is a fixed and true parameter vector $\hat{\mathbf{w}}$ (e.g., some fixed physical quantities). Because of the uncertainty within the data, the inference will yield a point estimate of $\hat{\mathbf{w}}$, namely $\tilde{\mathbf{w}}$. Confidence intervals (error-bars / uncertainties) in $\tilde{\mathbf{w}}$ can be determined by considering the distribution of further estimates obtained from independent and identically distributed (i.i.d.) data sets [129, §15.6].

Bayesian inference: It is considered that there is a distribution of possible parameter vectors before the data set $\mathcal{D}$ is observed, which represents the degree of belief about the parameters in oneself (i.e., the distribution represents the state of knowledge a person has about the parameters). This a priori distribution is then adjusted according to the observed data set, yielding an a posteriori distribution of the parameter vector (i.e, one gains refined knowledge).

The Bayesian inference formula can be derived from the product rule of probability theory[8], and reads as follows

$$\mathrm{pdf}(X|Y) = \frac{\mathrm{pdf}(Y|X)\mathrm{pdf}(X)}{\mathrm{pdf}(Y)}. \tag{28}$$

$\mathrm{pdf}(X|Y)$ can be interpreted as the a posteriori distribution function[9] of an event X (or a hypothesis being true) given that the event Y (data) has been observed. The a priori probability distribution function about X is given by $\mathrm{pdf}(X)$, and $\mathrm{pdf}(Y|X)$ is the conditional probability distribution function for Y under the assumption that X is true and is called likelihood (function). The denominator on the r.h.s. of this equation is the marginal probability of Y, which is obtained by integrating the numerator over X (i.e., applying the sum rule of probability theory); it ensures normalization. It is worth emphasizing that the a priori probability and the likelihood are built on some available background information I. Therefore one should indicate this issue by accurately writing the Bayes formula as $\mathrm{pdf}(X|Y,I) = \mathrm{pdf}(Y|X,I)\mathrm{pdf}(X|I)/\mathrm{pdf}(Y|I)$, but it is not done here, in order to keep the notation uncluttered. However, all the available information (i.e., constraints on X and dependence of Y given X) should be incorporated into these functions in order to perform the inference as good as possible.

If X and Y are substituted by $\mathbf{w}$ and D from the regression example above, the Bayes formula reads

$$\mathrm{pdf}(\mathbf{w}|\mathcal{D}) = \frac{\mathrm{pdf}(\mathcal{D}|\mathbf{w})\mathrm{pdf}(\mathbf{w})}{\mathrm{pdf}(\mathcal{D})}, \tag{29}$$

allowing us to infer the a posteriori parameter distribution of $\mathbf{w}$ given the data $\mathcal{D}$ and given some background information I (not explicitly mentioned here). In order to get a more appropriate likelihood function for our regression problem (in which the ordinate values are generated in dependence of the abscissa values and the parameter vector — case (*iii*), respectively case (*ii*), shown in Fig. 4), the product rule of probability theory can be applied on the r.h.s. of Eq. (29) in order to get

$$\frac{\mathrm{pdf}(\mathbf{x},\mathbf{y}|\mathbf{w})\mathrm{pdf}(\mathbf{w})}{\mathrm{pdf}(\mathbf{x},\mathbf{y})} = \frac{\mathrm{pdf}(\mathbf{y}|\mathbf{x},\mathbf{w})\mathrm{pdf}(\mathbf{x}|\mathbf{w})\mathrm{pdf}(\mathbf{w})}{\mathrm{pdf}(\mathbf{y}|\mathbf{x})\mathrm{pdf}(\mathbf{x})}, \tag{30}$$

[8] The product rule of probability theory reads as follows $\mathrm{pdf}(X,Y) = \mathrm{pdf}(Y|X)\mathrm{pdf}(X) = \mathrm{pdf}(X|Y)\mathrm{pdf}(Y)$, and there is another important rule, called sum rule, that reads $\mathrm{pdf}(X) = \sum_Y \mathrm{pdf}(X,Y)$ [16, §1]. Herein, $\mathrm{pdf}(X)$ and $\mathrm{pdf}(Y)$ are the probability distribution functions of X and Y, $\mathrm{pdf}(X,Y)$ is the joint distribution of X and Y, and $\mathrm{pdf}(Y|X)$ is the conditional distribution of Y given X. The rules can be seen [77] as an extension of the Boolean logic for the case where reasoning has to be done from incomplete information, and they are unique to perform consistent reasoning [31] — see also, e.g., [141, §1.2-1.3 and appendix B].

[9] With $\mathrm{pdf}(x)$ we mean the probability distribution function of x. However, in case of continuous x, it is actually a probability density function.

where the vector $\mathbf{x}$ contains the abscissa values $\{x_n\}_{n=1}^N$, and the vector $\mathbf{y}$ the ordinate values $\{y_n\}_{n=1}^N$. Since in the regression problem above, we are looking for a parametric model for $\mathbf{y}$ given $\mathbf{x}$, and the generation of $\mathbf{x}$ is not modeled by $\mathbf{w}$ ($\mathbf{x}$ is statistically independent of $\mathbf{w}$), i.e., $\mathrm{pdf}(\mathbf{x}|\mathbf{w}) = \mathrm{pdf}(\mathbf{x})$, the a posteriori distribution of the parameter vector becomes

$$\mathrm{pdf}(\mathbf{w}|\mathbf{x},\mathbf{y}) = \frac{\mathrm{pdf}(\mathbf{y}|\mathbf{x},\mathbf{w})\mathrm{pdf}(\mathbf{w})}{\mathrm{pdf}(\mathbf{y}|\mathbf{x})}. \tag{31}$$

The best likelihood $\mathrm{pdf}(\mathbf{y}|\mathbf{x},\mathbf{w})$ is given by the true conditional distribution $\widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x})$, cf. Fig. 4. If it is assumed that the data are independent, and, e.g., the true noise generation at position x is Gaussian $\mathcal{N}(\cdot)$ with mean given by the true deterministic function $\hat{y}(x)$ and having variance[10] σ^2, then the likelihood should be modeled as

$$\begin{aligned}\mathrm{pdf}(\mathbf{y}|\mathbf{x},\mathbf{w}) &= \prod_{n=1}^{N} \mathcal{N}(y_n|f(x_n,\mathbf{w}),\sigma^2) \\ &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_n - f(x_n,\mathbf{w}))^2}{\sigma^2}\right),\end{aligned} \tag{32}$$

and the optimal $f(x,\mathbf{w})$ is given by $\hat{y}(x)$. Moreover, if a priori there is reason to believe that the parameters in $\mathbf{w}$ are independently distributed as a Gaussian with mean zero and variance γ^{-1}, the a priori distribution would be

$$\begin{aligned}\mathrm{pdf}(\mathbf{w}) &= \prod_{m=1}^{M} \mathcal{N}(w_m|0,\gamma^{-1}) \\ &= \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\gamma^{-1}}} \exp\left(-\frac{\gamma}{2}w_m^2\right).\end{aligned} \tag{33}$$

Substituting Eq. (32),(33) in Eq. (31) and taking the negative logarithm[11] of Eq. (31) yields

$$-\ln \mathrm{pdf}(\mathbf{w}|\mathbf{x},\mathbf{y}) = \frac{1}{2}\sum_{n=1}^{N}\frac{(y_n - f(x_n,\mathbf{w}))^2}{\sigma^2} + \frac{\gamma}{2}\sum_{m=1}^{M} w_m^2 + const., \tag{34}$$

where *const.* includes remaining terms not involving $\mathbf{w}$. Thus, if the parameter vector that is looked for is one that has the highest a posteriori probability, i.e., $\mathbf{w}_{\mathrm{MAP}} = \arg\max_{\mathbf{w}} \mathrm{pdf}(\mathbf{w}|\mathbf{x},\mathbf{y})$, one can equivalently minimize Eq. (34), yielding

$$\mathbf{w}_{\mathrm{MAP}} = \arg\min_{\mathbf{w}} \left\{ \frac{1}{2}\sum_{n=1}^{N}\frac{(y_n - f(x_n,\mathbf{w}))^2}{\sigma^2} + \frac{\gamma}{2}\sum_{m=1}^{M} w_m^2 \right\}. \tag{35}$$

The subscribed index MAP means maximum a posteriori. Anyway, the MAP solution only reflects one specific point of the distribution, whereas there might be many further important features of the distribution such as width, multi-modality, skewness, etc., which can be calculated from expectation values of the form

$$\langle Q \rangle = \int Q(\mathbf{w})\mathrm{pdf}(\mathbf{w}|\mathbf{x},\mathbf{y})\ \mathrm{d}\mathbf{w}, \tag{36}$$

[10] Here, for exemplification, it is not considered that the variance σ^2 depends on x, i.e., homoscedastic noise is assumed.

[11] Note, the logarithm of the a posteriori distribution does not change the positions of the extrema, since it is a strictly monotonically increasing function.

where Q is a desired quantity. If $\mathbf{w}$ is high dimensional ($M \gtrsim 3$) the integration over $\mathbf{w}$ is usually analytically as well as by means of deterministic numerical integration methods not possible. Often, however, the expectation value can be determined via the Metropolis-MCMC method described in section 4.1.

In a frequentist approach, principles that lie outside the rules of probability theory have to be applied in order to perform inference. Here, e.g., the Kullback-Leibler Divergence (KLD) [89] from information theory is used to derive a frequentist inference method.[12] The KLD measures the amount of excess information that is required if the true distribution $\widehat{\text{pdf}}$ is modeled by pdf. The KLD is always larger than zero except if $\text{pdf} = \widehat{\text{pdf}}$:

$$\text{KLD}(\widehat{\text{pdf}}\|\text{pdf}) \geq 0, \qquad \text{and} \qquad \text{KLD}(\widehat{\text{pdf}}\|\text{pdf}) = 0 \Leftrightarrow \widehat{\text{pdf}} = \text{pdf}. \tag{37}$$

Hence, an optimal model can be determined from a class of models by selecting the model that has the lowest KLD.

Now, consider that data $(\mathbf{x}, \mathbf{y})$ are generated according to the true joint distribution $\widehat{\text{pdf}}(\mathbf{x}, \mathbf{y})$, and that $\text{pdf}(\mathbf{x}, \mathbf{y}|\mathbf{w})$ is a model of it — cf. case (i) in Fig. 4. In this case the KLD reads as follows

$$\text{KLD}(\widehat{\text{pdf}}\|\text{pdf}) = -\iint \widehat{\text{pdf}}(\mathbf{x}, \mathbf{y}) \ln\left(\frac{\text{pdf}(\mathbf{x}, \mathbf{y}|\mathbf{w})}{\widehat{\text{pdf}}(\mathbf{x}, \mathbf{y})}\right) \text{d}\mathbf{x}\ \text{d}\mathbf{y}, \tag{38}$$

and minimizing this term with respect to $\mathbf{w}$ would allow us to infer an optimal model for the overall generation of the data. However, here we are interested in the generation of $\mathbf{y}$ given $\mathbf{x}$ (cases (ii) and (iii) in Fig. 4)), therefore we use $\widehat{\text{pdf}}(\mathbf{x}, \mathbf{y}) = \widehat{\text{pdf}}(\mathbf{y}|\mathbf{x})\widehat{\text{pdf}}(\mathbf{x})$ and model $\text{pdf}(\mathbf{x}, \mathbf{y}|\mathbf{w})$ as $\text{pdf}(\mathbf{y}|\mathbf{x}, \mathbf{w})\widehat{\text{pdf}}(\mathbf{x})$, such that the KLD becomes

$$\begin{aligned}\text{KLD}(\widehat{\text{pdf}}\|\text{pdf}) &= -\iint \widehat{\text{pdf}}(\mathbf{x}, \mathbf{y}) \ln\left(\frac{\text{pdf}(\mathbf{y}|\mathbf{x}, \mathbf{w})}{\widehat{\text{pdf}}(\mathbf{y}|\mathbf{x})}\right) \text{d}\mathbf{x}\ \text{d}\mathbf{y} \\ &= -\iint \widehat{\text{pdf}}(\mathbf{x}, \mathbf{y}) \ln \text{pdf}(\mathbf{y}|\mathbf{x}, \mathbf{w})\ \text{d}\mathbf{x}\ \text{d}\mathbf{y} + \iint \widehat{\text{pdf}}(\mathbf{x}, \mathbf{y}) \ln \widehat{\text{pdf}}(\mathbf{y}|\mathbf{x})\ \text{d}\mathbf{x}\ \text{d}\mathbf{y},\end{aligned} \tag{39}$$

showing that only the first term on the r.h.s. in the last line of this equation, called generalization error, makes a cost contribution to the KLD when the KLD is minimized with respect to pdf. As aforementioned, pdf is described by a noise generating function and a regression model. If it is assumed that the noise generating function is fixed, then only the parameter vector $\mathbf{w}$ of the regression model has to be determined. Hence the generalization error has to be minimized with respect to $\mathbf{w}$

$$\arg\min_{\mathbf{w}} \left\{ \mathcal{E}_{\widehat{\text{pdf}}}(\mathbf{w}) := -\iint \widehat{\text{pdf}}(\mathbf{x}, \mathbf{y}) \ln \text{pdf}(\mathbf{y}|\mathbf{x}, \mathbf{w})\ \text{d}\mathbf{x}\ \text{d}\mathbf{y} \right\}, \tag{40}$$

in order to determine the optimal parameter vector. The integration expresses the general idea of the frequentist perspective: An infinite amount of data sets is required in order to perform inference. In practice, however, there is only a finite amount of data, therefore the

[12] Usually in other works the basis for determining some parameters starts directly with the maximum likelihood principle (see below). However, this principle "... is entirely based on intuition. It has no formal mathematical basis in and of itself."[129, §15.1]. Whereas the KLD from information theory seems to be a more fundamental principle, it does not only allow to derive optimal model parameters, but also allows to derive model selection methods.

generalization error cannot be minimized directly. As a substitute for it one may consider the empirical error

$$\begin{aligned} E_{\widehat{\text{pdf}}}(\mathbf{w}) &= -\frac{1}{N_{\mathcal{D}}} \sum_{n=1}^{N_{\mathcal{D}}} \ln \text{pdf}(\mathbf{y}_n|\mathbf{x}_n, \mathbf{w}) \\ &= -\frac{1}{N_{\mathcal{D}}} \ln \prod_{n=1}^{N_{\mathcal{D}}} \text{pdf}(\mathbf{y}_n|\mathbf{x}_n, \mathbf{w}), \end{aligned} \tag{41}$$

where $N_{\mathcal{D}}$ is the number of data sets. For independent data sets, it follows that with probability one it holds [87]

$$\lim_{N_{\mathcal{D}} \to \infty} E_{\widehat{\text{pdf}}}(\mathbf{w}) = \mathcal{E}_{\widehat{\text{pdf}}}(\mathbf{w}), \tag{42}$$

because of the strong law of large numbers. Minimization of the empirical error yields

$$\mathbf{w}_{\text{ML}} = \arg\min_{\mathbf{w}} E_{\widehat{\text{pdf}}}(\mathbf{w}), \tag{43}$$

which is called maximum likelihood (ML) solution. This is because the negative sum of logarithms of the likelihood functions is minimized in Eq. (43), which can be replaced by an equivalent optimization, the maximization of the product of likelihood functions, i.e., $\mathbf{w}_{\text{ML}} = \arg\max_{\mathbf{w}} \prod_{n=1}^{N_{\mathcal{D}}} \text{pdf}(\mathbf{y}_n|\mathbf{x}_n, \mathbf{w})$. In practice, there is often only one data set, $N_{\mathcal{D}} = 1$, containing N data points. Assuming this, and that the N data points in the available data set are independently generated, then the empirical error reads

$$\begin{aligned} E_{\widehat{\text{pdf}}}(\mathbf{w}) &= -\ln \text{pdf}(\mathbf{y}|\mathbf{x}, \mathbf{w}) \\ &= -\ln \prod_{n=1}^{N} \text{pdf}(y_n|x_n, \mathbf{w}) \\ &= -\sum_{n=1}^{N} \ln \text{pdf}(y_n|x_n, \mathbf{w}). \end{aligned} \tag{44}$$

Using in this formula, e.g., the Gaussian noise model from above it follows the optimization problem

$$\mathbf{w}_{\text{ML}} = \arg\min_{\mathbf{w}} \left\{ \frac{1}{2} \sum_{n=1}^{N} \frac{(y_n - f(x_n, \mathbf{w}))^2}{\sigma^2} \right\}. \tag{45}$$

Since, up to this point, no background information about the function $f(x, \mathbf{w})$ is taken into account in this approach, trouble can easily arise with the empirical (there are only a finite number of data points) frequentist inference, especially if a free-form solution is searched. In this case the function complexity should be high enough, such that the underlying $\hat{y}(x)$ can be described reasonably well; otherwise underfitting would occur, meaning that there would be a systematic error incorporated in the fitted function, see Fig. 5. Accordingly, there should be enough adjustable parameters controlling the complexity of the function. On the other side, if there are too many parameters, the function may follow the noise in the data, since this then decreases the value of the empirical error. This effect is known as overfitting. E.g., if there are N data points and if $f(x, \mathbf{w})$ is a polynomial with $M = N$ parameters, all points are interpolated by $f(x, \mathbf{w})$, and the empirical error is zero — cf. Fig. 5. In this case, the fitted function will in general not be a good approximation of $\hat{y}$, and the polynomial oscillates strongly. This problem can be easily revealed with a second i.i.d. data set (validation set)

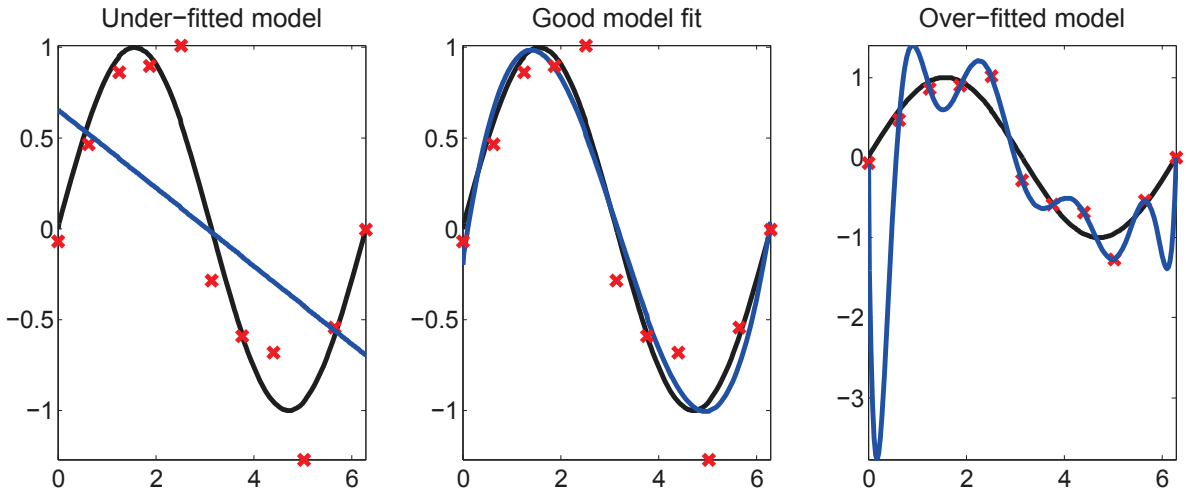


Figure 5: Regression of a data set consisting of $N = 11$ samples drawn from a noisy sine. Blue curves: regression models, black curves: noiseless sines, red crosses: noisy data. From left to right: fit via a line ($\mathbf{w}_{\mathrm{ML}}^T = (0.66, -0.22)$), fit via a polynomial of order three ($\mathbf{w}_{\mathrm{ML}}^T = (-0.2,\ 1.9, -0.85,\ 0.09)$) and fit via a polynomial of order ten ($\mathbf{w}_{\mathrm{ML}}^T = (-0.07, -48.8,\ 211.5, -356.9,\ 319.8, -170.5,\ 56.7, -11.8,\ 1.5, -0.11,\ 0.003)$).

on which the empirical error is determined, yielding a more objective estimate of the desired generalization error: the empirical error would give a value larger than zero for the second data set. In summary: the ML estimator is an estimator that minimizes the empirical error, but it has the tendency to perform overfitting (hence underestimating the generalization error) if the fit function is relatively complex and the amount of data is too small, cf. Fig. 6. However, one may apply the principle of 'Occam's razor' to prevent overfitting. It states, that the simplest model should be chosen that can explain the data. This idea is contained in the method of regularization that extends the empirical error for a penalty term $E_\gamma(\mathbf{w}) \geq 0$, which penalizes too complex models. The optimal parameter vector is then determined via

$$\mathbf{w}_R = \arg\min_{\mathbf{w}} \left\{ E_{\widehat{\mathrm{pdf}}}(\mathbf{w}) + E_\gamma(\mathbf{w}) \right\}. \tag{46}$$

In the polynomial regression example from above, one may notice that some parameter values may drastically increase with the number of fitted parameters, cf. Fig. 5, since a fitted parameter having a large positive value can be compensated by another parameter having a large negative value, and vice versa. Thus, ad hoc, the following penalizing function

$$E_\gamma(\mathbf{w}) = \frac{\gamma}{2} \sum_{m=1}^{M} w_m^2 \qquad \gamma > 0, \tag{47}$$

can be considered to prevent large absolute parameter values, if γ is sufficiently large, and hence gives a simple method to control the model complexity of the solution in Eq. (46). γ is called regularization parameter. Combining this penalizing term together with the squared error function (derived from the Gaussian noise model) gives the optimization problem

$$\mathbf{w}_{R,\gamma} = \arg\min_{\mathbf{w}} \left\{ \frac{1}{2} \sum_{n=1}^{N} \frac{(y_n - f(x_n, \mathbf{w}))^2}{\sigma^2} + \frac{\gamma}{2} \sum_{m=1}^{M} w_m^2 \right\}. \tag{48}$$

The problem of adjusting γ (hence fine tuning the model complexity) appropriately remains. As explained above, the minimization of the generalization error is the main goal, and it can

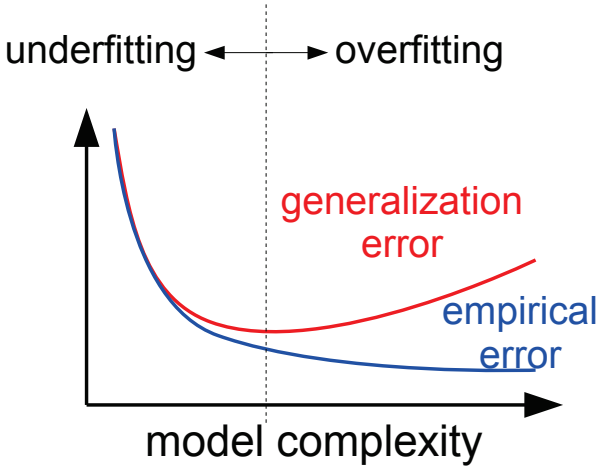


Figure 6: Under- and overfitting. Underfitting: If the model complexity is too low, the model is poorly approximated and the generalization error as well as the empirical error are high. Overfitting: If the model complexity is too high, the empirical error is low (since the model follows the noise in the data), but the generalization error is high. The optimal model complexity lies between the regime of underfitting and overfitting.

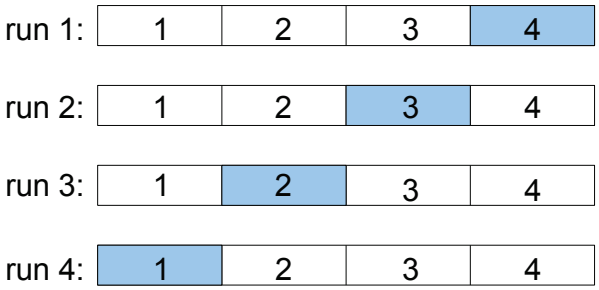


Figure 7: 4-folded cross-validation scheme. A data set is randomly divided into four disjoint data sets. In the first run model parameters are calculated on basis of the first three data sets and the error is determined on the fourth set. In the next three iterations the process is repeated with different disjoint sets for the calculation of the model parameters and the determination of the error. Finally, the errors are averaged.

be estimated via calculating the empirical error on a validation set. Consequently, a γ can be chosen that has the lowest empirical error on a second independent data set.
In case where there is only a single data set available, the generalization error can be estimated by means of the cross-validation method [111, §1.4.8]. This method works as follows — cf. Fig. 7 and algorithm 1: The original data set $\mathcal{D}$ is divided into K disjoint data sets, $\mathcal{D} = \dot{\bigcup}_{k=1}^{K} \mathcal{D}_k$. Then there are K runs. In the l-th run, the minimizer $\mathbf{w}_{R,\gamma}^{[-\mathcal{D}_l]}$ is determined on the data set $\dot{\bigcup}_{k \neq l} \mathcal{D}_k$, and the empirical error $E_{\widehat{\text{pdf}}}(\mathbf{w}_{R,\gamma}^{[-\mathcal{D}_l]})$ is calculated on the l-th data set. Finally, the average of the K empirical errors gives an estimate of the generalization error for a certain model (e.g., for a certain regularization parameter). Note, if the set $\mathcal{D}$ cannot be divided into equally sized disjoint data sets, the average calculation has to take into account the number of elements in each set.
In case the noise is a Gaussian one, having variance σ^2, and the data are independently distributed, another simple method to determine an optimal γ is the following one: At first,

Algorithm 1 The principle cross-validation algorithm (applied to the problem of selecting the optimal regularization parameter), cf. also Fig. 7.

1: Create a set of different models. Here, for exemplification, only a set of different regularization parameters $\{\gamma_j\}_{j=1}^{J}$ is considered that specify the model complexity.
2: Divide the original data set $\mathcal{D}$ into K disjoint sets $\{\mathcal{D}_k\}_{k=1}^{K}$, i.e., $\mathcal{D} = \dot{\bigcup}_{k=1}^{K} \mathcal{D}_k$.
3: **for** $j = 1$ **to** J **do**
4: Select model j, $\gamma = \gamma_j$.
5: **for** $l = 1$ **to** K **do**
6: $\mathbf{w}_{R,\gamma_j}^{l} = \arg\min_{\mathbf{w}} \left\{ E_{\widehat{\mathrm{pdf}}}(\mathbf{w}|\dot{\bigcup}_{k\neq l}\mathcal{D}_k) + E_{\gamma_j}(\mathbf{w}) \right\}$
7: **end for**
8: $\overline{E}_{\widehat{\mathrm{pdf}}}(\gamma_j) = \frac{1}{K}\sum_{l=1}^{K} E_{\widehat{\mathrm{pdf}}}(\mathbf{w}_{R,\gamma_j}^{l}|\mathcal{D}_l)$
9: **end for**
10: Select the optimal model complexity, $\gamma_j^* = \arg\min_{\gamma_j} \overline{E}_{\widehat{\mathrm{pdf}}}(\gamma_j)$.

the mean of the conditional distribution $\widehat{\mathrm{pdf}}(y|x)$ is defined as

$$\hat{y}(x) := \int y\, \widehat{\mathrm{pdf}}(y|x)\ \mathrm{d}y, \tag{49}$$

and then it is noticed that the generalization error can be rewritten as — neglecting terms independent of $\mathbf{w}$, which are irrelevant for a minimization of the generalization error with respect to $\mathbf{w}$ —

$$\begin{aligned}
\mathcal{E}'_{\widehat{\mathrm{pdf}}}(\mathbf{w}) &= N \iint \widehat{\mathrm{pdf}}(x,y) \left(\frac{1}{2} \frac{(y - f(x,\mathbf{w}))^2}{\sigma^2} \right) \mathrm{d}x\ \mathrm{d}y \\
&= \frac{N}{2} \iint \widehat{\mathrm{pdf}}(x,y) \frac{[y - \hat{y}(x) + \hat{y}(x) - f(x,\mathbf{w})]^2}{\sigma^2}\ \mathrm{d}x\ \mathrm{d}y \\
&= \frac{N}{2} \iint \widehat{\mathrm{pdf}}(x,y) \frac{(y - \hat{y}(x))^2}{\sigma^2}\ \mathrm{d}x\ \mathrm{d}y + \frac{N}{2} \iint \widehat{\mathrm{pdf}}(x,y) \frac{(\hat{y}(x) - f(x,\mathbf{w}))^2}{\sigma^2}\ \mathrm{d}x\ \mathrm{d}y \\
&\quad + N \iint \widehat{\mathrm{pdf}}(x,y) \frac{(y - \hat{y}(x))(\hat{y}(x) - f(x,\mathbf{w}))}{\sigma^2}\ \mathrm{d}x\ \mathrm{d}y \\
&= \frac{N}{2} + \frac{N}{2} \int \widehat{\mathrm{pdf}}(x) \frac{(\hat{y}(x) - f(x,\mathbf{w}))^2}{\sigma^2}\ \mathrm{d}x + 0,
\end{aligned} \tag{50}$$

thus showing that if $f(x,\mathbf{w}) = \hat{y}(x)$ the minimum of $\mathcal{E}'_{\widehat{\mathrm{pdf}}}(\mathbf{w})$ is given by $N/2$. In the following $\mathcal{E}'_{\widehat{\mathrm{pdf}}}(\mathbf{w})$ is also called generalization error. Accordingly, the optimal model should have the appropriate empirical error $E'_{\widehat{\mathrm{pdf}}}(\mathbf{w}_{R,\gamma^*}) \approx N/2$.

Note, in the frequentist discussion above the true joint distribution $\widehat{\mathrm{pdf}}(\mathbf{x},\mathbf{y})$ is used, since it was assumed that $(\mathbf{x},\mathbf{y})$ is a random sample from $\widehat{\mathrm{pdf}}(\mathbf{x},\mathbf{y})$ (general case). However, for SAS data, the random character of $\mathbf{x}$ is encoded into $\mathbf{y}$ after data pre-processing, and data are measured at discrete abscissa values. Accordingly, the generalization error in Eq. (40) is approximated by a generalization error whose expectation is calculated with respect to the conditional distribution $\widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x})$, i.e.,

$$\mathcal{E}_{\widehat{\mathrm{pdf}}}(\mathbf{w}) \approx \mathcal{E}_{\widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x})}(\mathbf{w}) := -\int \widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x}) \ln \mathrm{pdf}(\mathbf{y}|\mathbf{x},\mathbf{w})\ \mathrm{d}\mathbf{y}, \tag{51}$$

and minimizing $\mathcal{E}_{\widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x})}(\mathbf{w})$ is equivalent to minimizing the Kullback-Leibler divergence $\mathrm{KLD}(\widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x})|\mathrm{pdf}(\mathbf{y}|\mathbf{x},\mathbf{w}))$. Using a finite, discrete data set, the empirical error can be further on calculated according to Eq. (41), but it then gives an estimate of $\mathcal{E}_{\widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x})}(\mathbf{w})$, and for independently distributed Gaussian data $(y_n|x_n \sim \mathcal{N}(y|\hat{y}(x_n),\sigma^2))$ the generalization error from which the constant terms with respect to $\mathbf{w}$ are dropped, $\mathcal{E}'_{\widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x})}(\mathbf{w})$, is similar to the result given in Eq. (50):

$$\begin{aligned}\mathcal{E}'_{\widehat{\mathrm{pdf}}(\mathbf{y}|\mathbf{x})}(\mathbf{w}) &= \sum_{n=1}^{N} \int \widehat{\mathrm{pdf}}(y_n|x_n) \left(\frac{1}{2}\frac{(y_n - f(x_n,\mathbf{w}))^2}{\sigma^2}\right) \mathrm{d}y_n \\ &= \frac{N}{2} + \frac{1}{2}\sum_{n=1}^{N}\left(\frac{(\hat{y}(x_n) - f(x_n,\mathbf{w}))^2}{\sigma^2}\right),\end{aligned} \tag{52}$$

which is equal to $N/2$ if $\hat{y}(x_n) = f(x_n,\mathbf{w})\ \forall n$.

Last but not least, it is mentioned that there are further theoretical concepts (e.g., Akaike information criterion, Bayesian information criterion, structural risk minimization) that allow to estimate the generalization performance in dependence of the actual empirical error and a term that incorporates the complexity of the used function (class), see, e.g., [11].

Noteworthy, the optimization problem for determining an optimal parameter vector $\mathbf{w}_{R,\gamma}$ in Eq. (48) in the regularized frequentist approach is identical to the optimization problem given in Eq. (35) for determining the MAP solution $\mathbf{w}_{\mathrm{MAP}}$ in the Bayesian approach. However, the paths to arrive at these optimization problems are conceptually totally different: In the Bayesian approach all a priori information is used to build the a priori distribution of the parameters as well as the likelihood, and subsequently the rules from probability theory are used to perform a structured inference in order to get the a posteriori distribution. In the frequentist approach different concepts and ad-hoc principles have been applied to infer a satisfying solution in the case of a finite number of data points.[13]. Note, if the number of data points increases in a Bayesian approach, the likelihood term will dominate the prior term, hence the prior will become irrelevant. Thus, taking the negative logarithm of the a posteriori distribution and assuming an infinite amount of data, the Bayesian approach yields essentially the generalization error. Moreover, in the frequentist inference procedure the optimal parameter vector is always determined by means of solving an optimization problem, whereas in the Bayesian inference the a posteriori distribution is given by using basic arithmetic operations (additions and multiplications are required for applying the sum and product rules of probability theory) and only needs an optimization if, e.g., the MAP solution shall be determined.

So far not considered is the problem of subjectivity in the Bayesian approach: If, e.g., there are two persons having the same background information about the a priori distribution of the parameters, they may incorporate the information differently, yielding different a priori distributions. E.g., if both persons would know the mean vector $\boldsymbol{\mu}_{\mathbf{w}}$ and the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{w}}$ of a parameter distribution, one person may find an unimodal distribution in accordance

[13] Cox's theorem [31] states that there is only one unique way of performing consistent statistical inference (Jaynes [77] remarks that "this article was the most important advance in the conceptual ... formulation of probability theory since Laplace"). The rules for doing this are encoded in the sum and product rules of probability theory (that then yield the Bayes formula). Any other method, as done in the 'ad-hoc devices' [77] in frequentist approaches, thus lead inevitably to inconsistencies. Jaynes devoted a whole chapter in his book [77] to inconsistencies of frequentist statistics, showing paradoxes of unbiased estimators, confidence intervals, statistical tests, etc. (see also [111, §6.6]). In [129, §15.0.1] "Either you become a Bayesian or else you must live in a world with no general calculus of inference.".

with $\boldsymbol{\mu}_\mathbf{w}$ and $\boldsymbol{\Sigma}_\mathbf{w}$, while the other person may take a different distribution, e.g., a bimodal one, which satisfies $\boldsymbol{\mu}_\mathbf{w}$ and $\boldsymbol{\Sigma}_\mathbf{w}$, too. In order to overcome such ambiguities, the maximum entropy method (MEM) [76] can be applied to get a unique distribution that is in accordance with the constraints. People following this approach belong to the objective Bayesian school, while the others belong to the subjective Bayesian school. The MEM maximizes the entropy [139] of the probability distribution with respect to the given constraints. E.g., if the mean vector $\boldsymbol{\mu}_\mathbf{w}$ and the covariance matrix $\boldsymbol{\Sigma}_\mathbf{w}$ of the distribution are known, the distribution that is found by the MEM is the Gaussian one $\mathcal{N}(\mathbf{w}|\boldsymbol{\mu}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{w})$ [54, appendix E].[14,15]

Another aspect that has not been discussed so far is the issue of model comparison / model selection. E.g., if there are two different physical models or two different classes of free-form functions (e.g., one class may consist of a number of adjustable sine functions and the other class may consist of a number of adjustable b-splines) and the 'optimal' model shall be selected.

As discussed above, in frequentist inference the fundamental idea is to minimize the KLD, or equivalently the generalization error, and therefore an optimal model is one that has the lowest generalization error. Thus, the model selection procedure in a frequentist approach only requires an estimation procedure of the generalization error in order to perform model comparison. Often cross-validation is a good choice.[16] However, if there are two different models yielding the same value of the KLD, additional knowledge is required in order to perform model selection.

In the Bayesian inference, the principle idea is to assign a posteriori probabilities, in accordance with the rules of probability theory, to the models and their parameters given that some data are observed. It essentially adjusts the a priori probabilities (of the models and their parameters) in accordance with the observed data to yield the joint a posteriori distribution of the models and their parameters. In the following it is assumed that there is a set of K model classes $\{H_k\}_{k=1}^K$, and each model can have L_k hyperparameters[17], summarized in the vector $\boldsymbol{\theta}_k$, and has a parameter vector $\mathbf{w}_k$ with M_k elements. Then, the a posteriori distribution is $\mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_\mathbf{k}, H_k|\mathcal{D})$, where $\mathcal{D}$ is the observed data set. The product rule of probability theory can be sequentially applied in order to rewrite $\mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_\mathbf{k}, H_k|\mathcal{D})$: Firstly,

$$\mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_k, H_k|\mathcal{D}) = \mathrm{pdf}(\mathbf{w}_k|\boldsymbol{\theta}_k, H_k, \mathcal{D})\,\mathrm{pdf}(\boldsymbol{\theta}_k, H_k|\mathcal{D}), \tag{53}$$

where

$$\mathrm{pdf}(\boldsymbol{\theta}_k, H_k|\mathcal{D}) = \mathrm{pdf}(\boldsymbol{\theta}_k|H_k, \mathcal{D})\,\mathrm{pdf}(H_k|\mathcal{D}), \tag{54}$$

and hence

$$\mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_k, H_k|\mathcal{D}) = \mathrm{pdf}(\mathbf{w}_k|\boldsymbol{\theta}_k, H_k, \mathcal{D})\,\mathrm{pdf}(\boldsymbol{\theta}_k|H_k, \mathcal{D})\,\mathrm{pdf}(H_k|\mathcal{D}). \tag{55}$$

Each term on the r.h.s. of Eq. (55) can be interpreted as an a posteriori distribution in a hierarchical inference procedure [98],[131, §5.2]. On the lowest level the probability distribution of $\mathbf{w}_k$ is inferred given $\boldsymbol{\theta}_k$, H_k and $\mathcal{D}$

$$\mathrm{pdf}(\mathbf{w}_k|\boldsymbol{\theta}_k, H_k, \mathcal{D}) = \frac{\mathrm{pdf}(\mathcal{D}|\mathbf{w}_k, \boldsymbol{\theta}_k, H_k)\,\mathrm{pdf}(\mathbf{w}_k|\boldsymbol{\theta}_k, H_k)}{\mathrm{pdf}(\mathcal{D}|\boldsymbol{\theta}_k, H_k)}, \tag{56}$$

[14] In section 6.1.5, the MEM is used to determine an orientational distribution function.

[15] A distribution that maximizes the entropy subject to a set of constraints can be seen as the most general (i.e., maximally non-committal) distribution that is consistent with the constraints.

[16] There are also other model comparison methods like the well known Akaike Information Criterion (AIC) that is asymptotically equivalent to the cross-validation method [145]. However, in practice it can also easily give misleading results [16, §4.4.1] if the model parameters are not "well-determined".

[17] E.g., the width of a Gaussian a priori distribution can be an uncertain hyperparameter.

with the marginal likelihood

$$\mathrm{pdf}(\mathcal{D}|\boldsymbol{\theta}_k, H_k) = \int \mathrm{pdf}(\mathcal{D}|\mathbf{w}_k, \boldsymbol{\theta}_k, H_k)\,\mathrm{pdf}(\mathbf{w}_k|\boldsymbol{\theta}_k, H_k)\;\mathrm{d}\mathbf{w}_k. \tag{57}$$

On the second level of inference, the a posteriori distribution of the hyperparameters $\boldsymbol{\theta}_k$ given H_k and $\mathcal{D}$ is determined by calculating

$$\mathrm{pdf}(\boldsymbol{\theta}_k|H_k, \mathcal{D}) = \frac{\mathrm{pdf}(\mathcal{D}|\boldsymbol{\theta}_k, H_k)\,\mathrm{pdf}(\boldsymbol{\theta}_k|H_k)}{\mathrm{pdf}(\mathcal{D}|H_k)} = \int \mathrm{pdf}(\boldsymbol{\theta}_k, \mathbf{w}_k|H_k, \mathcal{D})\;\mathrm{d}\mathbf{w}_k, \tag{58}$$

where the marginal likelihood is

$$\mathrm{pdf}(\mathcal{D}|H_k) = \int \mathrm{pdf}(\mathcal{D}|\boldsymbol{\theta}_k, H_k)\,\mathrm{pdf}(\boldsymbol{\theta}_k|H_k)\;\mathrm{d}\boldsymbol{\theta}_k. \tag{59}$$

At the top level, the model a posteriori probabilities are inferred:

$$\mathrm{pdf}(H_k|\mathcal{D}) = \frac{\mathrm{pdf}(\mathcal{D}|H_k)\,\mathrm{pdf}(H_k)}{\mathrm{pdf}(\mathcal{D})} = \iint \mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_k, H_k|\mathcal{D})\;\mathrm{d}\mathbf{w}_k\;\mathrm{d}\boldsymbol{\theta}_k, \tag{60}$$

where the marginal likelihood is

$$\mathrm{pdf}(\mathcal{D}) = \sum_{k=1}^{K} \mathrm{pdf}(\mathcal{D}|H_k)\,\mathrm{pdf}(H_k). \tag{61}$$

As it is shown here, the Bayesian inference procedure can be a well structured one, and can be easily extended to even more nested model classes. In accordance with this hierarchical inference procedure, Eq. (55)-(61), hierarchical model selection can be done by first selecting the most probable model class given the data $\mathcal{D}$

$$k^* = \arg\max_k \{\mathrm{pdf}(H_k|\mathcal{D})\}, \tag{62}$$

i.e., the MAP solution (of the joint a posteriori distribution in which $\mathbf{w}_k$ and $\boldsymbol{\theta}_k$ are marginalized out) for the model class. Next, the MAP solution for the hyperparameters given H_{k^*} and $\mathcal{D}$ reads

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \{\mathrm{pdf}(\boldsymbol{\theta}|H_{k^*}, \mathcal{D})\}, \tag{63}$$

and finally the MAP solution for the parameters given $\boldsymbol{\theta}^*, H_{k^*}$ and $\mathcal{D}$ is

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \{\mathrm{pdf}(\mathbf{w}|\boldsymbol{\theta}^*, H_{k^*}, \mathcal{D})\}. \tag{64}$$

Anyway, performing the whole inference procedure given in Eq. (55)-(61) is usually analytically as well as by means of deterministic numerical integrations not possible because of the nested and often high-dimensional integrals.[18] In this case Monte-Carlo methods may be used to sample from the distributions. E.g., by means of the Metropolis-MCMC method described in chapter 4.1 it is possible to draw samples from a distribution that is proportional to a desired distribution, and determining the appropriate normalization constant (marginal likelihoods) may be achieved by applying an extended Metropolis-MCMC method [163] or by using nested sampling methods [142, 42]. Complementary to these statistical sampling

[18] This is still a reason why simple concepts for performing model selection (e.g., cross-validation) are often taken.

methods are deterministic approximate inference methods. E.g., there is the Laplace approximation of a distribution around its mode by a Gaussian (cf. section 3.2, page 33) or, e.g., the involved distributions are assumed to be well approximated by factorizing distributions (this is a special case of variational inference [16, §10.1]), such that the nested integrals can be written as a product of integrals.

Of course, model selection via $\mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_k, H_k|\mathcal{D})$ can mathematically also be performed by determining the global maximum of the joint a posteriori distribution

$$(\mathbf{w}_{k^*}^*, \boldsymbol{\theta}_{k^*}^*, H_{k^*}) = \arg \max_{(\mathbf{w}_k, \boldsymbol{\theta}_k, H_k)} \{\mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_k, H_k|\mathcal{D})\}, \tag{65}$$

neglecting any hierarchical inference interpretation. Note here that the Bayes formula only describes the mathematical procedure of how consistent inference is performed, but it does not make any statement of how the variables are to be interpreted or how to encode them in the formula. As explained above background information should be encoded objectively in the formula, and usually the hierarchical variable model is chosen, since then the hierarchical model selection given in Eq. (62)-(64) prevents from selecting an overfitted model. This is because in a hierarchical model selection procedure the principle of Occam's razor is intrinsically contained within the marginalization process (see, e.g., [16, §3.4]): Consider different model functions/classes. Then, if a model function is more complex than another one, it can create more different data sets $\mathcal{D}$ than the simpler model, i.e., $\mathrm{pdf}(\mathcal{D}|H_k) = \int\int \mathrm{pdf}(\mathcal{D}|\mathbf{w}_k, \boldsymbol{\theta}_k, H_k)\, \mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_k|H_k)\, \mathrm{d}\mathbf{w}_k\, \mathrm{d}\boldsymbol{\theta}_k$ is broader for the more complex model, and thus it is very likely that the probability $\mathrm{pdf}(\mathcal{D}|H_k)$ for a specific data set $\mathcal{D}'$ will be less for the more complex model (since the total area under the distribution $\mathrm{pdf}(\mathcal{D}|H_k)$ is normalized), see Fig. (8). Therefore, if the a priori distribution of models $\mathrm{pdf}(H_k)$ does not influence Eq. (60) much, the model that is selected by means of Eq. (62) is the one that has the highest $\mathrm{pdf}(\mathcal{D}'|H_k)$, and is likely to be the simplest one that is in agreement with the data; hence the principle of Occam's razor is implemented. On the second level of inference one can argue as on the top level. Hence, the hierarchical model selection procedure can be performed, in order to prevent from selecting an overfitted model. However, if overfitting does not occur, and if there is no need for a hierarchical model description, the maximization of the joint a posteriori distribution, Eq. (65), is computationally less demanding than performing the hierarchical model selection via Eq. (62)-(64), and can be applied as well — e.g., it is applied in section 3.2 in Eq. (72)-(74).

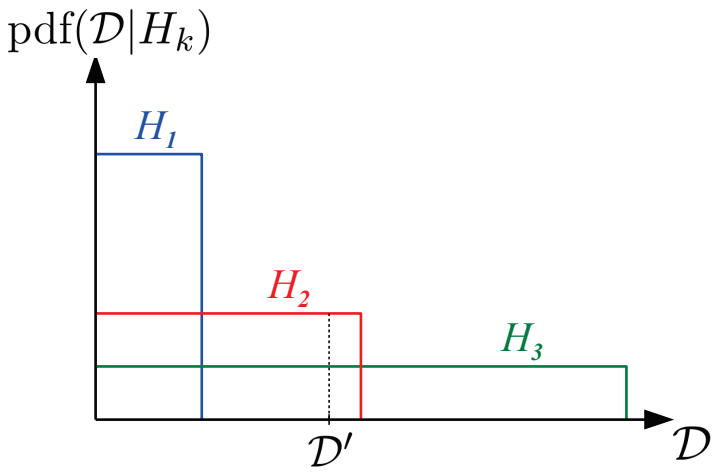


Figure 8: Hierarchical model selection incorporates the principle of Occam's razor — a schematic illustration. Three different model classes $\{H_k\}_{k=1}^3$ are considered, where H_1 is the simplest model and H_3 the most complex one. Model class H_1 is only able to generate a few different data sets $\mathcal{D}$, hence the distribution pdf($\mathcal{D}|H_1$) is confined on a small interval, and in this interval pdf($\mathcal{D}|H_1$) is relatively high (i.e., the probability that model H_1 generates a data set from this interval is relatively high). Model class H_3 is the most complex one, and thus can generate many different data sets. Therefore the distribution pdf($\mathcal{D}|H_3$) is very broad, and hence pdf($D|H_3$) is relatively low. The model complexity of H_2 lies between the complexities of H_1 and H_3, and H_2 has for the data set $\mathcal{D}'$ the highest likeliness. Hence, if the data set $\mathcal{D}'$ is observed, this intermediate model will be preferred to the more complex model H_3 (if there is no further impact by a weighting with the priors, cf. Eq. (60)) in the hierarchical model selection scheme. Therefore, the hierarchical model selection procedure incorporates the principle of Occam's razor.

3. SASET and Model Fitting

This section discusses the issue of fitting physical models to scattering data in order to infer structural information from a SAS experiment. As outlined in section 2.1 (approach 3 (page 4), and text thereafter), a robust way to infer structural information from a scattering experiment is to build a physical model of the scattering system (as well as to assume a noise model), and then to use an optimization routine to fit the model to the scattering data, i.e., to perform regression (cf. section 2.2), in order to determine optimal model parameters. The development of the fitting program SASET is a centerpiece of this thesis. This program is unique in that it allows to perform fast and simultaneous model fitting of large SAS data sets/series with flexible models and optimization options; it is presented in the next subsection. The subsequent subsection, discusses the case of model fitting in the context of the Bayesian and frequentist statistical inference, which sheds light on the trustworthiness of fitted models and the uncertainties of model parameters. In the last subsection simultaneous model fitting is performed in order to exemplify the potency of SASET as well as to discuss the results from section 3.2 in the context of real data fits.

3.1. SASET

Many different SAS analysis programs are used today [88, 132, 127, 150, 61, 32, 179, 36], either proposed by facilities for their users or developed in groups specializing in SAS. However, a systematic trend of such software is that they are developed by scientists with little programming knowledge, initially for a very specific, narrow and typically not very challenging task, with new functions added progressively as 'patchs' on the original code. As a result, relatively powerful options, such as simultaneous or series fitting, are under-performing, and only work well for quite simple cases. For example, SASfit [88], originally developed by J. Kohlbrecher during his PhD thesis in the early 90s, can handle simultaneous data set fitting since many years and series fitting since a couple of years. However, the architecture of the software, originally conceived for single data set fitting, creates restrictions for both features. Simultaneous fitting of more than 10 spectra, even with a simple model and few fit parameters, becomes extremely time consuming, parameter ranges cannot be constrained, and only a simple relationship between global parameters can be handled. Series fitting requires that all spectra are very similar so that minor changes in parameters are a prerequisite, and any failure to fit will result in the interruption of the series fit. Yet this program is probably so far the most used, and the 'best', proposed to the SAS community. Even more recent projects, such as SASview, are lagging behind in terms of functionalities.

At the opposite, the new and homemade program SASET (SAS Evaluation Tool, SASET) has been purposely designed to handle large amounts of data efficiently, and it taps into the vast capabilities of MATLAB for minimization algorithms; thus allowing for performing model fitting with virtually unlimited constraints between parameters. SASET is published in [113], and the software is available from [112]. A preliminary version of SASET was implemented by the author in his diploma thesis, and it could mainly accomplish the following: it could perform 1-dimensional model fitting on a whole series of scattering curves via the Levenberg-Marquardt algorithm (without constraints), and it could extract 1-dimensional scattering curves from 2-dimensional images. In the this work, new features, which are listed below, have been added to SASET in order to allow for more complex, comprehensive and efficient SAS data analysis, and therefore state of the art experiments, where a lot of information has to be processed, can be easily evaluated. All features are implemented user-friendly. Moreover, many physical model functions have been implemented, and are fully documented

within the help of SASET — available from within the program. Tutorials with step-by-step instructions for all major features/topics are accessible from the main interface.

An essential aspect of the development of SASET is that it results from the interconnection between a software engineer and scientists from physical-chemistry, resulting in a sophisticated code tailor-made for the practical problems of end users. E.g., the option of simultaneous fitting (see below) has been developed in direct cooperation with Dr. Katharina Bressel, who gave a lot of useful ideas to make the program interaction very user-friendly. Dr. Bressel used SASET extensively for analyzing time-series of scattering data of self-aggregating colloidal systems (see also section 5), obtained from different experimental SAXS configurations, by means of performing simultaneous series fitting. The results of these analyses are published in [23, 24, 25].

The realization of users wishes resulted in a versatile, efficient and user-friendly program.

In the present work following main features have been added to SASET:

1. New optimization procedures are available for performing (series) model fitting. These procedures are

   (*i*) the `lsqcurvefit()` procedure of MATLAB, and

   (*ii*) the `fmincon()` procedure of MATLAB,

   which allow to perform model fitting with constraints. Constraints allow to incorporate information derived from other experiments or the molecular or mesoscopic build-up of the system, thereby enhancing the reliability of the inferred structural information, cf. section 2.2. The originally implemented Levenberg-Marquardt algorithm does not allow to use constraints in its fitting procedure. The two newly included algorithms allow to define lower and upper parameter bounds: Let $\mathbf{a}$ the parameter vector controlling the model function $f(q, \mathbf{a})$, where q is the magnitude of the scattering vector, and let $\mathcal{D} = \{(q_n, I_n, \sigma_n)\}_{n=1}^{N}$ the observed scattering data set, where I_n is the n-th intensity value measured at position q_n and having the uncertainty σ_n, then the following optimization problem can be solved by the new algorithms in SASET

$$\arg\min_{\mathbf{a}} \left\{ \sum_{n=1}^{N} \frac{\left[I_n - f(q_n, \mathbf{a})\right]^2}{2\sigma_n^2} \right\} \quad \text{subject to} \quad \mathbf{lb} \leq \mathbf{a} \leq \mathbf{ub}, \tag{66}$$

   where $\mathbf{lb}$ is a vector defining the lower bounds of the parameters in $\mathbf{a}$, and $\mathbf{ub}$ is a vector defining the upper bounds of the parameters. Additionally, the `fmincon()` procedure allows to perform model fitting with even more general constraints in Eq. (66), which are given by

$$\begin{aligned} \mathbf{c}(\mathbf{a}) &\leq \mathbf{0}, \\ \mathbf{ceq}(\mathbf{a}) &= \mathbf{0}, \\ \mathbf{A} \cdot \mathbf{a} &\leq \mathbf{b}, \quad \text{and} \\ \mathbf{Aeq} \cdot \mathbf{a} &= \mathbf{beq}, \end{aligned} \tag{67}$$

   where $\mathbf{c}(\mathbf{a})$ and $\mathbf{ceq}(\mathbf{a})$ are (nonlinear) vectorial functions of the parameter vector $\mathbf{a}$, $\mathbf{A}$ and $\mathbf{Aeq}$ are matrices, and $\mathbf{b}$ and $\mathbf{beq}$ are vectors. Nonlinear constraints become useful in cases such as fitting the scattering intensity of a multi-component complex fluid where only overall values are experimentally known, e.g., in bimodal mini-/micro-emulsions where the total concentrations of oil, surfactants and cosurfactants are fixed [69]. Therein nonlinear constraints allow to insure that the total volume fractions of

oil (inside mini-/micro-emulsion droplets), surfactants (at the interface of the droplets) and cosurfactant (possibly partitioned between the interface, the oil and the water, where interfacial cosurfactant amount depends on the curvature) are fixed. Inequality constraints are useful to insure that quantities of interest stay below or above certain threshold values.

2. Different data sets can be fitted simultaneously by SASET. Simultaneous fitting means that there are different data sets with corresponding models, and that the models are coupled by common/global parameters within the fitting procedure. Such a situation arises frequently in SANS experiments for the case of external or internal contrast variation, but also if there are SANS and SAXS data for the same sample, or when using Anomalous SAXS (ASAXS). By means of simultaneous fitting one then is able to infer more reliably structural parameters and can also employ refined structural models that use the enhanced information content of such experiments. Then, the fitting routine minimizes over a parameter space that is given by the global parameters as also by the non-global parameters: Given K data sets $\{D_k\}_{k=1}^{K}$, where $D_k = \{(q_{nk}, I_{nk}, \sigma_{nk})\}_{n=1}^{N_k}$, and corresponding models $\{f_k(\cdot,\cdot)\}_{k=1}^{K}$, the optimization problem reads as follows

$$\arg \min_{\mathbf{b},\{\mathbf{a}'_k\}_{k=1}^{K}} \left\{ \mathcal{E}_s := \sum_{k=1}^{K} w_k \sum_{n=1}^{N_k} \frac{\left[I_{nk} - f_k\big(q_{nk}, \mathbf{g}_k(\mathbf{a}_k, \mathbf{h}_k(\mathbf{b}, \mathbf{c}_k))\big)\right]^2}{2\sigma_{nk}^2} \right\} \tag{68}$$

with

$$\begin{aligned} \mathbf{g}_k : \quad & \mathbb{R}^{M_k} \times \mathbb{R}^{L_k} \to \mathbb{R}^{M_k} \\ & (\mathbf{a}_k, \mathbf{H}_k) \mapsto \tilde{\mathbf{a}}_k \end{aligned} \tag{69}$$

and

$$\begin{aligned} \mathbf{h}_k : \quad & \mathbb{R}^{b} \times \mathbb{R}^{d} \to \mathbb{R}^{L_k} \\ & (\mathbf{b}, \mathbf{c}_k) \mapsto \mathbf{H}_k. \end{aligned} \tag{70}$$

$\mathbf{a}_k$ is the primal parameter vector of the k-th model. The mapping $\mathbf{g}_k(\cdot,\cdot)$ returns $\tilde{\mathbf{a}}_k$ that is identical to $\mathbf{a}_k$ except that L_k entries are bijectively replaced by elements of $\mathbf{H}_k$. The parameters that are not replaced are collected in $\mathbf{a}'_k$. The elements of $\mathbf{H}_k = \mathbf{h}_k(\mathbf{b}, \mathbf{c}_k)$ are calculated from the global parameter vector $\mathbf{b}$ and some fixed parameters in $\mathbf{c}_k$ that are associated with the k-th data set. $\mathbf{c}_k$ can directly be defined from the header of a BerSANS [83] file, but can also be set manually. $\{w_k | w_k \in \mathbb{R}_+\}_{k=1}^{K}$ is a set of fixed weights, which allow to weight the k-th contribution in the cost function.

The user can interactively define the global parameters $\mathbf{b}$ as well as the functional behavior, i.e., the functions $\left\{\big(\mathbf{g}_k(\cdot,\cdot), \mathbf{h}_k(\cdot,\cdot)\big)\right\}_{k=1}^{K}$, in a table. Thus, a coupling of the model functions can be performed quickly and adapted to the problem at hand. Consequently, contrast variation data sets as also data coming from different instrumental setups (e.g., different sample-to-detector distances resulting in different instrumental resolution functions) can easily be analyzed.

The optimization procedure that is used for simultaneous fitting is `lsqcurvefit()`. It is planned that `fmincon()` will be available as well in a later SASET version.

3. Parallel methods have been implemented in order to speed up the fitting procedure: (*i*) integration procedures can be performed in parallel mode, (*ii*) if a simultaneous fit is performed, the model functions can be calculated in parallel mode.

These new parallel program features can be used, if there is an open pool of MATLAB workers. A MATLAB worker is an additional running MATLAB instance that runs in the background of the main instance. It should have its own resources, i.e., it should use its own core and enough memory should be available for it. Then, the program parallelization is achieved by sending independent statements (e.g., the body of a parallel for-loop) from the main MATLAB instance to the different workers. Of course there is a communication overhead when the main MATLAB instance communicates with the workers, due to sending the statements to the workers and sending the results back from the workers to the main MATLAB instance. Thus, the statements that shall be calculated in parallel have to be sufficiently time consuming to at least compensate the communication overhead.

Following benchmarks are done — on a PC with an Intel® Core™ i7 CPU 860 at 2.8 GHz (4 cores, 8 threads), running Matlab2012b, and eight local Matlab workers are used in parallel mode — with compute-intensive models to highlight time differences:

(*i*): A typical case is the scattering of a polydisperse, lognormally distributed collection of cylinders, which can be computed by employing the model `lognpdfInRCylinders()` in SASET. The evaluation of this function takes in non-parallel mode ca. 1.5 s for 200 different q-values, whereas the calculation takes ca. 0.4 s in parallel mode with eight workers (100 integration support points for the integration over the lognormal distribution are used, and 50 integration support points for the orientational average of a cylinder). Hence, the speedup[19] is ca. 370 %.

(*ii*): For the more compute-intensive model `homogeneousSpheresCLSF()`, which calculates the scattering of spheres interacting with a crystalline structure factor, the execution speed in non-parallel mode is ca. 6.1 s, and in parallel mode ca. 1.3 s (for 200 different q-values, and 250 integration support points for each of the two inherent integrations). Hence, the speedup is ca. 470 %.

(*iii*): Fitting a polydisperse core-shell-(linear corona) model with polydisperse hard sphere structure factor (see section 3.3.1) to the data set 3 (containing four measured intensity curves) from section 3.3 takes ca. 600 s in non-parallel mode and ca. 185 s if the functions are calculated in parallel. Hence, the speedup is ca. 320 %. (If the content of the functions is calculated in parallel mode the fit takes ca. 245 s.)

These benchmarks show typical speedups. Moreover, the benchmarks show that even if there are eight local workers available, the speedup typically does not increase much more than 400 % if there are only four cores available. In principle, one would expect that, e.g., with eight cores and eight threads the performed benchmarks (*i*) and (*ii*) would have a speedup of ca. 700 – 800 % in parallel mode, whereas for case (*iii*) one would not expect an increase of the speedup in parallel mode when the functions are calculated in parallel, since there are only four different functions to be calculated. More detailed analyses of the benchmarks are difficult, since, e.g., CPU frequency (in particular 'turbo-boost' mode with modulation of the frequency depending on the type and number of tasks), cache memories of various levels and also hyper-threading effects have to be taken into account (hyper-threading being known for having sometimes counter-productive effects).

Note that MathWorks restricts the number of MATLAB workers in the Parallel Computing Toolbox, but MathWorks regularly increases it when new processors with more

[19] The speedup is defined as $T_{\text{old}}/T_{\text{new}}$ [172], where T_{old} is the old execution time, i.e., without the improvement, and T_{new} is the new execution time, i.e., with the improvement.

cores are released. Thus, over the next few years SASET will be able to exploit new CPUs with more cores and thus it can become even faster for complex calculations.

4. A model function file parser has been implemented. This parser does the following:

   a) Scattering model function files being in certain directories of the SASET parent folder are automatically available from a pop-up menu in the fitting Graphical User Interface (GUI). Consequently, this allows users to write their own model functions, save them in a separate directory, and access them from within the GUI.

   b) If a model function is selected from the pop-up menu, the function file is analyzed for consistency and a function definition is created (which includes the definition of fit start values, lower and upper parameter bounds as well as an information whether a parameter shall be fitted or not) that is then set into some edit-fields of the fitting GUI. Because the function definition is editable within the edit-fields, this easily allows to manipulate it for minor changes. Moreover, there are further tools, which allow, e.g., (*i*) to edit conveniently the fit start parameters in a table, as well as their bounds and the information whether a parameter shall be fixed or not, and (*ii*) to easily add a smearing function to a current function definition.

   c) A model function file can be endowed with a result function and a plot function (i.e., a certain string information can be added as a comment that is then interpreted by SASET as a special function call). This allows to evaluate further functions given the parameters of the current model, e.g., in order to calculate meaningful structural parameters (e.g., the gyration radius) or the forward intensity of the current model, or to create a plot in dependence of the current model (size distribution, scattering length density profile, effective structure factor, 2-dimensional cross section or 3-dimensional representation of the shape of the scatterer, etc.).

   The principle form of a model function file is described in Fig. 9. As explained there, the code may also contain further comments and specific commands, which then allow to generate an associated function description/help (containing, e.g., hyperlinks, images, and LaTeX formulas) out of it, cf. Fig. 10.

   The implemented parser uses a lot of regular expressions in order to look for patterns in the function file.

5. A file parser, similar as the one described in the last point, has been implemented in order to access and use anisotropy function files within SASET (cf. section 6 for the anisotropy measurement methods that are by default available in SASET).

6. The fit start parameters of a whole 1-dimensional scattering data series can be defined in a table.

7. Different radial or azimuthal intensity curves can be extracted from a 2-dimensional image and attached with a label. Hereafter, the labeled set of intensity curves can be fitted simultaneously. This extraction and labeling of intensity curves works also on series of scattering images.

Before in section 3.3 a simultaneous fitting evaluation is performed on an interpolyeletrolyte complex system, which has been done with SASET, a general discussion about fitting of physical models is given in the next section.

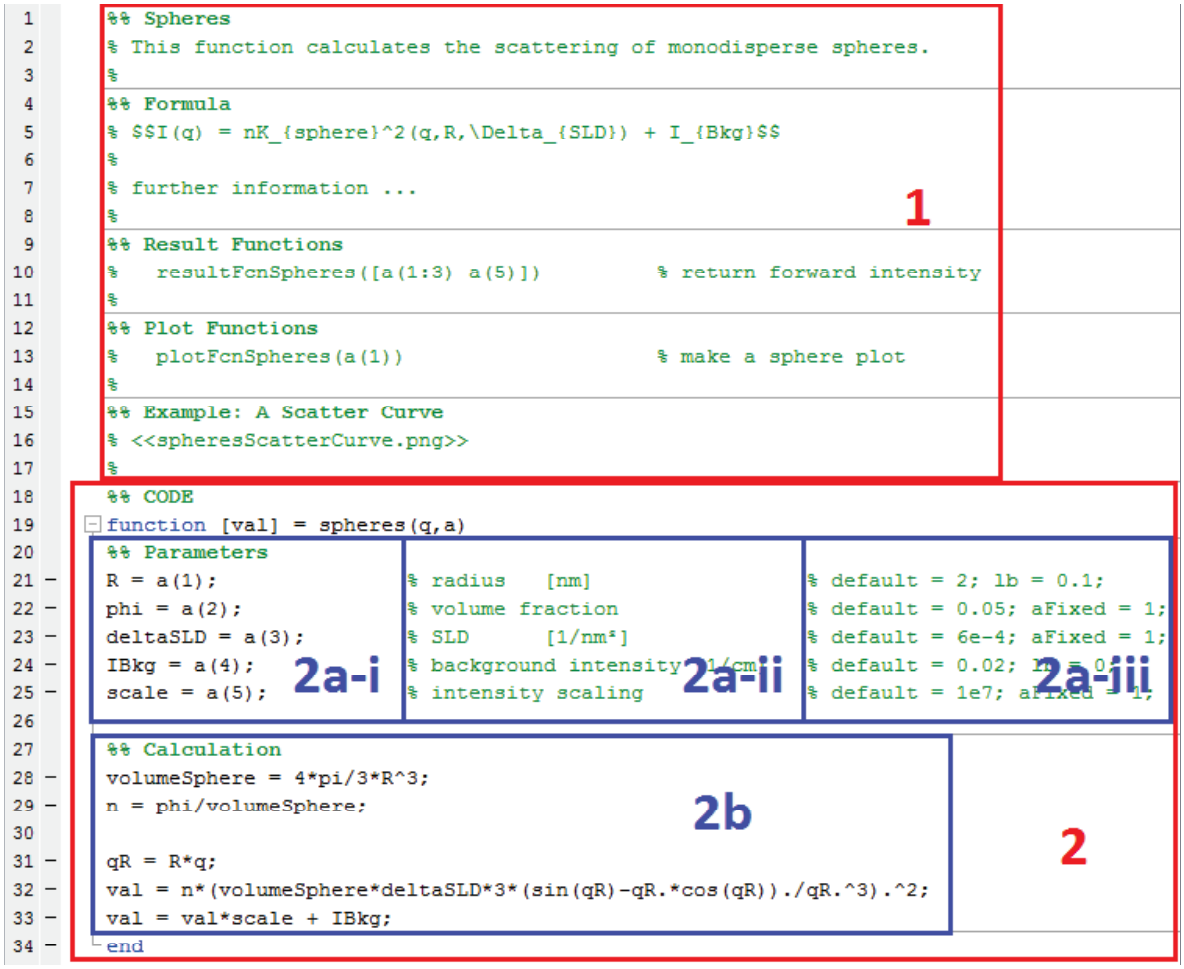


Figure 9: An exemplary model function file, the associated help is depicted in Fig. 10. The function file is divided into two main blocks. Block 1 is used to create the main function help/description. This block is partitioned into cell blocks, where a cell starts with '%%'. Each cell is a separate section within the function help, where the string behind '%%' gives the section title. In the first cell (from top) a general description of the model function file is given. In the second cell a formula is written in LaTeX, which produces a formula in the help file. In the 5th cell a command is used that includes an image in the help file. The 3rd and 4th cells are information that allow to call result functions or plot functions — see the text. Block 2 defines the sphere function. Herein **a** is a parameter vector with five elements that are specified in block 2a-i, and the appropriate parameter descriptions are given in block 2a-ii. In block 2a-iii start fit values (default = ...;), lower (lb = ...;) and upper (ub = ...;) bounds are defined as well as an information whether a parameter shall be fitted or not (aFixed = ...;). The sphere function is calculated in block 2b.

Spheres

This function calculates the scattering of monodisperse spheres.

Contents

Formula

$I(q) = nK^2_{sphere}(q, R, \Delta_{SLD}) + I_{Bkg}$

further information ...

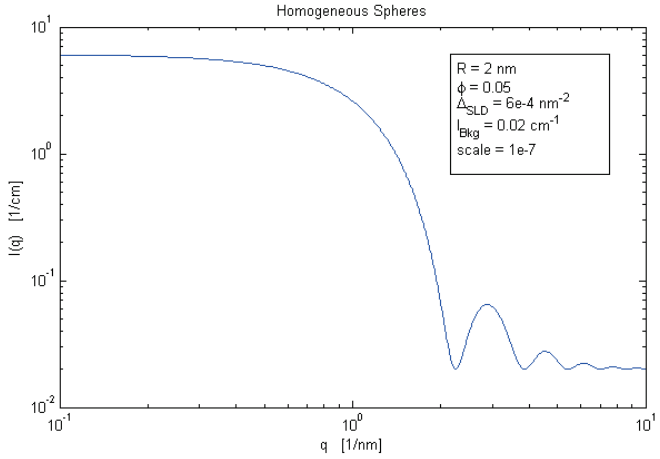
Result Functions

```
resultFcnSpheres([a(1:3) a(5)])          % return forward intensity
```

Plot Functions

```
plotFcnSpheres(a(1))                     % make a sphere plot
```

Example: A Scatter Curve



CODE

```
function [val] = spheres(q,a)
```

Parameters

```
R = a(1);               % radius    [nm]               % default = 2; lb = 0.1;
phi = a(2);             % volume fraction              % default = 0.05; aFixed = 1;
deltaSLD = a(3);        % SLD       [1/nm^4]           % default = 6e-4; aFixed = 1;
IBkg = a(4);            % background intensity [1/cm]  % default = 0.02; lb = 0;
scale = a(5);           % intensity scaling            % default = 1e7; aFixed = 1;
```

Calculation

```
volumeSphere = 4*pi/3*R^3;
n = phi/volumeSphere;

qR = R*q;
val = n*(volumeSphere*deltaSLD*3*(sin(qR)-qR.*cos(qR))./qR.^3).^2;
val = val*scale + IBkg;
```

```
end
```

Figure 10: The help file associated to the function file in Fig. 9.

3.2. Fitting of Physical Models

Fitting a physical model to some scattering data is an instance of regression. As outlined in section 2.2, there are two statistical inference approaches, a frequentist one and a Bayesian one, that can be used to perform regression. In both approaches a likelihood function needs to be defined, and as explained in section 2.2, it can be described by a noise model together with a model function. In SASET a weighted square sum of errors is used as the objective function, cf. Eq. (66),(68), and thus a heteroscedastic Gaussian noise model is implicitly assumed. In practice, the Gaussian noise model is often a good choice, because of the Central Limit Theorem, which states that a sum of independent random numbers becomes more and more Gaussian distributed as the number of random numbers increases. In SANS single neutrons are counted, and hence the intensity measured by a detector pixel can be described by a Poisson distribution. Hence, if many neutrons ($\gtrsim 20$) are counted the distribution is well described by the Gaussian one. Also if less neutrons are counted the involved transmission correction procedure additionally leads to a more Gaussian like distribution. For SAXS (or light scattering) one can usually consider a flow of photons (very high number of photons) and the Gaussian noise assumption should be a rather good one (aside from detector non-ideal behavior).[20]

In the frequentist inference the goal is to minimize the generalization error, Eq. (40) (respectively Eq. (51)), and as long as no overfitting occurs this can be accomplished by minimizing the empirical error, Eq. (41), respectively encoded in Eq. (66),(68). If overfitting occurs the generalization error can be estimated by methods like cross-validation, see section 2.2.

As explained in section 2.2 overfitting describes the case when the model function follows the noise in the data. In case of fitting a physical model function to a scattering curve overfitting usually does not occur[21], since

(i) a scattering curve is 'oversampled', meaning that the data curve is still very well represented, if, e.g., only every second data point from the original data curve is taken into account, and

(ii) the number of fit parameters is small, and additionally often the parameters are very likely to be correlated, meaning that there is less freedom to vary independently the parameters in order to get a better fit.

Consequently, the generalization error can be estimated by the empirical error. Accordingly, if there are a number of different physical model fits, the fit having the smallest empirical error will define the optimal model. Here, the classical frequentist inference finishes.

However, performing model selection only on the criterion of having the lowest generalization error may not be a good approach, since no a priori knowledge is taking into account. This can then lead to obscurities and inconsistencies (cf. also footnote 13, page 18), e.g.,

(a) in general two different scattering models may produce the same intensity curve, but then there is no clue what model is the true or better one, and

[20] Also if the Gaussian noise assumption is often a good one in practice, sometimes one may experience outlier points. In this case the Gaussian noise model is not a good choice, since the Gaussian has rather short distribution tails, which then lead to a huge contribution of the outlier points to the objective function, and which may then lead to a wrong inference of the shape of the regression function. This problem can be circumvented with distributions having longer tails (e.g., the Lévy distribution or the Cauchy (Lorentzian) distribution), and hence give a more robust estimation [129, §15.7].

[21] For the IFT method in section 7 the case is different: The data curves are still oversampled, but there are many free-form parameters to be adjusted that are only weakly interdependent. In this case overfitting can occur and cross-validation becomes useful.

(b) a model may have a lower fitting error as another model, yet be physically much less likely.

For case (a) the principle of Occam's razor (cf. section 2.2) can be applied, hence the simplest model should be selected; but simplicity can be considered within a mathematical model description (e.g., a model could have less parameters as the other one) as well as within a physical model description (e.g., the physical surface of a model could be less smooth as the one from the other model, and thus it might have a higher surface energy, and therefore this model might be less plausible). Indeed, model selection should respect the physical model assessment and the mathematical one, but there is no consistent approach to accomplish this within the frequentist school. In case (b) the fits must be analyzed with respect to where the mismatch of a model is coming from in order to subsequently perform model selection. As explained above the mismatch is usually not resulting from an overfitted model function, instead it is because of underfitting occurs, meaning that there is a systematic error within the fit, which results from a wrong physical model description[22] (often a too simple one) of the scattering system (e.g., assuming spheres instead of ellipsoids) or of the scattering conditions (e.g., wrong or missing resolution function, existence of multiple scattering, aging of sample during measurement). E.g., if a fit is wrong in the low q-regime, and the measured intensity decreases in this regime, it might be that a structure factor of interacting particles is badly modeled instead of having big core-shell objects that produce a similar intensity but are very unlikely. Then, based on these considerations (i) model selection should be done or (ii) the structure factor should be better modeled and then, this new physical model has to be refitted to the data.

Last but not least, the question arises how trustworthy the inferred model parameters are. Since a frequentist approach is considered many i.i.d. data sets need to be generated (as discussed in section 2.2), which can be obtained by repeating the experiment many times or by an artificial generation via, e.g., the bootstrap method [129, §15.6.2]. Then, on each of these data sets a model fit must be performed, which finally yields a joint parameter distribution. The uncertainty of a parameter is then obtained by taking the standard deviation of its marginalized distribution. Furthermore, the correlation of two parameters can be gained by determining the covariance of the joint two-parameter distribution (in which the other parameters are marginalized out). However, since the scattering data curves are oversampled, the width of the distribution can be expected to be very small, except if the scattering data do not contain any well defined features at all (e.g., if the scattering data values are nearly constant and a Gaussian distribution shall be fitted to them, then the width of the Gaussian distribution can be quite uncertain).

Another frequentist approach to describe the goodness of the determined parameters is given by inspecting the curvature of the likelihood function at its maximum [122, §2.5, 3.4]. Therein, the idea is that the logarithm of the likelihood function, called log-likelihood, can often be well described by a quadratic function around its maximum likelihood solution. Then, if the modulus of the curvature is high, it corresponds to a strong peak, and hence 'intuitively' [122, §2.5] indicates less uncertainty about the parameters as if the curvature is low. The curvature of the negative log-likelihood is given by the Hessian matrix of the negative logarithm of the likelihood, Eq. (44), i.e., $-\nabla\nabla \ln \mathrm{pdf}(\mathbf{y}|\mathbf{x}, \mathbf{w})$, which is called observed (Fisher) information matrix [180, §3.2.1], and the square root of its inverse can be related to the width of the peak, hence can be used to define parameter uncertainties. No more details

[22] A poor data reduction can also lead to poor data. In this case the data reduction has to be improved — for a discussion on the many correction steps (a number of them are actually rarely performed) in data reduction for SA(X)S can be found in [121] and [146].

are given here, since in the Bayesian approach below, a very similar discussion is given about uncertainties of the determined parameters; then from Eq. (79) by setting the second term on the r.h.s. to zero (i.e., the Bayesian prior is removed) parameter uncertainties are quantified for the frequentist approach.

In the Bayesian inference, the main goal is the determination of the joint a posteriori distribution $\mathrm{pdf}(\mathbf{w}_k, \boldsymbol{\theta}_k, H_k|\mathcal{D})$, Eq. (55). Subsequently, based on the determined a posteriori distribution model selection can be performed. Since a physical model is considered, it seems logical to perform hierarchical model selection via Eq. (62)-(64), as discussed in section 2.2. However, such an approach can easily be very compute-intensive, since the involved likelihood functions are nonlinear (because of the inherent nonlinear physical models), and the formula requires the calculation of many nested integrals: E.g., consider that a 'simple' physical model function has two inherent integrations (e.g., one for the model itself and another one for a resolution function), and that the model has two adjustable parameters, then the inference formula may end up with six nested integrations if addtionally two hyperparameters are considered for the two parameters. Usually, only deterministic approximate inference methods as well as MC methods can solve such a problem, cf. section 2.2 and 4.1. Note that further integrations are required for the other considered models. However, the number of integrations can be reduced if the hyperparameters are considered to be fixed. Then, in the given example only four nested integrations need to be calculated, but this is still a demanding problem.

In order to get a more tractable solution the following is considered: As explained above, the models are usually underfitted, i.e., there is a systematic error, and as discussed in section 2.2, the global maximum, Eq. (65), can then be used to identify the optimal model, hyperparameters, and parameters simultaneously. For simplicity it is assumed that the hyperparameters are fixed (respectively do not exist), then the a posteriori distribution can be written as

$$\begin{aligned}\mathrm{pdf}(\mathbf{w}_k, H_k|\mathcal{D}) &= \frac{\mathrm{pdf}(\mathcal{D}|\mathbf{w}_k, H_k)\mathrm{pdf}(\mathbf{w}_k, H_k)}{\mathrm{pdf}(\mathcal{D})} \\ &= \frac{\mathrm{pdf}(\mathcal{D}|\mathbf{w}_k, H_k)\mathrm{pdf}(\mathbf{w}_k|H_k)\mathrm{pdf}(H_k)}{\mathrm{pdf}(\mathcal{D})}.\end{aligned} \tag{71}$$

Hence, its maximization gives the joint MAP solution

$$(\mathbf{w}_{k^*}^*, H_{k^*}) = \arg \max_{(\mathbf{w}_k, H_k)} \left\{\mathrm{pdf}(\mathcal{D}|\mathbf{w}_k, H_k)\mathrm{pdf}(\mathbf{w}_k|H_k)\mathrm{pdf}(H_k)\right\}, \tag{72}$$

which can be accomplished by a sequential optimization where at first for each model the optimal parameter vector is determined

$$\mathbf{w}_k^* = \arg\max_{\mathbf{w}_k} \{\mathrm{pdf}(\mathcal{D}|\mathbf{w}_k, H_k)\mathrm{pdf}(\mathbf{w}_k|H_k)\}, \tag{73}$$

and then the optimal model can be selected via

$$H_k^* = \arg\max_{H_k} \{\mathrm{pdf}(\mathcal{D}|\mathbf{w}_k^*, H_k)\mathrm{pdf}(\mathbf{w}_k^*|H_k)\mathrm{pdf}(H_k)\}. \tag{74}$$

Eq. (73) can be seen a regularized optimization (cf. the Gaussian example in Eq. (35)), which becomes a non-regularized optimization if $\mathrm{pdf}(\mathbf{w}_k|H_k)$ is considered to be constant: in this case maximization is accomplished equivalently by minimizing the empirical error, Eq. (41), respectively encoded in Eq. (66),(68). Based on physical considerations and information gained from other experiments, upper and lower parameter bounds can often be used

in the conditional a priori probability distribution $\mathrm{pdf}(\mathbf{w}_k|H_k)$. Specifying $\mathrm{pdf}(H_k)$ is usually not so easy, since this requires a detailed physical understanding of the system together with a method of how to encode this knowledge as a probability. However, if there is knowledge from other experiments stating that the relative frequency of occurrence of model H_k is r_k, then $\mathrm{pdf}(H_k) = r_k$.

The trustworthiness of an optimally determined parameter vector $\mathbf{w}_k^*$ for a given model H_k is fully described by the conditional a posteriori distribution $\mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})$. The functional behavior of $\mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})$ around the mode $\mathbf{w}_k^*$ can be locally approximated via a Taylor expansion. However, a common practice is to perform a Taylor expansion on the logarithm of $\mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})$, since $\mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})$ is often a very 'peaky' function around the mode [141, §2.2] that can often be well described by a Gaussian distribution. Hence, the local approximation gives

$$\ln \mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D}) \approx \ln \mathrm{pdf}(\mathbf{w}_k^*|H_k, \mathcal{D}) + \frac{1}{2}(\mathbf{w}_k - \mathbf{w}_k^*)^T \left[\nabla\nabla \ln \mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})\right]_{\mathbf{w}_k^*} (\mathbf{w}_k - \mathbf{w}_k^*), \tag{75}$$

and approximately yields

$$\mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D}) \propto \exp\left\{\frac{1}{2}(\mathbf{w}_k - \mathbf{w}_k^*)^T \left[\nabla\nabla \ln \mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})\right]_{\mathbf{w}_k^*} (\mathbf{w}_k - \mathbf{w}_k^*)\right\} \tag{76}$$

around $\mathbf{w}_k^*$. Herein, $\left[\nabla\nabla \ln \mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})\right]_{\mathbf{w}_k^*}$ is the Hessian matrix of $\ln \mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})$ evaluated at $\mathbf{w}_k^*$, and defines the curvature of $\ln \mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})$ at $\mathbf{w}_k^*$. Eq. (76) has the form of a multidimensional Gaussian distribution [126]. Therefore, the negative of the inverse Hessian matrix can be identified as the covariance matrix of a Gaussian distribution

$$\mathrm{cov}(\mathbf{w}_k^*) = -\left(\left[\nabla\nabla \ln \mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})\right]_{\mathbf{w}_k^*}\right)^{-1}, \tag{77}$$

and the missing constant of proportionality in Eq. (76) is

$$\sqrt{\frac{\det\left(\left[\nabla\nabla \ln \mathrm{pdf}(\mathbf{w}_k|H_k, \mathcal{D})\right]_{\mathbf{w}_k^*}\right)}{(2\pi)^{M_k}}}, \tag{78}$$

where M_k is the number of elements in $\mathbf{w}_k$. This approach (fitting a Gaussian to a mode of a distribution) is called Laplace approximation [16, §4.4]. The variance of the m-th parameter is given by $[\mathrm{cov}(\mathbf{w}_k^*)]_{mm}$, hence its standard deviation is $\sqrt{[\mathrm{cov}(\mathbf{w}_k^*)]_{mm}}$. The off-diagonal elements in the covariance matrix define parameter-parameter covariances. As an example of the Laplace approximation, consider the Gaussian posterior in Eq. (34) of section 2.2: At first, a mode needs to be found by a nonlinear optimization procedure. Then, given a mode the covariance reads

$$\mathrm{cov}(\mathbf{w}_k^*) = \left(\left[\nabla\nabla\left(\frac{1}{2}\sum_{n=1}^{N}\frac{(y_n - f(x_n, \mathbf{w}_k))^2}{\sigma^2}\right)\right]_{\mathbf{w}_k^*} + \gamma \mathbf{1}_{M_k}\right)^{-1}, \tag{79}$$

where $\mathbf{1}_{M_k} \in \mathbb{R}^{M_k \times M_k}$ is the identity matrix of dimension M_k.

Remarks:

(i) the Gaussian approximation of the a posteriori distribution around a mode is often a reasonable assumption, but

(*ii*) because of the non-linearity of the physical models many modes might exist — in this case separate Laplace approximations can be performed for each mode, in order to describe the complete posterior distribution. Alternatively, MC methods can be used to sample from the posterior distribution, and hence to infer the shape of it, cf. section 4.1.

Issue (*i*) can be explained with the fact that if the data are independent, the logarithm of the likelihood defines a sum of independent random numbers, which yields — due to the Central Limit Theorem — a Gaussian variable if the number of data points increases. Hence, if the prior is relatively flat or if the number of data points N is so high that the likelihood dominates the prior, the posterior will become a Gaussian distribution. Moreover, with an increase of data points the likelihood increases and will ultimately dominate the prior term, and hence the covariance will decrease[23] (because of the inversion of the Hessian matrix in Eq. (77)). Because of these facts, and because of scattering curves are oversampled and the physical models are not overfitted, it can be expected that the covariances are relatively small.

Last but not least, from the above frequentist and Bayesian discussions it follows that the global minimum of the generalization error or the global maximum of the joint MAP defines the optimal solution; however, in practice the global optimum might not be desired, since:

(*i*) Often, because of simplicity, only lower and upper parameter bounds are used in order to describe the domain of the objective function, whereas in general only sophisticated non-linear constraints would be able to describe the domain properly. As a consequence of this, an optimization routine may find a solution that lies outside of the physical domain of the model function, and hence has to be rejected.

(*ii*) Usually, underfitting occurs, i.e., the physical model is wrong even if the fitted model corresponds to a global optimum. In this case another locally optimal solution can still be physically more meaningful than the globally optimal solution.

Of course, in the Bayesian approach these two points define inconsistencies, but only occur if the a priori information is not properly encoded in the likelihood and the a priori distribution.

3.2.1. Results

Inferring physical models and model parameters within frequentist and Bayesian inference is discussed. It has been shown that:

(*i*) Because of the usual underfitting of the physical models

  (*a*) the empirical error is sufficient as objective function (for determining the mathematical optimal model and its parameters) in the frequentist approach or alternatively

  (*b*) the joint maximum a posteriori distribution in the Bayesian approach can be used to infer the optimal model and parameters.

(*ii*) A priori knowledge about physical models is essential to perform (physical) model selection, and it is especially important here, since an inverse (scattering) problem needs to be solved. In the Bayesian approach a priori knowledge has to be encoded in the a priori distributions, while only ad-hoc/intuitive methods can accomplish this in the frequentist approach.

[23] An illustrative example is given in [16, §3.3].

(*iii*) Parameter uncertainties are usually expected to be quite small. However, these uncertainties might not have much/any meaning if the model is wrong [129, §15.0], which is usually the case because of an underfitting (there is a systematic error in the description of the physical model).

A full hierarchical Bayesian model approach does not only allow to take into account physical a priori knowledge, but the model selection scheme also penalizes mathematically complex models more than simpler ones — hence implementing the principle of Occam's razor in a mathematical sense. In any case, this full scheme is usually not tractable, but the simpler joint maximum a posteriori approach can in principle be performed, because of the expected underfitting of the physical models. However, the problem of having physical a priori knowledge at hand, allowing to quantify a priori model and parameter probabilities, is often missing (usually, only parameter bounds can be easily specified) and knowledge can be difficult to encode in the formula. Because of these practical reasons the frequentist approach is still commonly chosen.

3.3. Example: Simultaneous Fitting of SANS Data of PIB-PMAA / P4VPQ IPECs

In this section, evaluations of SANS data in external contrast variation are described for micelles resulting from the mixing of oppositely charged polymers that are known to build InterPolyElectrolyte Complexes (IPECs). This evaluation is done as a demonstration of the powerfulness of the program SASET. The main questions that should be clarified in this work is how the considered IPEC structures look like, e.g., (*i*) are they core-shell or core-shell-corona structures, and in the latter case can the density profile of the corona be determined, and (*ii*) how compact is the shell?

The considered systems consist of the diblock copolymer polyisobutylene-b-poly(methacrylic acid) (IB_x−MAA_y, PIB-b-PMAA) where the polyacid has been neutralized into the sodium salt of the polyanion PMANa via equimolar addition of NaOH, and the polymer P4VP (poly(4-vinylpyridine), which is quaternized at $\alpha = 90\,\%$ with C_2D_5Br into the polycationic poly(N-ethyl-4-vinylpyridinium bromide) (d5-P4VPQBr_α) (i.e., not all 4VP units bear a charge).

The PMANa block of the copolymer is negatively charged, therefore strong electrostatic interactions will take place between this block and the positively charged polycation d5-P4VPQBr_α, yielding a (PMANa+d5-P4VPQBr_α) complex. The PIB block is hydrophobic. The complex is characterized by the ratio Z between the number of positive and negative polymer charges, Q^+, Q^-:

$$Z = \frac{Q^+}{Q^-} = \frac{\text{moles of 4VPQ units}}{\text{moles of MA units}}. \tag{80}$$

In all sample systems the overall charge ratio is at 0.4, i.e., PMANa is in large excess. The aqueous solution resulting from the mixing of these two polymers was observed to be transparent and stable against sedimentation for low Z. It is expected, since already observed for a very similar system in [125], that colloidal nanoparticles are formed which have a spherically symmetrical structure. Accordingly, spherically symmetrical core-shell, core-corona or core-shell-corona structures are expected and modeled as schematically illustrated in Fig. 11, and following model assumptions are made: The core is hydrophobic and made of PIB. In the core-shell and core-corona models the outer part (i.e., the shell or corona) consists of the complex and hydration. In the core-shell-corona models the shell consists of the complex at charge stoichiometry (and hydration), and the corona consists of PMANa (and hydration).

Three different systems are analyzed consisting of different degrees of polymerization of PMAA and of PIB. The appropriate data sets are listed in table 1 together with the employed D_2O:H_2O ratios. In short, three diblock copolymers are used: for sets 1 and 2 the hydrophobic blocks are the same (75 units), and for sets 1 and 3 the hydrophilic blocks are the same (190 units), while for set 2 the hydrophilic block is almost an order of magnitude larger. It is assumed that there is no effect from isotopic substitution in the solvent, such that one specific geometric aggregate model can be employed for the different D_2O:H_2O ratios.

All samples are at $10\,\mathrm{g{\cdot}dm^{-3}}$ in PIB-PMAA, in a buffer of NaCl $0.1\,\mathrm{mol{\cdot}dm^{-3}}$ with TRIS (tris(hydroxymethyl)aminomethane, for pH adjustment) $0.01\,\mathrm{mol{\cdot}dm^{-3}}$. The total volume fraction of polymer (PIB-PMANa and d5-P4VPQBr_α) is around 1 – 2 %.

The samples were measured on D11 at ILL by Markus Burkhardt in March 2005 [27]. The apparent aqueous densities of polymers were not measured at that time. However, densitometry measurements were performed by Dr. Sylvain Prévost (his results are described below) in the same concentration range in aqueous (H_2O) buffer on materials as similar as

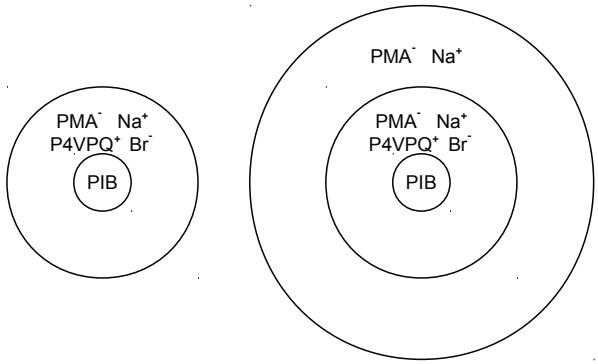


Figure 11: Schematic IPEC model structures. Left: core-shell / core-corona model: the core consists of PIB, and the shell, respectively the corona, consists of the complex (and hydration). Right: core-shell-corona model: the core consists of PIB, the shell consists of the complex (and hydration), and the corona consists of PMANa (and hydration).

possible, but not on the diblock copolymer PIB-PMAA whose synthesis, purification and characterization is time-consuming.

The density of PIB was taken from the thesis of Markus Burkhardt [27] as being $0.918\,\mathrm{g{\cdot}cm^{-3}}$. The scattering length of IB (one unit of PIB) for neutrons is -3.33 fm, giving with $0.918\,\mathrm{g{\cdot}cm^{-3}}$ a SLD of $-0.33{\cdot}10^{-4}\,\mathrm{nm^{-2}}$.

The density of PMANa was measured on commercially available polymer[24] at 0 – 3 % and found to be $2.4\,\mathrm{g{\cdot}cm^{-3}}$ in buffer. The scattering length of MANa (one unit of PMANa) for neutrons is 23.13 fm, giving with $2.4\,\mathrm{g{\cdot}cm^{-3}}$ a SLD of $3.2{\cdot}10^{-4}\ \mathrm{nm^{-2}}$.

Since P4VPQBr is commercially not available, the quaternization of P4VP with (hydrogenated) ethylene bromide has been carried out by Michaela Dzionara[25]. The NRM spectrum suggest that around 90 % of polymer units have been effectively quaternized, see appendix C.1. The density of $\mathrm{P4VPQBr}_\alpha$ was measured in buffer and found to be $1.44\,\mathrm{g{\cdot}cm^{-3}}$, corresponding to $1.5\,\mathrm{g{\cdot}cm^{-3}}$ for d5-$\mathrm{P4VPQBr}_\alpha$. The scattering length of d5-4VPQBr$_\alpha$ (one unit of d5-$\mathrm{P4VPQBr}_\alpha$) for neutrons is 77.807 fm, giving with $1.5\,\mathrm{g{\cdot}cm^{-3}}$ a SLD of $3.38{\cdot}10^{-4}\ \mathrm{nm^{-2}}$.

Solutions of the complex (PMANa+$\mathrm{P4VPQBr}_\alpha$) are cloudy and start phase-separating within minutes; two clear phases (low viscous upper phase, viscous lower phase) are obtained within hours. Densitometry measurements indicate a strongly non-ideal behavior. At $Z = 0.4$, the apparent density of the stoichiometric complex (assuming that the excess of PMANa has the same apparent density as without $\mathrm{P4VPQBr}_\alpha$) is found to be $1.5\,\mathrm{g{\cdot}cm^{-3}}$, too. If subtracting from the complex the apparent volume of NaBr (which has an apparent density of $4.35\,\mathrm{g{\cdot}cm^{-3}}$), the apparent density of the complex is $1.13\,\mathrm{g{\cdot}cm^{-3}}$ (for hydrogenated $\mathrm{C_2H_5}$) or $1.16\,\mathrm{g{\cdot}cm^{-3}}$ (for deuterated $\mathrm{C_2D_5}$). Assuming 90 % of quaternization, the deuterated version will have a neutron coherent scattering length of 109.58 fm including counterions $\mathrm{Na^+}$ and $\mathrm{Br^-}$ (99.15 fm without counterions), giving with $1.5\,\mathrm{g{\cdot}cm^{-3}}$ (respectively $1.16\,\mathrm{g{\cdot}cm^{-3}}$) a SLD

[24] Chemical from Aldrich. Information from the provider: reference 434507-1L lot # MKBM1345V, solution of poly(methacrylic acid, sodium salt) at 30 % in $\mathrm{H_2O}$ by weight, number average molecular weight $\mathrm{M_n} \approx 5400\,\mathrm{g{\cdot}mol^{-1}}$, mass average molecular weight $\mathrm{M_w} \approx 9500\,\mathrm{g{\cdot}mol^{-1}}$ by GPC (Gel Permeation Chromatography).

[25] P4VP was from Sigma-Aldrich (ref.:25232-41-1, weight-average molecular mass 60 000 $\mathrm{g{\cdot}mol^{-1}}$) and used as received. 10 g (0.0950 mmol) of the polymer and 11.39 g (0.1045 mmol) bromoethane were solubilized in 80 mL methanol and heated up at 70 °C for 5 days under reflux. After cooling, the polymer was precipitated with 100 mL diethylether, and the precipitate was dried under vacuum at 30 °C for one day.

	DP_k(PIB)	DP_k(PMAA)	D_2O:H_2O ratios
set 1	75	190	100:0, 80:20, 60:40, 50:50, 0:100
set 2	75	1590	100:0, 80:20, 60:40, 40:60, 0:100
set 3	30	190	100:0, 80:20, 60:40, 40:60

Table 1: Used IPEC systems. DP_k: degree of polymerization (k is the average number of units).

of $2.90{\cdot}10^{-4}$ nm^{-2} (respectively $2.93{\cdot}10^{-4}$ nm^{-2}). Hence, for neutron scattering, it does not matter where the sodium and bromide counterions are; note that the situation would be completely different for X-rays, which are sensitive to electronic densities. In the following, the counter-ions will be considered as part of the complex.

Densitometry data seem to indicate that SLDs of PMANa ($3.2{\cdot}10^{-4}$ nm^{-2}), d5-P4VPQBr$_\alpha$ 90 % ($3.38{\cdot}10^{-4}$ nm^{-2}) and the stoichiometric complex ($2.90{\cdot}10^{-4}$ nm^{-2} or $2.93{\cdot}10^{-4}$ nm^{-2} respectively including or excluding counterions) are hardly distinguishable by neutrons.

3.3.1. Models

In order to analyze the SANS curves, different spherically symmetrical amplitude form factors have been implemented in SASET, which model the scattering amplitudes of core-shell, core-(linear corona), core-(algebraic corona), core-shell-(linear corona) and core-shell-(algebraic corona) objects. The scattering length density (SLD) profiles of such objects are exemplified in Fig. 12. Moreover, following assumptions are made:

1. It is assumed that within one data set (cf. table 1) the IPECs are reproducible. This essentially means that there is no effect from isotopic substitution[26], and that the other framework conditions are as similar as possible (stoichiometric composition, temperature, etc.).

2. H and D are homogeneously distributed in the solvent.

Core: In all models it is assumed that the core solely consists of hydrophobic PIB. Therefore, the SLD in the core, having radius R_i, is the one of PIB, $\mathrm{SLD_{PIB}}$, and the amplitude form factor is a homogeneous sphere one:

$$\begin{aligned} A_{\mathrm{sphere}} &= \int_0^{R_i} 4\pi r^2 \frac{\sin(qr)}{qr} [\mathrm{SLD_{PIB}} - \mathrm{SLD_{matrix}}]\, \mathrm{d}r \\ &= 4\pi(\mathrm{SLD_{PIB}} - \mathrm{SLD_{matrix}}) \frac{\sin(qR_i) - qR_i\cos(qR_i)}{q^3}, \end{aligned} \tag{81}$$

where $\mathrm{SLD_{matrix}}$ is the SLD of the matrix.

Shell: For the shell, the following two extreme cases are considered:

(*i*) Either the shell consists of all polyelectrolytes (i.e., the total PMANa and d5-P4VPQBr$_\alpha$) with solvent, i.e., it has a Z ratio of 0.4, or

(*ii*) the shell consists of the solvated stoichiometric complex (PMANa+d5-P4VPQBr$_\alpha$) with $Z = 1$, and the excess of PMANa ($\approx 60\,\%$) is contained in an outer corona.

[26] Effects from isotopic substitution have been observed, e.g., in [147].

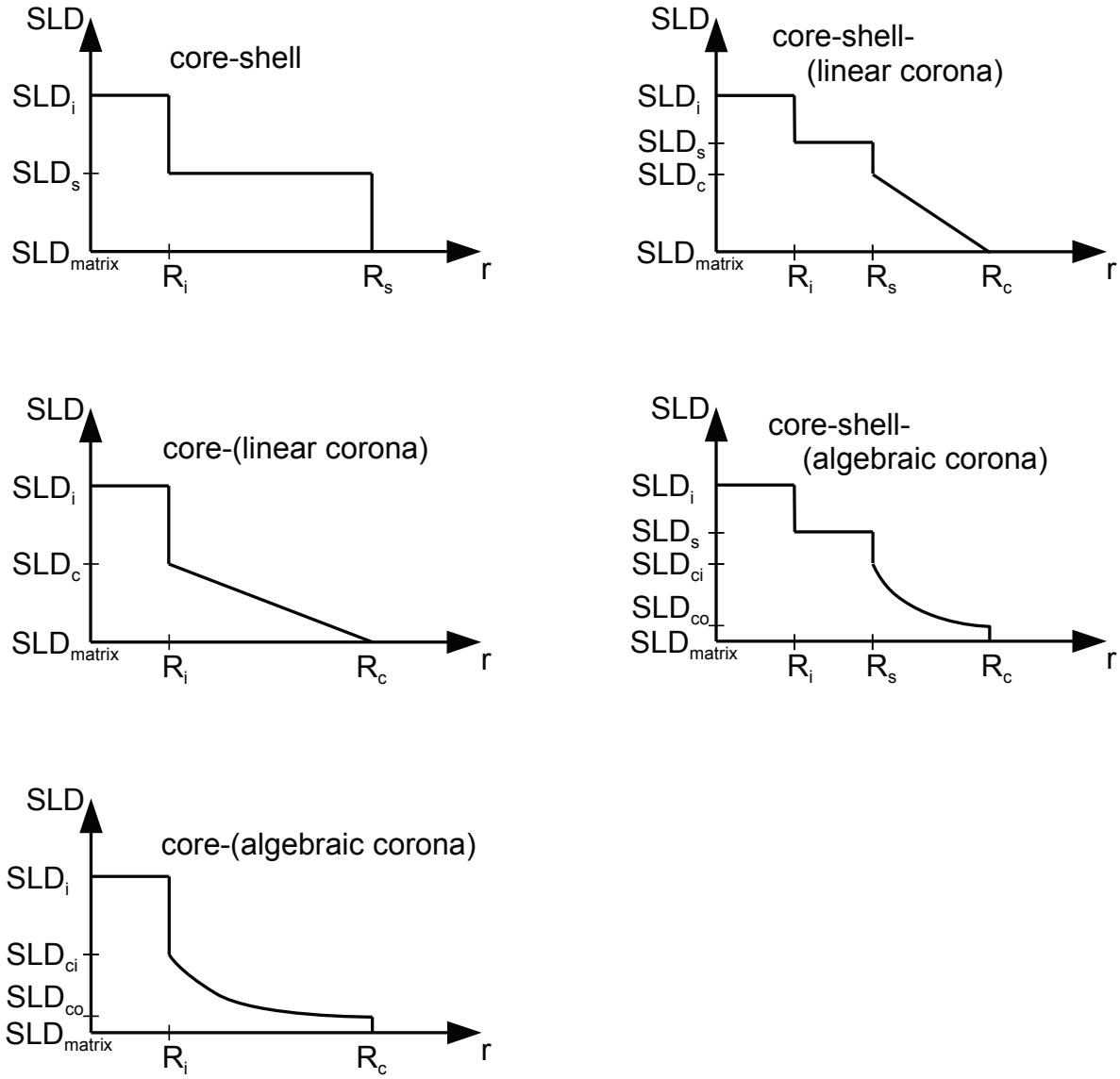


Figure 12: Schematic SLD profiles of the assumed IPEC objects. SLD_i: SLD in the core. SLD_s: SLD in the shell. SLD_c, and SLD_{ci}: SLD in the corona at the interface core-corona or shell-corona. SLD_{co}: SLD in the corona at the interface corona/matrix. SLD_{matrix}: SLD of the matrix. R_i: core radius. R_s: shell radius. R_c: corona radius.

The volume fraction of material (PMANa+P4VPQBr) in the shell is

$$\varphi_m = 1 - \varphi_h, \tag{82}$$

where φ_h is the volume fraction of shell hydration. Hence, the SLD in the shell is

$$\mathrm{SLD}_{\mathrm{shell}}(r) = \varphi_m \mathrm{SLD}_m + \varphi_h \mathrm{SLD}_{\mathrm{matrix}}, \qquad R_i \le r \le R_s, \tag{83}$$

where SLD_m is the SLD of the material, i.e., the complex. Accordingly, the shell amplitude form factor reads

$$\begin{aligned} A_{\mathrm{shell}} &= \int_{R_i}^{R_s} 4\pi r^2 \frac{\sin(qr)}{qr} [\mathrm{SLD}_{\mathrm{shell}}(r) - \mathrm{SLD}_{\mathrm{matrix}}] \, \mathrm{d}r \\ &= 4\pi(1-\varphi_h)(\mathrm{SLD}_m - \mathrm{SLD}_{\mathrm{matrix}}) \Bigg\{ \frac{\sin(qR_s) - qR_s\cos(qR_s)}{q^3} \\ &\qquad - \frac{\sin(qR_i) - qR_i\cos(qR_i)}{q^3} \Bigg\}. \end{aligned} \tag{84}$$

Corona: In the models that have a corona but no shell, the corona consists of the complex (PMANa+P4VPQBr) and is hydrated. In the models with a corona and a shell, the corona only consists of PMANa and is hydrated.

Consider a linear corona in $[R_1, R_c]$, where $R_1 = R_i$ in case of the core-(linear corona) model, and $R_1 = R_s$ in case of the core-shell-(linear corona) model. The linear corona is characterized by the fact that the volume fraction of material in the corona (i.e., PMANa+P4VPQBr for the core-(linear corona) model and PMANa for the core-shell-(linear corona) model) is linearly decreasing with increasing r:

$$\varphi_{\mathrm{lin.}}(r) = \varphi'_{\mathrm{int.}} \left(1 - \frac{r - R_1}{R_c - R_1} \right) \quad \text{for } R_1 \le r \le R_c, \tag{85}$$

where $\varphi'_{\mathrm{int.}}$ is the interfacial volume fraction[27] (i.e., surface fraction) of material at R_1+0. $\varphi'_{\mathrm{int.}}$ is given by the fraction of a sphere (with radius R_i, respectively R_s for the core-shell-corona models) that is penetrated by PMANa chains (the PMANa cross-section is calculated from the molecular volume (which is determined from the molecular mass and the fitted density) and the PMANa length (the monomer unit length is taken to be 0.252 nm [78]). Thus, the SLD in the corona reads

$$\begin{aligned} \mathrm{SLD}_{\mathrm{lin.}}(r) &= \varphi_{\mathrm{lin.}}(r)\mathrm{SLD}_m + [1 - \varphi_{\mathrm{lin.}}(r)]\mathrm{SLD}_{\mathrm{matrix}} \\ &= \varphi_{\mathrm{lin.}}(r)\Delta\mathrm{SLD} + \mathrm{SLD}_{\mathrm{matrix}}, \end{aligned} \tag{86}$$

where

$$\Delta\mathrm{SLD} := \mathrm{SLD}_m - \mathrm{SLD}_{\mathrm{matrix}}, \tag{87}$$

and SLD_m is the SLD of the considered material. Accordingly, the amplitude form factor is

$$\begin{aligned} A_{\mathrm{lin.\ corona}} &= \int_{R_1}^{R_c} 4\pi r^2 \frac{\sin(qr)}{qr} [\mathrm{SLD}_{\mathrm{lin.}}(r) - \mathrm{SLD}_{\mathrm{matrix}}] \, \mathrm{d}r \\ &= 4\pi\varphi'_{\mathrm{int.}}\Delta\mathrm{SLD} \int_{R_1}^{R_c} r^2 \frac{\sin(qr)}{qr} \frac{r - R_c}{R_1 - R_c} \, \mathrm{d}r \\ &= 4\pi\varphi'_{\mathrm{int.}}\Delta\mathrm{SLD} / \big\{ (R_1 - R_c) q^4 \big\} \\ &\times \big\{ qR_c\sin(qR_c) + 2\cos(qR_c) + qR_c\sin(qR_1) - q^2 R_1 R_c \cos(qR_1) \\ &\quad + q^2 R_1^2 \cos(qR_1) - 2\cos(qR_1) - 2qR_1\sin(qR_1) \big\}. \end{aligned} \tag{88}$$

[27] Note: For the core-corona models $\varphi'_{\mathrm{int.}} = 1$.

Now, an algebraic corona in $[R_1, R_c]$ is considered, where $R_1 = R_i$ in case of the core-(algebraic corona) model, and $R_1 = R_s$ in case of the core-shell-(algebraic corona) model. The algebraic corona is characterized by the fact that the volume fraction of material (i.e., PMANa+d5-P4VPQBr$_\alpha$ for the core-(algebraic corona) model and PMANa for the core-shell-(algebraic corona) model) is decreasing according to

$$\varphi_{\text{alg.}}(r) = \varphi'_{\text{int.}} \left(\frac{r}{R_1}\right)^{-a}, \quad \text{for } R_1 \leq r \leq R_c. \tag{89}$$

$0 < a \leq 2$ is an exponent describing the shape of the corona function [45], and is related to the so-called 'Flory exponent': for $a = 2$ the polymer chain is in stretched conformation, for $a = 1$ the polymer is in theta conditions[28], and for $a = 0$ the chain is collapsed. The corresponding SLD profile of the corona is calculated analogously to the linear one given by Eq. (86). Thus, the excess SLD reads

$$\Delta\text{SLD}_{\text{alg.}}(r) = \varphi_{\text{alg.}}(r)\Delta\text{SLD}, \tag{90}$$

where ΔSLD is given in Eq. (87), and the amplitude form factor is

$$\begin{aligned} A_{\text{alg. corona}} &= \int_{R_1}^{R_c} 4\pi r^2 \frac{\sin(qr)}{qr} \Delta\text{SLD}_{\text{alg.}}(r)\ \mathrm{d}r \\ &= 4\pi\Delta\text{SLD}\varphi'_{\text{int.}} \int_{R_1}^{R_c} r^2 \frac{sin(qr)}{qr} \left(\frac{r}{R_1}\right)^{-a}\ \mathrm{d}r. \end{aligned} \tag{91}$$

Defining $c := 4\pi\Delta\text{SLD}\phi'_{\text{int.}}R_1^a$, then [176]

$$\begin{aligned} A_{\text{alg. corona}} = c &\int_{R_1}^{R_2} \frac{r^{1-a}\sin(qr)}{q}\ \mathrm{d}r \\ &- i\frac{cq^{a-3}}{2}\Big\{\left[(-i)^a\Gamma(2-a,-iqR_2) - i^a\Gamma(2-a,iqR_2)\right] \\ &\quad - \left[(-i)^a\Gamma(2-a,-iqR_1) - i^a\Gamma(2-a,iqR_1)\right]\Big\}, \end{aligned} \tag{92}$$

where $\Gamma(\cdot,\cdot)$ is the (upper) incomplete gamma function. The incomplete gamma function can be expressed by means of the confluent hypergeometric function of first kind (Kummer function, ${}_1F_1(\cdot)$) as well as the gamma function $\Gamma(\cdot)$ [6, Eq. (6.5.3) and Eq. (6.5.12)]

$$\Gamma(b,x) = \Gamma(b) - [b^{-1}x^b{}_1F_1(b,1+b,-x)]. \tag{93}$$

However, the incomplete gamma function available in MATLAB does not allow for complex arguments and the confluent hypergeometric function does not exist. In general the convergence of such functions is numerically not easy to accomplish. Therefore the integration in Eq. (91) has been performed numerically via a Gaussian quadrature method using a fixed number of integration support points [33, §2.7].

[28] From Wikipedia [173]: The "polymer coils act like ideal chains, assuming exactly their random walk coil dimensions. Thermodynamically, the excess chemical potential of mixing between a polymer (without mass and volume) and a theta solvent is zero."

Scattering amplitudes: Using the basic amplitude form factors given above, the overall amplitude form factors for the different models are:

1. core-shell model:
$$A_{\text{core-shell}} = A_{\text{sphere}} + A_{\text{shell}} \tag{94}$$

2. core-(linear corona) model:
$$A_{\text{core-(lin. corona)}} = A_{\text{sphere}} + A_{\text{lin. corona}} \tag{95}$$

3. core-(algebraic corona) model:
$$A_{\text{core-(alg. corona)}} = A_{\text{sphere}} + A_{\text{alg. corona}} \tag{96}$$

4. core-shell-(linear corona) model:
$$A_{\text{core-shell-(lin. corona)}} = A_{\text{sphere}} + A_{\text{shell}} + A_{\text{lin. corona}} \tag{97}$$

5. core-shell-(algebraic corona) model:
$$A_{\text{core-shell-(alg. corona)}} = A_{\text{sphere}} + A_{\text{shell}} + A_{\text{alg. corona}} \tag{98}$$

If the system is monodisperse and if there are no interactions between the particles, the normalized scattering intensity reads

$$I^n_{\text{mono}}(q) = nA^2(q), \tag{99}$$

where n is the particle number density and $A(q)$ is one of the above-mentioned amplitude form factors.

For a polydisperse[29] system, where the dispersity is taken over the aggregation number $N_{\text{agg.}}$ of PIB-PMANa, the normalized intensity reads as follows

$$I^n_{\text{poly}}(q) = n\big\langle A^2(q; N_{\text{agg.}})\big\rangle_{N_{\text{agg.}}}, \tag{100}$$

where the bracket indicates an average over the aggregation number. Here, it is assumed that the aggregation number can be well described by a probability density function pdf$(\cdot)$. Usually, the effect of different uni-modal distributions is negligible in SAS, and a log-normal probability density distribution was taken:

$$\text{pdf}(N_{\text{agg.}}) = \text{lognpdf}(N_{\text{agg.}}, u, v) = \frac{1}{N_{\text{agg.}} v\sqrt{2\pi}} \exp\left(-\frac{\ln^2(N_{\text{agg.}}/u)}{2v^2}\right). \tag{101}$$

u and v are parameters that are related to the mean $\overline{N}_{\text{agg.}}$ and the standard deviation std$(N_{\text{agg.}})$ of the distribution by

$$u = \frac{\overline{N}^2_{\text{agg.}}}{\sqrt{(\text{std}(N_{\text{agg.}}))^2 + \overline{N}^2_{\text{agg.}}}}, \tag{102}$$

[29] Here, and in the following sections the notion 'polydispersity' means the quantity $p = \sigma/\mu$, where σ and μ are the standard deviation and mean of a considered distribution. The quantity p is also called 'polydispersity parameter' [62] and 'coefficient of variation' [4].

and

$$v = \sqrt{\ln\left(\frac{(\mathrm{std}(N_{\mathrm{agg.}}))^2}{\overline{N}_{\mathrm{agg.}}^2} + 1\right)}. \tag{103}$$

Particle interactions have been taken into account by an effective hard sphere potential, such that the considered model intensities read

$$I^n_{\mathrm{mono,HS}}(q) = I^n_{\mathrm{mono}}(q) S_1(q), \qquad \text{and} \tag{104}$$

$$I^n_{\mathrm{poly,HS}}(q) = I^n_{\mathrm{poly}}(q) S_1(q), \tag{105}$$

where here $S_1(q)$ is the structure factor in the approximation for monodisperse hard spheres with the Percus-Yevick closure relation [154, 170, 85] (cf. appendix D.1 Eq. (421)). The formula for the monodisperse case, Eq. (104), is exact within the Percus-Yevick closure approximation [124]. However, the structure factor for the polydisperse case, Eq. (105), only describes an approximation to the effective structure factor within a complete polydisperse Percus-Yevick closure approximation, i.e., of $S^{\mathrm{eff.}}_{\mathrm{HS,compl.}}(q)$ in

$$\begin{aligned} I^n_{\mathrm{poly,HS,compl.}}(q) = n &\int_0^\infty A^2(q; N_{\mathrm{agg.}}) \mathrm{pdf}(N_{\mathrm{agg.}})\, \mathrm{d}N_{\mathrm{agg.}} \\ + n^2 &\int_0^\infty \int_0^\infty A(q; N'_{\mathrm{agg.}}) A(q; N''_{\mathrm{agg.}}) H\big(q; \sigma(N'_{\mathrm{agg.}}), \sigma(N''_{\mathrm{agg.}})\big) \\ &\times \mathrm{pdf}(N'_{\mathrm{agg.}}) \mathrm{pdf}(N''_{\mathrm{agg.}})\, \mathrm{d}N'_{\mathrm{agg.}}\, \mathrm{d}N''_{\mathrm{agg.}} \\ = n &\big\langle A^2(q; N_{\mathrm{agg.}}) \big\rangle_{N_{\mathrm{agg.}}} S^{\mathrm{eff.}}_{\mathrm{HS,compl.}}(q), \end{aligned} \tag{106}$$

where $H\big(q; \sigma(N'_{\mathrm{agg.}}), \sigma(N''_{\mathrm{agg.}})\big)$ is the partial structure function for $N'_{\mathrm{agg.}}$ and $N''_{\mathrm{agg.}}$ (cf. appendix D.2), and $\sigma(N_{\mathrm{agg.}})$ is a function[30] that returns the hard sphere diameter for $N_{\mathrm{agg.}}$, and

$$\begin{aligned} S^{\mathrm{eff.}}_{\mathrm{HS,compl.}}(q) = 1 + \Bigg\{ n^2 &\int_0^\infty \int_0^\infty A(q; N'_{\mathrm{agg.}}) A(q; N''_{\mathrm{agg.}}) H\big(q; \sigma(N'_{\mathrm{agg.}}), \sigma(N''_{\mathrm{agg.}})\big) \\ &\times \mathrm{pdf}(N'_{\mathrm{agg.}}) \mathrm{pdf}(N''_{\mathrm{agg.}})\, \mathrm{d}N'_{\mathrm{agg.}}\, \mathrm{d}N'_{\mathrm{agg.}} \Bigg\} \\ \Bigg/ &\int_0^\infty A^2(q; N_{\mathrm{agg.}}) \mathrm{pdf}(N_{\mathrm{agg.}})\, \mathrm{d}N_{\mathrm{agg.}}. \end{aligned} \tag{107}$$

Here, Eq. (106) was employed, and the partial structure factor functions are taken from[31] Blum and Stell [17] and its errata paper [18] and further corrections given by [56] Still, [56] contains typos and are corrected in Eq. (106 and the partial structure factor functions given in the appendix D.2.

Following quantities are fitted:

- the density of the complex ρ(complex),

[30] In the implemented core-shell model function the core radius $R_i(N_{\mathrm{agg.}})$ and the shell thickness $t_s(N_{\mathrm{agg.}})$ are determined for a given aggregation number $N_{\mathrm{agg.}}$. Then, the hard sphere diameter is calculated as $\sigma(N_{\mathrm{agg.}})/2 = R_i(N_{\mathrm{agg.}}) + x_r t_s(N_{\mathrm{agg.}})$, where $x_r \geq 0$ is an adjustable parameter allowing to vary the hard sphere diameter. An analogous calculation is done for the core-shell-(linear corona) model. The determined mean hard sphere radii $\overline{r}_{HS}$ are listed in the tables 2 - 4.

[31] Vrij also derived a formula independently [167, 168].

- the aggregation number $N_{\text{agg.}}$ for the monodisperse case, respectively the mean aggregation number $\overline{N}_{\text{agg.}}$ and the polydispersity $p(N_{\text{agg.}})$ for the polydisperse case,
- the hydration fraction within the shell φ_h,
- the shape parameter of the algebraic corona a, and
- an effective hard sphere structure radius r_{HS}, respectively the parameter x_r (see footnote 30 on page 43) in the polydisperse structure factor approach (from which an effective mean hard sphere radius $\overline{r}_{\text{HS}}$ is calculated).

Note that the hydration and thickness of the corona are implicitly determined by the constraints given by the definition of the linear/algebraic corona and the volume of the complex or of PMANa — see appendix C.2. The material equations that are used within the amplitude form factors are given in appendix C.2.

3.3.2. Evaluations

Before the fits are discussed in some detail below, it should be mentioned that many of the measured intensity curves have an increase of the intensity in the low q-regime, see Fig. 13-17. In the literature such an upturn is already observed for pure polyelectrolytes in solution [101, 38, 30] as a signature of fluctuations of SLD over very large distances. The reason for it is still being under discussion, and possibly related to same-charge attractivity [100]. In the systems presented here, the upturn of the intensity in the low q-regime might also arise from the existence of some clusters of micelles linked together via d5-P4VPQBr$_\alpha$ (creating necklace structures).

For all fits, the optimization algorithm converged and the Hessian matrix of the misfit error (Hessian matrix of $\mathcal{E}_s$ in Eq. (68)) was positive definite (i.e., a minimum has been found by the algorithm). For all models $\mathcal{E}_s$ was determined in the optimum, see tables 2-4. However, because of the underfitting it may be more instructive to look at the obtained fits and to decide which one is good — often called 'chi-by-eye' [129, §15.0] — since the misfit error does not contain the information from where the mismatch is coming from. The uncertainties of all determined parameters are in the per-mill regime or less.

Set 1: Fig. 13 - 15 show the data set 1 together with the different model fits, and the fitted parameters are listed in table 2. Fig. 13 shows the monodisperse fits of the different models. The model fits are quite good for all models except for the core-(linear corona) model, and the core-(algebraic corona) model. These two models cannot capture the features of the observed intensity curves, hence also the model parameters are not listed in table 2. The other three models (core-shell, core-shell-(linear corona), and core-shell-(algebraic corona)) fit rather well, except from a relatively slight mismatch in the low q-regime. Therefore, polydisperse fits have been considered for them, and are shown in Fig. 14. These polydisperse models improve the fits in the mid and high q-regime (the original oscillatory curve behavior is suppressed), but not in the low q-regime. The core-shell-(linear corona) model and the core-shell-(algebraic corona) model fit equally well the data, hence only the core-shell-(linear corona) model is considered in the following. Fig. 15 shows fits of the core-shell model and of the core-shell-(linear corona) model, where a monodisperse hard sphere structure factor is applied, as well as a polydisperse one. There is nearly no difference between the monodisperse structure factor model and the polydisperse one. Best fits are obtained when a model contains a shell, polydispersity, and a structure factor. The inclusion of a corona (one more fitted parameter) enhances the fit ($\mathcal{E}_s$ decreases by a factor 3).

Set 2: Fig. 16 shows the data set 2 together with different model fits, and the fitted parameters are listed in table 3. The core-shell model as well as the core-shell-(linear corona) model capture the main features of the scattering intensity in the mid and high q-regime nearly equally well, but there are similar mismatches in the low q-regime. The other basic models (core-(linear corona), and core-(algebraic corona)) are not represented, since no good model fits could be achieved for them. The polydisperse model version of the core-shell model and the core-shell-(linear corona) model improve the fits in the low and mid q-regime (the oscillatory behavior of the curves is suppressed), but the low q-regime is still not well fitted. Using an additional structure factor (monodisperse or polydisperse) for the considered models does not really improve the fits. Overall, adding a corona does not lead to any improvement, visually as well as for $\mathcal{E}_s$. Hence, all models are equally good as long as a polydispersity is included, but no model can capture all features.

Set 3: Fig. 17 shows the data set 3 together with the different model fits, and the fitted parameters are listed in table 4. Here, the core-shell model as well as the core-shell-(linear corona) model can both capture the main features of the intensity curves in the mid and high q-regime. The other basic models (core-(linear corona), and core-(algebraic corona)) are not represented, since no good model fits could be achieved for them. Anyway, in the mid q-regime the core-shell-(linear corona) model slightly fits the intensity curves better than the core-shell model. The same fact holds for the polydisperse variants of these two models. The point seems to be the following one: Monodisperse models exhibit too strong features; then, adding a polydispersity but no corona leads to too smooth models that cannot simultaneously reproduce the smoothness of the curves at high-to-mid q and the minimum in intensity observed at $q \approx 0.2\,\mathrm{nm}^{-1}$ for the curve with a D_2O:H_2O ratio of 40:60. The presence of the corona allows to get smoother curves with less polydispersity (ca. 0.8 instead of ca. 1.3), allowing to fit the minimum. Taking into account a hard sphere structure factor (monodisperse or polydisperse) only very slightly improves the fits in the low q-regime, but there is nearly no difference between the monodisperse structure factor and the polydisperse one. In this case the inclusion of a corona does improve the fits visually and it leads to a small reduction of $\mathcal{E}_s$, but overall the reproduction of experimental features is not properly achieved.

In the following, essential outcomes of the fits are discussed:

The aggregation number: The comparison between sets 1 and 3 shows that reducing the hydrophobic block from 75 to 30 units, with constant hydrophilic block (190 units), results in a decrease of the aggregation number from ca. 300 to ca. 100. The comparison between set 1 and 2 shows that increasing the hydrophilic block from 190 to 1590 units, with constant hydrophobic block (75 units), results in a decrease of the aggregation number from ca. 300 to ca. 20.

The density of the complex is almost constant for a given Z: Whenever fits are good, the fitted value of the density of the complex at Z=1 is systematically close to $2\,\mathrm{g{\cdot}cm}^{-3}$, and the fitted value of the density of the complex at Z=0.4 is systematically close to $2.2\,\mathrm{g{\cdot}cm}^{-3}$. Given that PMANa has a density of $2.4\,\mathrm{g{\cdot}cm}^{-3}$, and that d5-P4VPQBr$_\alpha$ has a density of $1.5\,\mathrm{g{\cdot}cm}^{-3}$, these values are perfectly reasonable.

The shell thickness is slightly larger for set 1 (ca. 12 nm) than for set 3 (ca. 9 nm), while the hydrophilic block is identical, in agreement with the smaller aggregation number (same

volume of complex, similar hydration around 80-85 %, but the core/shell volume ratios are larger for smaller micelles). The value for set 2 (ca. 15 nm), with a much larger hydrophilic block length (1590 units) is barely twice the value for set 1, but this can again be explained by the smaller core and aggregation number.

The shell hydrations for set 2 are higher (ca. 95 %) than for sets 1 and 3 (ca. 80 %). Since, data set 2 does not contain much information/features, fits have also been performed with the hydration fraction fixed at 80 %, but these fit are not good at all. Consequently, it can be stated that there is the correlation: the longer the polyelectrolyte chain is the less compact the shell is.

The number density n can be used to estimate the average distance d between centers of micelles (assuming a simple cubic packing, it follows $d \approx n^{-1/3}$). In all cases, this average distance is comparable to the diameter of micelles with a core-shell-corona structure. For models with a corona used for fitting sets 2 and 3, the interparticle distance is actually less than the size of the objects, which means that the overlapping concentration has been reached or possibly exceeded. This overlap is not taken into account by the applied models (both in terms of SLD profile and of structure factor), i.e., these models are partially inconsistent.

The aggregation number and the shell thickness are controlled by the hydrophobic and hydrophilic blocks in a way that is qualitatively explained by the packing parameter [75] concept: a surfactant with larger head-group and/or smaller tail leads to smaller aggregates with larger shell thickness.

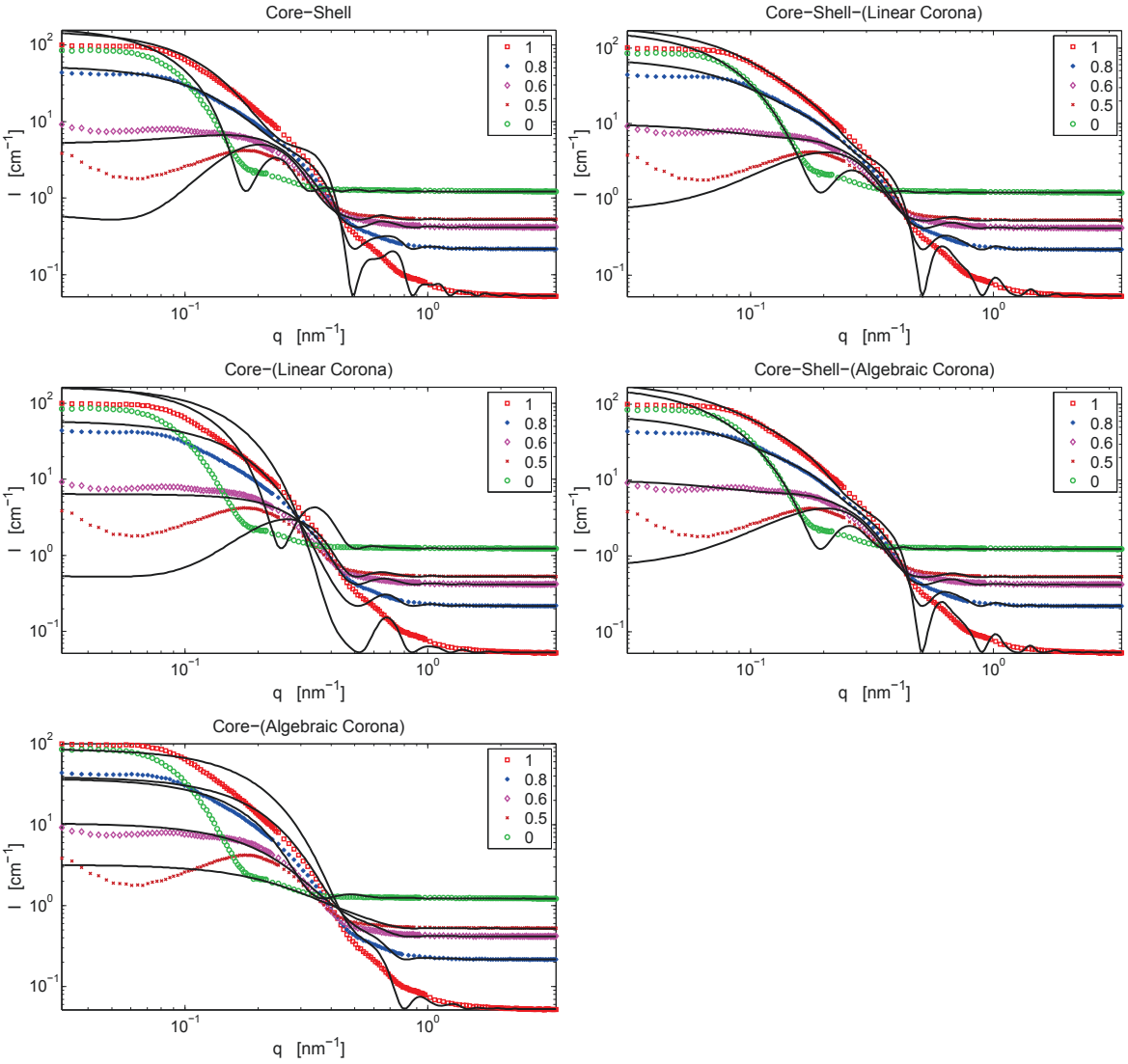


Figure 13: Monodisperse model fits achieved for set 1 (IB_{75}−$MANa_{190}$). No structure factor taken into account. Legend entries show the percentage of D_2O.

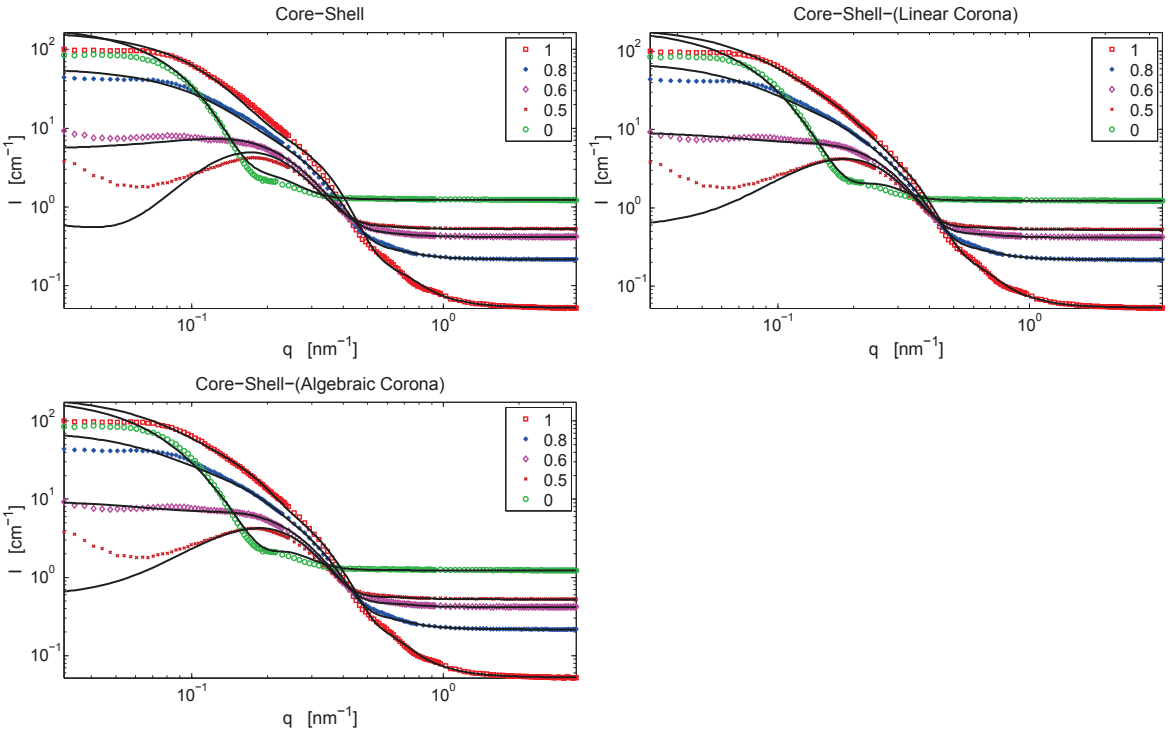


Figure 14: Polydisperse model fits achieved for set 1 ($IB_{75}-MANa_{190}$). No structure factor taken into account. Legend entries show the percentage of D_2O.

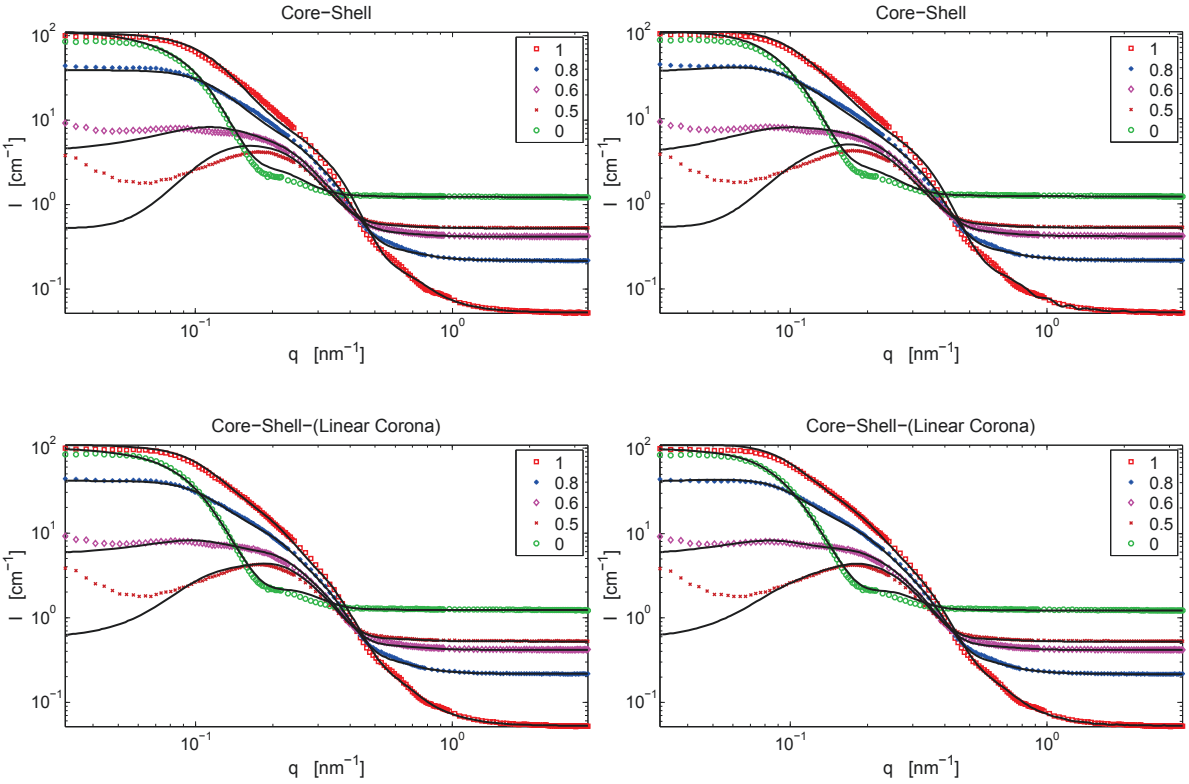


Figure 15: Polydisperse model fits achieved for set 1 ($IB_{75}-MANa_{190}$). Structure factor taken into account. Left side: monodisperse hard sphere structure factor. Right side: polydisperse hard sphere structure factor. Legend entries show the percentage of D_2O.

	CS	CSP	CSLC	CSAC	CSLCP	CSACP	CSPHS	CSPHSP	CSLCPHS	CSLCPHSP
ρ(complex) / (g·cm^{-3})	2.28	2.29	1.95	1.95	2.02	2.02	2.24	2.26	2	2
$N_{\text{agg.}}, \overline{N}_{\text{agg.}}$	374	279	362	367	297	303	289	285	308	303
$p(N_{\text{agg.}})$		0.72			0.62	0.61	0.72	0.73	0.66	0.66
φ_h	0.82	0.84	0.81	0.81	0.81	0.81	0.84	0.84	0.8	0.8
a				2		2				
$r_{\text{HS}}, \overline{r}_{\text{HS}}$ / nm							25.1	25.7	27.9	28.3
$R_i, \overline{R}_i$ / nm	8.79	7.61	8.7	8.74	7.85	7.91	7.7	7.7	7.91	7.87
$t_s, \bar{t}_s$ / nm	15.4	14.3	13.2	13.2	11.5	11.7	14.3	14.3	11.4	11.4
$t_c, \bar{t}_c$ / nm			27.7	29.3	26.2	29.3			26.2	26.2
n / (10^{-7}nm^{-3})	7.97	10.6	8.16	8.06	9.96	9.77	10.3	10.4	9.61	9.77
d / nm	108	98.1	107	107	100	101	99	98.7	101	101
$\mathcal{E}_s$ /(10^5)	7.71	2.15	4.14	4.04	1.36	1.3	1.57	1.55	0.49	0.5

Table 2: Model parameters obtained for set 1 ($\text{IB}_{75}-\text{MANa}_{190}$). Model abbreviations: CS = core-shell; CSP = CS polydisperse; CSPHS = CSP with monodisperse hard sphere structure factor, CSPHSP = CSP with polydisperse hard sphere structure factor; CSLC = core-shell-(linear corona); CSAC = core-shell-(algebraic corona); CSLCP = CSLC polydisperse, CSACP = CSAC polydisperse; CSLCPHS = CSLCP with monodisperse hard sphere structure factor, CSLCPHSP = CSLCP with polydisperse hard sphere structure factor. ρ(complex): apparent density of the complex. $N_{\text{agg.}}$: aggregation number. $p(N_{\text{agg.}})$: polydispersity of the aggregation number. φ_h: hydration fraction (within the shell). a: shape parameter of the algebraic corona. r_{HS}: hard sphere radius. R_i: core radius. t_s: shell thickness. t_c: corona thickness. n: number density. d: average particle distance (assuming a simple cubic packing). $\mathcal{E}_s$: misfit error as defined in Eq. (68). A bar over one of these symbols means average value. The first column block contains fitted parameters. The bottom column block contains parameters determined from the fit.

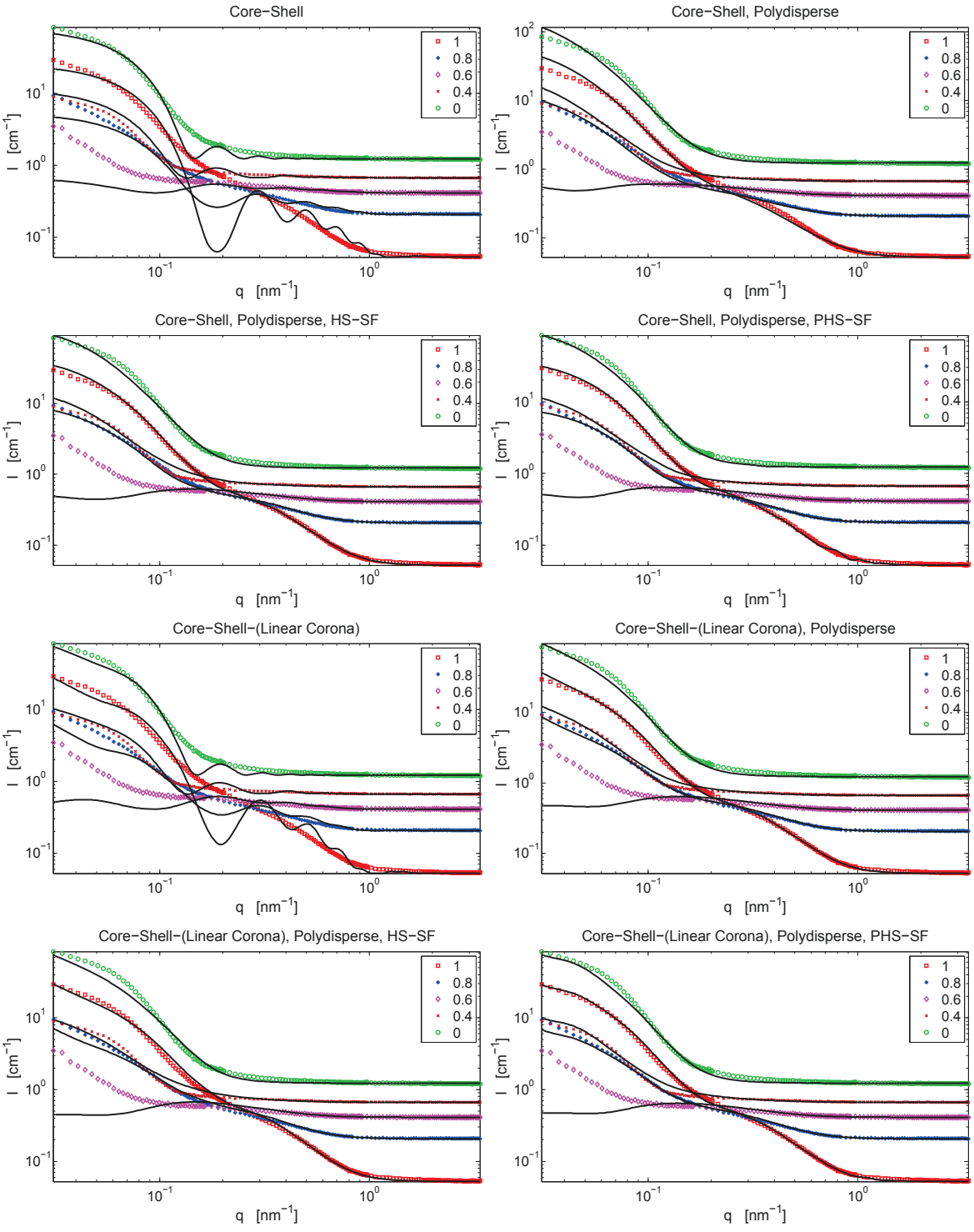


Figure 16: Model fits achieved for set 2 (IB_{75}−$MANa_{1590}$). HS-SF: hard sphere structure factor. PHS-SF: polydisperse hard sphere structure factor. Legend entries show the percentage of D_2O.

	CS	CSP	CSLC	CSLCP	CSPHS	CSPHSP	CSLCPHS	CSLCPHSP
ρ(complex) / (g·cm^{-3})	2.4	2.34	2.11	2.06	2.09	2.35	2.05	2.07
$N_{\text{agg.}}, \overline{N}_{\text{agg.}}$	22.1	6.89	41.2	17	14.4	12.9	19.6	17
$p(N_{\text{agg.}})$		3.25		1.95	2.09	2.49	2	1.98
φ_h	0.96	0.95	0.95	0.93	0.95	0.95	0.93	0.94
$r_{\text{HS}}, \overline{r}_{\text{HS}}$ / nm					16.7	15.9	17.6	78.6
$R_i, \overline{R}_i$ / nm	3.48	1.76	4.21	2.63	2.46	2.26	2.75	2.59
$t_s, \bar{t}_s$ / nm	27.3	14.8	25.5	15	19.6	16	14.8	14.7
$t_c, \bar{t}_c$ / nm			103	75.2			75	74.3
n / (10^{-6}nm^{-3})	1.95	6.4	1.05	2.57	3.05	3.6	2.24	2.62
d / nm	80	53.9	98.4	73	69	65.2	76.4	72.5
$\mathcal{E}_s$ /(10^4)	36.5	6.39	28.1	4.64	4.66	4.28	7.68	4.46

Table 3: Model parameters obtained for set 2 (IB$_{75}$−MANa$_{1590}$). Model abbreviations: CS = core-shell; CSP = CS polydisperse; CSPHS = CSP with monodisperse hard sphere structure factor, CSPHSP = CSP with polydisperse hard sphere structure factor; CSLC = core-shell-(linear corona); CSLCP = CSLC polydisperse, CSLCPHS = CSLCP with monodisperse hard sphere structure factor, CSLCPHSP = CSLCP with polydisperse hard sphere structure factor. ρ(complex): apparent density of the complex. $N_{\text{agg.}}$: aggregation number. $p(N_{\text{agg.}})$: polydispersity of the aggregation number. φ_h: hydration fraction (within the shell). a: shape parameter of the algebraic corona. r_{HS}: hard sphere radius. R_i: core radius. t_s: shell thickness. t_c: corona thickness. n: number density. d: average particle distance (assuming a simple cubic packing). $\mathcal{E}_s$: misfit error as defined in Eq. (68). A bar over one of these symbols means average value. The first column block contains fitted parameters. The bottom column block contains parameters determined from the fit.

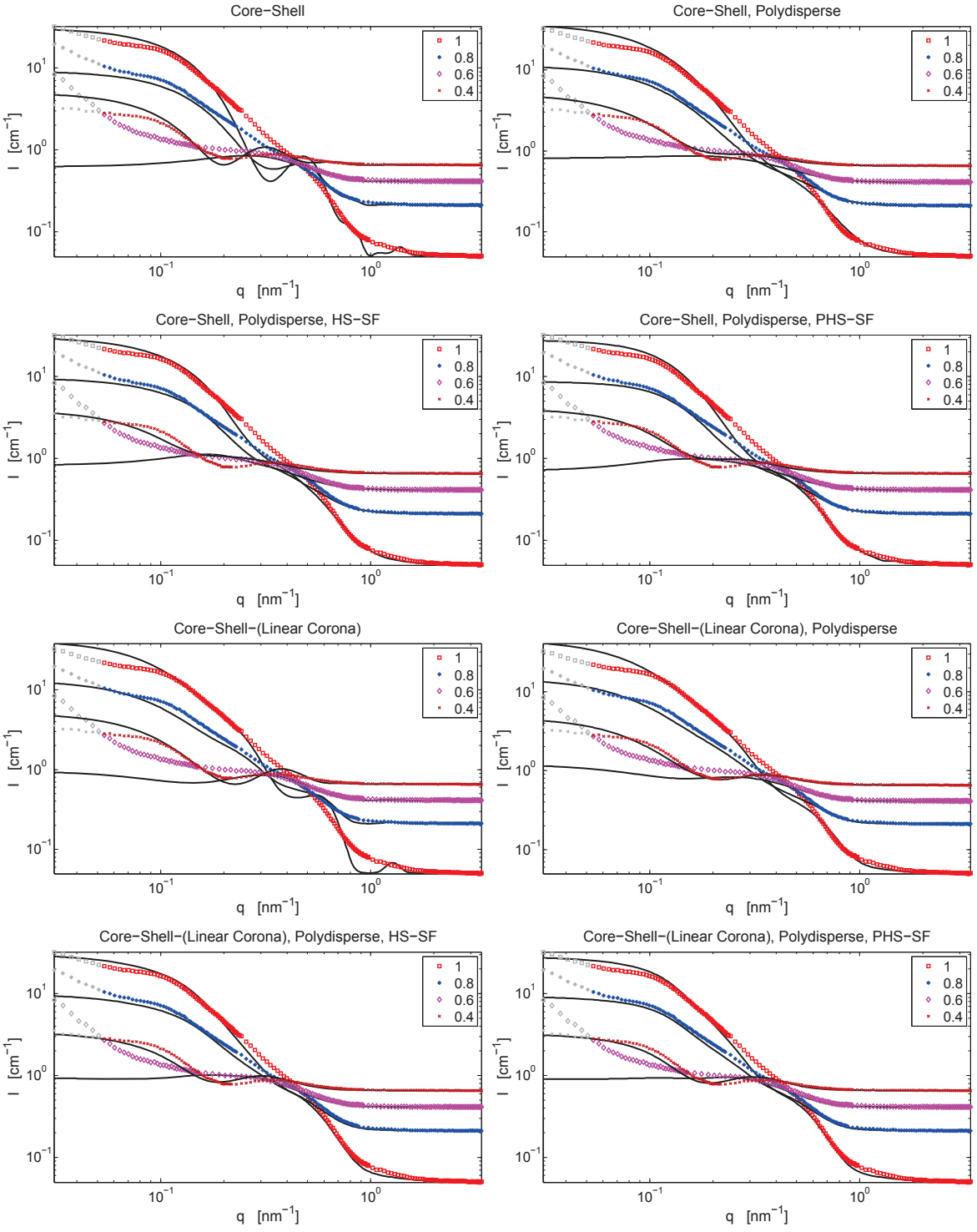


Figure 17: Model fits achieved for set 3 (IB_{30}−$MANa_{190}$). HS-SF: hard sphere structure factor. PHS-SF: polydisperse hard sphere structure factor. The gray curve parts are not used in the fits, since they cannot be explained by the considered models. Legend entries show the percentage of D_2O.

	CS	CSP	CSLC	CSLCP	CSPHS	CSPHSP	CSLCPHS	CSLCPHSP
ρ(complex) / (g·cm^{-3})	2.22	2.17	2	1.94	2.14	2.18	1.94	1.95
$N_{\text{agg.}}, \overline{N}_{\text{agg.}}$	115	48.6	135	82.4	68.1	68.6	113.2	107
$p(N_{\text{agg.}})$		1.29		0.81	1.37	1.17	0.83	0.8
φ_h	0.87	0.83	0.8	0.79	0.85	0.85	0.82	0.81
$r_{\text{HS}}, \overline{r}_{\text{HS}}$ / nm					15.5	13.8	19.7	19.4
$R_i, \overline{R}_i$ / nm	4.38	2.94	4.62	3.7	3.26	3.36	4.1	4.04
$t_s, \bar{t}_s$ / nm	13.5	11.1	9.9	7.82	9.47	9.71	9.23	9.01
$t_c, \bar{t}_c$ / nm			24.4	21.8			23.2	23
n / (10^{-6}nm^{-3})	2.92	6.96	2.49	4.1	4.98	4.88	2.98	4.88
d / nm	70	52.4	73.8	62.5	58.6	59	69.5	59
$\mathcal{E}_s$ /(10^5)	5.18	2.82	3.6	2.33	2.02	2.24	1.54	1.61

Table 4: Model parameters obtained for set 3 (IB_{30}−MANa_{190}). Model abbreviations: CS = core-shell; CSP = CS polydisperse; CSPHS = CSP with monodisperse hard sphere structure factor, CSPHSP = CSP with polydisperse hard sphere structure factor; CSLC = core-shell-(linear corona); CSLCP = CSLC polydisperse, CSLCPHS = CSLCP with monodisperse hard sphere structure factor, CSLCPHSP = CSLCP with polydisperse hard sphere structure factor. ρ(complex): apparent density of the complex. $N_{\text{agg.}}$: aggregation number. $p(N_{\text{agg.}})$: polydispersity of the aggregation number. φ_h: hydration fraction (within the shell). a: shape parameter of the algebraic corona. r_{HS}: hard sphere radius. R_i: core radius. t_s: shell thickness. t_c: corona thickness. n: number density. d: average particle distance (assuming a simple cubic packing). $\mathcal{E}_s$: misfit error as defined in Eq. (68). A bar over one of these symbols means average value. The first column block contains fitted parameters. The bottom column block contains parameters determined from the fit.

3.3.3. Results

For all IPEC data sets, the core-shell model, the core-shell-(linear corona) model, and the variants of these two models (polydisperse models, and with monodisperse or polydisperse structure factor) fit well the observed intensities in absolute scale, except in the very low q-regime, and in particular for data set 3. The core-corona models did not fit the data at all. Hence the first result is that a shell with a constant SLD is needed to reproduce the experimental data. The polydisperse models are essentially necessary to explain the smoothness of the curves. The inclusion of a structure factor slightly improves the fits in the low q-regime. The fitted parameters for different models for a given data set typically end up in a narrow range. In particular, the aggregation numbers, densities of the complex, shell hydrations and corona thicknesses are quite similar.

The compactness of the shell depends on the length of the PMANa block. For a short PMANa block of 190 units, the shell hydration is ca. 80-85 %; for the long PMANa block (1590 units) the hydration increases at 95 %. In both cases, the amount of solvent is very large, which is interesting given that the complex made of homopolymers is not water-soluble, and one could expect such a neutral complex to be rather hydrophobic.

The question of whether the IPEC structures are core-shell objects or core-shell-corona objects cannot fully be answered, since both models capture nearly equally well the main features of the scattering intensities. The main problem is that the measured intensity curves are quite smooth as a consequence of the polydispersity of the objects, meaning that distinct features of a core-shell or of a core-shell-corona structure are smeared out, and hence a reliable determination of such structures is not possible. Therefore, also the shape of a possible corona cannot be inferred. However, data set 1 and data set 3 contain more distinct features than data set 2, and at the same time for set 1 and 3 the core-shell-corona models fit slightly better than the core-shell models. This might give an indication that the built structures, at least for set 1 and 3, contain a corona. Data set 2 does not have enough features to make any selection at all.

This evaluation has also shown the importance of physical a priori knowledge in order to perform model selection. In the fits, the dominating error results from a systematic error, i.e., underfitting exists. The obtained fitting errors can be used to compare models with each other, but they do not contain information of where a mismatch is resulting from. The obtained parameter uncertainties are all in the per-mill regime or less. However, these parameters and their uncertainties are meaningless if the assumed model is wrong as already stressed in the last subsection.

3.4. Conclusion

In this section the case of physical model fitting of SAS data has been discussed.

The bottleneck of evaluating comprehensive 1- and 2-dimensional scattering data sets/series is now overcome and efficiently possible with the newly developed features of SASET. Outstanding is the feature to easily perform simultaneous model fitting (of data sets as well as of coupled series), hence allowing to conveniently increase the information content of an analysis and therefore to determine more reliably models.

The frequentist and Bayesian statistical inference discussions have shown how model selection can be performed in the two different inference approaches, where limitations occur in the inference procedures, how uncertainties of model parameters are determined, and it is also shown that physical a priori knowledge is indispensable in order to perform model selection.

Last but not least a simultaneous fitting example of contrast variation data of a PIB-PMAA/P4VPQ interpolyelectrolyte complex system has been presented. The model fits were quickly obtained via SASET (a few seconds (for monodisperse models) up to a few minutes (for polydisperse models with polydisperse structure factor) on a PC with an Intel® Core^{TM} i7 CPU 860 at 2.8 GHz (4 cores, 8 threads) and using parallel computing capacities of SASET). The model fits have shown that spherical core-shell and core-shell-corona structures are possible ones.

4. Complex Colloidal Systems

This chapter analyzes the scattering of complex cluster systems, which consist of vesicles that are decorated by particles and/or may contain internalized particles, cf. Fig. 18, 19, 20. Such hierarchical morphologies cannot be modeled by analytical formulas. Actual structures must be generated by physically consistent random placement of the different components, which can be achieved by means of Monte Carlo simulations, but they usually need a lot of computational power.

Here, the vesicles consist of a soft bilayer membrane (phospholipids) and the (solid) particles are silica nanoparticles. Such systems have attracted growing attention in the last years, since vesicles (closed bilayer membranes) can be considered as model for cells (cell membranes) and the particles as interacting test specimen, hence these systems are very important for studies in nanomedicine, and nanotoxicology [105], as there is an increasing number of applications of nanoparticles. In this work, the scattering of such systems is simulated in order to identify model parameters that can reliably be determined from scattering experiments.

Two systems are considered here — detailed information have been friendly provided by Dr. Raphaël Michel: in one case DPPC (1,2-dipalmitoyl-sn-glycero-3-phosphocholine) builds closed phospholipid bilayer with membranes in a fluid state; in the other case DOPC (1,2-dioleoyl-sn-glycero-3-phosphocholine) builds membranes in a gel state[32]. The matrix is H_2O and the silica (nano-)particles consist of Ludox® HS 40. The particles are assumed to be spherical, having a mean radius of $R_p = 8.36\,\mathrm{nm}$ and a polydispersity of $p_p = 0.14$. The vesicles (closed phospholipid bilayer membranes) are considered to be spherical, too, and having different size distributions for the different experimental setups (cf. text below), which were inferred from earlier experiments, namely cryo-TEM (cryogenic transmission electron microscopy), light scattering, and SANS [104]. For instance, Fig. 18 and 19 show cryo-TEM pictures of decorated vesicles and of vesicles with internalized particles. Moreover, from these earlier experiments it was deduced for the DOPC vesicle systems that the particles are first on the outside of the vesicle shell, and gradually become incorporated into the vesicle interior. Furthermore, it was inferred that if a particle goes into a vesicle, it removes a part of the vesicle membrane in order to build its own membrane shell encapsulation — this shell is called 'supported lipid bilayer'. Fig. 20 depicts a schematic drawing of a decorated vesicle with internalized particles that have an additional membrane encapsulation.

However, the performed cryo-TEM micrographs may contain artifacts, because of the involved freezing procedure, and additionally this method can only show a few localized, static objects. Complementary to this microscopic method are scattering methods, which are in situ methods that allow to infer structural information from an intensity average of the scattering objects. However, the performed SANS experiments could not be used to deduce detailed structural information about the particles, because of the poor contrast conditions (SANS is mostly sensitive to the hydrogenated phospholipidic vesicles).

In this work, SAXS experiments are simulated in order to comprehend how well structural information can be inferred if such an experiment is performed. For SAXS, the contrast conditions stress the focus more on the particles and their interactions in comparison with SANS (which has been used in earlier evaluations [104]), and therefore SAXS gives a more detailed overall picture of the particle structure within the considered systems.

The aim is to compare experimental SAXS data with the simulations, to assess our full understanding of the system. However, for technical reasons, the SAXS measurements could not be performed within the time of this thesis.

[32]The transition temperature of DPPC is 41 °C and the one of DOPC is −18 °C.

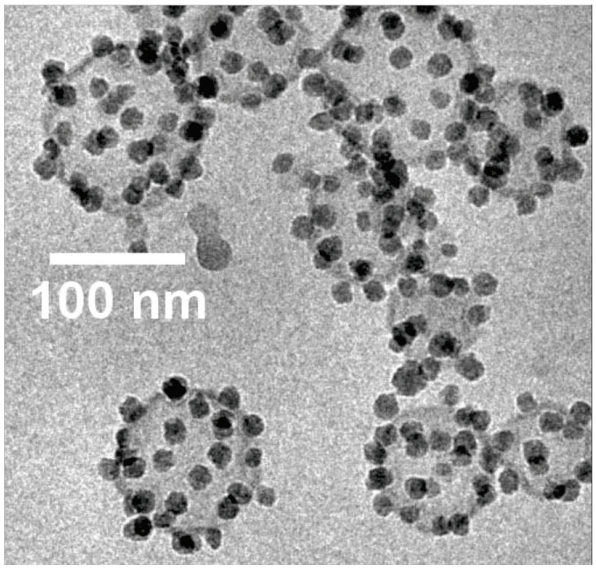


Figure 18: Cryo-TEM image featuring decorated DPPC vesicles in the mixed system $DP_{1.0}$. Figure adapted and reprinted with permission from [104, 107].

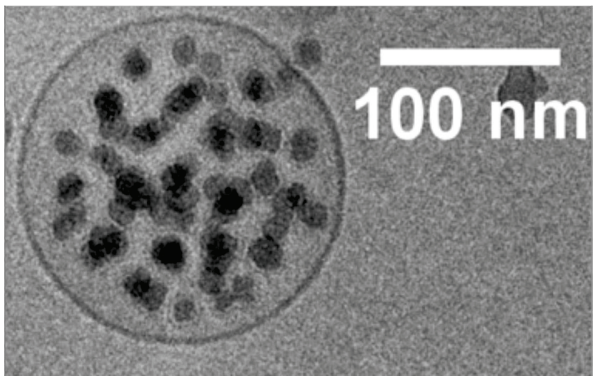


Figure 19: Cryo-TEM image featuring a DOPC vesicle with internalized (nano-)particles in the mixed system $DO_{0.5}$. Figure adapted and reprinted with permission from [104, 106].

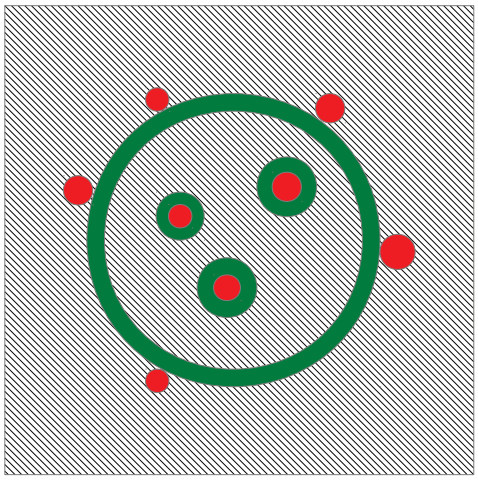


Figure 20: Schematic drawing of a decorated vesicle with internalized particles (i.e., a complex cluster). The particles within the vesicle are surrounded by a shell. The hatching indicates the matrix.

The main program to simulate such complex clusters consists of four parts as described in algorithm 2. The two main sub procedures are the creation of the complex clusters and the calculation of the appropriate intensity. The creation of the complex clusters is described in section 4.2, and is based on Monte Carlo (MC) simulations, since a full model description is analytically as well as by means of deterministic numerical methods not tractable. An introduction into Monte Carlo (MC) simulations is given in the next section. In section 4.3 the scattering formulas for the complex clusters are derived as well as the formulas for a few simpler analytical models. In section 4.4 the results are presented.

Algorithm 2 The complex cluster algorithm.

1: initialize algorithm parameters
2: create complex cluster size parameters and appropriate configuration of particles (via algorithm 4)
3: calculate the appropriate scattering intensity
4: plot the scattering intensity

4.1. Introduction to Monte Carlo Simulations

In the following a summary of the theory of MC simulations is presented, which mainly consists of the contents in [16, 129, 116].

Monte Carlo (MC) methods, in earlier times also known as statistical sampling methods, are used to approximately solve problems that are analytically not solvable as well as by means of deterministic numerical integration methods not tractable. A typical example is the evaluation of multidimensional integrals, e.g., in order to determine the expectation value of a quantity $Q(\mathbf{x})$ with respect to the probability density function $\mathrm{pdf}(\mathbf{x})$ [16, §11]:

$$\langle Q \rangle = \int Q(\mathbf{x}) \, \mathrm{pdf}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{108}$$

cf. Fig. 21. Another example [129, §15.8] is the sampling from a probability density function itself, e.g., in order to infer the a posteriori distribution in a Bayesian approach.

In this work expectations of the form given in Eq. (108) are considered. If the sampling procedure creates N independent samples $\{\mathbf{x}_n\}_{n=1}^{N}$ from the distribution pdf($\mathbf{x}$), then the expectation of $Q(\mathbf{x})$ is approximated by the mean

$$\overline{Q} = \frac{1}{N}\sum_{n=1}^{N} Q(\mathbf{x}_n) \tag{109}$$

and its standard deviation is [16, §11]

$$\mathrm{std}[\overline{Q}] = \sqrt{\frac{1}{N}\Big\langle \big(Q - \langle Q\rangle\big)^2 \Big\rangle}, \tag{110}$$

thus showing that the standard deviation of the estimator $\overline{Q}$ (i) decreases proportionally to $1/\sqrt{N}$ (variance reduction), and (ii) does not depend on the dimensionality of $\mathbf{x}$. Therefore, often only a few samples (≈10-20) are required in order to estimate the expectation of $Q(\mathbf{x})$ with sufficient accuracy. However, the sample size may have to be larger in cases where (i) the samples are not drawn independently from pdf($\mathbf{x}$) or (ii) where the function $|Q(\mathbf{x})|$ is low in regions where pdf($\mathbf{x}$) is large, and vice versa [16, §11].

In order to exemplify the need for a non-deterministic numerical integration method, one may consider the following: let $\mathbf{x}$ have 80 dimensions, e.g., in order to describe the positions of 40 particles on a vesicle (via an azimuthal angle and a polar one for each particle), and let pdf($\mathbf{x}$) a non-factorizable distribution, such that the nested integration cannot be written as a product of single integrations. If the deterministic integration procedure shall only evaluate the function at 10 different points per dimension (i.e., the integration is probably quite rough) the nested integration would yield 10^{80} function evaluations — a number that equals the estimated number of nucleons in the whole universe! However, the Monte Carlo approach suggests that a good estimation may be achievable by only drawing a few samples from the underlying probability density function.

Nowadays, there are pseudorandom number generators available for the uniform probability distribution and the normal distribution in almost all computer languages. Pseudorandom means that a number is not really randomly generated, but created according to a deterministic algorithm. Thus if the algorithm is initialized with the same seed number it generates the same sequence of pseudorandom numbers. Creating pseudorandom numbers having good statistics (e.g., less correlations between consecutive numbers, similarity to the target distribution) is a research field for itself and very important in encryption methods. Moreover, all methods that allow to sample from more complicated probability density distributions, like the probability density transformation method (appendix A.1), inversion sampling (appendix A.2), and the Metropolis algorithm described below are usually build on the pseudo-

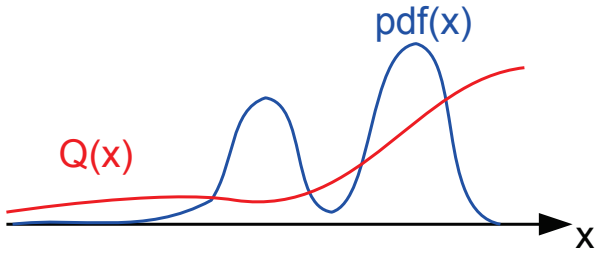


Figure 21: A given function/quantity $Q(x)$ is given and the expectation with respect to the probability density function pdf(x) shall be evaluated, i.e., $\langle Q\rangle = \int Q(x)\mathrm{pdf}(x)\ \mathrm{d}x$.

random number generators for the uniform distribution or the normal distribution. For an overview of methods of pseudorandom number generations the reader is referred to [129, §7].

In this work the Metropolis Markov Chain Monte Carlo (Metropolis MCMC) method is used in order to create sample points $\{\mathbf{x}_n\}_{n=1}^N$ in proportion to $\mathrm{pdf}(\mathbf{x})$ [129, §15.8]. This means that the points are sampled from a distribution $\pi(\mathbf{x})$ that is not normalized but proportional to the desired probability density function, i.e., $\pi(\mathbf{x}) \propto \mathrm{pdf}(\mathbf{x})$, having the advantage that no high dimensional integrals must be evaluated for determining a normalization constant. The algorithm explores the data space sequentially by a first-order Markov chain, i.e., a sequence of points $< \mathbf{x}_n >_{n=1}^N$ is created, where the probability for creating a point $\mathbf{x}_{n+1}$ depends only on the point $\mathbf{x}_n$. Thus the transition probability is given by the conditional distribution $\mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_0, \ldots, \mathbf{x}_n) = \mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n)$. Since the sequence shall be created according to the desired distribution $\pi(\mathbf{x})$, it follows that if $\mathbf{x}_n$ is sampled from $\pi(\mathbf{x})$ then the succeeding $\mathbf{x}_{n+1}$ has to be a sample from the distribution $\pi(\mathbf{x})$, too. This means that the transition probability distribution $\mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n)$ has to fulfill the equation[33]

$$\int \mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n)\ \mathrm{d}\mathbf{x}_n = \pi(\mathbf{x}_{n+1}) \tag{111}$$

such that the distribution $\pi(\mathbf{x})$ stays invariant. The distribution is then said to be invariant with respect to the chosen transition distribution. The required invariance can be imposed if the detailed balance principle

$$\mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n) = \mathrm{pdf}(\mathbf{x}_n|\mathbf{x}_{n+1})\pi(\mathbf{x}_{n+1}) \tag{112}$$

is fulfilled[34] (sufficient condition), as it can be easily seen by integrating Eq. (112) over $\mathbf{x}_n$[35]:

$$\int \mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n)\ \mathrm{d}\mathbf{x}_n = \pi(\mathbf{x}_{n+1}) \int \mathrm{pdf}(\mathbf{x}_n|\mathbf{x}_{n+1})\ \mathrm{d}\mathbf{x}_n = \pi(\mathbf{x}_{n+1}). \tag{113}$$

The detailed balance principle, also called reversibility, states that the overall transition rate from a point $\mathbf{x}_n$ to a point $\mathbf{x}_{n+1}$ equals the overall transition rate from the point $\mathbf{x}_{n+1}$ to the point $\mathbf{x}_n$. Metropolis et al. [103] were the first to suggest a way of how the transition probability distribution can be chosen. Later on, this method was generalized by Hastings [66], yielding the Metropolis-Hastings MCMC algorithm, which is as follows: the transition probability is considered to be a product of a proposal distribution $q(\mathbf{x}_{n+1}|\mathbf{x}_n)$ times an acceptance probability $\alpha(\mathbf{x}_n, \mathbf{x}_{n+1})$:

$$\mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n) = q(\mathbf{x}_{n+1}|\mathbf{x}_n)\alpha(\mathbf{x}_n, \mathbf{x}_{n+1}), \tag{114}$$

where

$$\alpha(\mathbf{x}_n, \mathbf{x}_{n+1}) := \min\left(1, \frac{q(\mathbf{x}_n|\mathbf{x}_{n+1})\pi(\mathbf{x}_{n+1})}{q(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n)}\right). \tag{115}$$

[33] Eq. (111) can also be seen as an eigenvalue equation, where the eigenfunction is $\pi(\cdot)$, the eigenvalue is one and the integral kernel $\mathrm{pdf}(\cdot|\cdot)$ is searched.

[34] Eq. (112) can be written with the normalized distributions $\mathrm{pdf}(\mathbf{x}_n)$ and $\mathrm{pdf}(\mathbf{x}_{n+1})$ as well. If done so, one can write the detailed balance principle as $\mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n)\mathrm{pdf}(\mathbf{x}_n) = \mathrm{pdf}(\mathbf{x}_{n+1}, \mathbf{x}_n) = \mathrm{pdf}(\mathbf{x}_n, \mathbf{x}_{n+1}) = \mathrm{pdf}(\mathbf{x}_n|\mathbf{x}_{n+1})\mathrm{pdf}(\mathbf{x}_{n+1})$, thus showing that detailed balance requires symmetry of the arguments in the joint probability distribution $\mathrm{pdf}(\mathbf{x}_n, \mathbf{x}_{n+1})$.

[35] If the sample distribution $\pi(\mathbf{x})$ consists of Dirac-distributions, the detailed balance principle has to be proven by showing that $\int_A \int_B \mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n)\ \mathrm{d}\mathbf{x}_{n+1}\ \mathrm{d}\mathbf{x}_n = \int_A \int_B \mathrm{pdf}(\mathbf{x}_n|\mathbf{x}_{n+1})\pi(\mathbf{x}_{n+1})\ \mathrm{d}\mathbf{x}_n\ \mathrm{d}\mathbf{x}_{n+1}$ holds for any two space regions A and B [116]. However, in this work Dirac-distributions are not considered.

Algorithmically, this can be implemented by first creating a point $\mathbf{x}_{n+1}$ according the proposal distribution, and then accepting $\mathbf{x}_{n+1}$ with probability $\alpha(\mathbf{x}_n, \mathbf{x}_{n+1})$. By substituting Eq. (114),(115) into Eq. (112) it is shown that this ansatz fulfills the detailed balance principle:

$$\begin{aligned}
\mathrm{pdf}(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n) &= q(\mathbf{x}_{n+1}|\mathbf{x}_n)\min\left(1, \frac{q(\mathbf{x}_n|\mathbf{x}_{n+1})\pi(\mathbf{x}_{n+1})}{q(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n)}\right)\pi(\mathbf{x}_n) \\
&= \min\left(q(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n),\, q(\mathbf{x}_n|\mathbf{x}_{n+1})\pi(\mathbf{x}_{n+1})\right) \\
&= q(\mathbf{x}_n|\mathbf{x}_{n+1})\min\left(\frac{q(\mathbf{x}_{n+1}|\mathbf{x}_n)\pi(\mathbf{x}_n)}{q(\mathbf{x}_n|\mathbf{x}_{n+1})\pi(\mathbf{x}_{n+1})}, 1\right)\pi(\mathbf{x}_{n+1}) \\
&= \mathrm{pdf}(\mathbf{x}_n|\mathbf{x}_{n+1})\pi(\mathbf{x}_{n+1}).
\end{aligned} \tag{116}$$

Algorithmically, the decision whether a candidate point $\mathbf{x}_{n+1}$ is accepted or rejected is implemented via a random number generator: let u a realization of a uniformly distributed random variable in $[0, 1]$ (i.e., $u \sim U(0, 1)$), then, if $u < \alpha(\mathbf{x}_n, \mathbf{x}_{n+1})$ the move is accepted, otherwise rejected. If the candidate point $\mathbf{x}_{n+1}$ is rejected by the algorithm, the algorithm sets $\mathbf{x}_n$ as $\mathbf{x}_{n+1}$ (if not done so, the sampling statistics would be wrong). Moreover, it is noticed here that since $\mathbf{x}_{n+1}$ and $\mathbf{x}_n$ are not independent, instead being strongly correlated, many more data points are required in Eq. (109) in order to determine the estimate of $\langle Q \rangle$ with sufficient accuracy.

The earlier choice of Metropolis et al. was to choose the proposal distribution symmetrically, i.e., $q(\mathbf{x}_n|\mathbf{x}_{n+1}) = q(\mathbf{x}_{n+1}|\mathbf{x}_n)$, thus accepting a new candidate point $\mathbf{x}_{n+1}$ with probability

$$\alpha(\mathbf{x}_n, \mathbf{x}_{n+1}) = \min\left(1, \frac{\pi(\mathbf{x}_{n+1})}{\pi(\mathbf{x}_n)}\right). \tag{117}$$

In their work, $\mathbf{x}$ is a point within a configuration space of particles (as it is also the case in this work, cf. algorithm 3 in section 4.2) and each particle is chosen with equal probability and is randomly moved in any direction and for a random short distance to a new position. Furthermore, their algorithm uses a maximum particle displacement distance (also used in this work), such that in general there is no direct transition from a state $\mathbf{x}_n$ to any other state $\mathbf{x}_{n+1}$. However, the algorithm fulfills the detailed balance principle, since (*i*) selecting a particle at random and moving it randomly to a new position has the same probability as selecting the particle in the new position at random and moving it randomly to its old position, and (*ii*) also if it is not possible to reach a point $\mathbf{x}_{n+1}$ from $\mathbf{x}_n$ within a single Markov chain step, the point can be reached by a number of extra intermediate steps.

Another important aspect is that whatever the initial distribution $\tilde{\pi}(\mathbf{x})$ in the algorithm for $\mathbf{x}_0$ is, the distribution $\tilde{\pi}(\mathbf{x}_n)$ should converge for $n \to \infty$ to the invariant distribution $\pi(\mathbf{x})$. If this is the case, the Markov chain is ergodic[36] and the invariant distribution is called the equilibrium distribution. Here, $\mathbf{x}_0$ is created according to $\tilde{\pi}(\mathbf{x}_0)$, but the domains, $\mathcal{D}(\cdot)$, of $\tilde{\pi}(\mathbf{x}_0)$ and $\pi(\mathbf{x})$ are considered the same. Thus $\mathbf{x}_0 \in \mathcal{D}(\pi(\mathbf{x}))$ and the Markov chain will sample from the equilibrium distribution. However, usually $\pi(\mathbf{x}_0)$ is relatively low for the initial $\mathbf{x}_0$, such that it may take a number of iterations to also reach regions in space where $\pi(\mathbf{x}_n)$ becomes relatively large. Accordingly, a good method is to discard the beginning of the Markov chain until the chain starts to explore a part of the space that occurs with higher probability, i.e., to wait until the burn-in phase[37] is over before sampling is performed. The decision as to whether the burn-in phase is over or not should always be taken by looking at the evolution of the desired quantity or of other parameter evolutions (cf. next section and

[36] There are different definitions of the technical term ergodicity, cf. [116]. Here, the definition from [16, §11] is taken.

[37] Often, the burn-in phase is also called equilibration phase.

and Fig. 23), even if there are diagnostics tools available for checking the convergence of the algorithm [129, §15.8.4].

In practice, choosing the right proposal distribution, especially its width, is often not easy. If, e.g., the proposal distribution only allows relatively small steps, then the space exploration can become extremely slow. If on the contrary, the proposal steps are created too large, most of the steps are rejected and the algorithm also only performs a slow space exploration. Often a good rule of thumb for having an acceptable fast space exploration is by choosing the proposal distribution such that the acceptance probability is roughly in $[0.1, 0.5]$ [7, §4.4], e.g., by changing appropriately the maximum step length.[38]

Another issue is the following one: Often there are regions of high probability density, called probability islands, that are separated by regions of low probability density and since the Metropolis-Hastings MCMC algorithm performs a (weighted) random walk space exploration [46, §3.2.1], it can take very long MCMC runs to sample from all islands. Possible remedies can be to use slice sampling or the hybrid Monte Carlo method [16, §11] or the technique of parallel tempering or cluster moves [46, §14].

4.2. Creating Complex Clusters

In this work, the interesting quantity in Eq. (108) is the scattering intensity of a vesicle that is decorated with particles or has internalized particles, i.e., the considered system is a complex cluster. The system is considered to be in thermodynamic equilibrium, such that an intensity average over the time can be replaced by a canonical ensemble average of the system. A canonical ensemble is an imaginary, infinite replication of the system, where all systems have the same temperature T, the same volume V, and K identically replicated particles, but the total energy of a system may change according to the positions $\mathbf{r}^K = (\mathbf{r}_1^T, \dots, \mathbf{r}_K^T)^T$ and momenta $\mathbf{p}^K = (\mathbf{p}_1^T, \dots, \mathbf{p}_K^T)^T$ of the particles. The canonical ensemble is then described statistically by the Boltzmann distribution (probability density function)

$$\mathrm{pdf}(\mathbf{r}^K, \mathbf{p}^K) = \frac{1}{Z} \exp\left(-\frac{H(\mathbf{r}^K, \mathbf{p}^K)}{k_B T}\right), \tag{118}$$

where $H(\mathbf{r}^K, \mathbf{p}^K)$ is the Hamiltonian at point $(\mathbf{r}^K, \mathbf{p}^K)$ in the phase space, and Z is the partition function

$$Z = \iint \exp\left(-\frac{H(\mathbf{r}^K, \mathbf{p}^K)}{k_B T}\right) \mathrm{d}\mathbf{r}^K \, \mathrm{d}\mathbf{p}^K, \tag{119}$$

ensuring the normalization of the probability density function. k_B is the Boltzmann constant. Thus, the canonical ensemble average of any quantity $Q(\mathbf{r}^K, \mathbf{p}^K)$ is

$$\langle Q \rangle = \frac{\iint Q(\mathbf{r}^K, \mathbf{p}^K) \exp\left(-\frac{H(\mathbf{r}^K, \mathbf{p}^K)}{k_B T}\right) \mathrm{d}\mathbf{r}^K \, \mathrm{d}\mathbf{p}^K}{\iint \exp\left(-\frac{H(\mathbf{r}^K, \mathbf{p}^K)}{k_B T}\right) \mathrm{d}\mathbf{r}^K \, \mathrm{d}\mathbf{p}^K}. \tag{120}$$

The integration over the momenta in the numerator can be nullified by the integration over the momenta in the denominator, since (i) the Hamiltonian is given as the sum of the kinetic $T(\mathbf{p}^K)$ and potential $E(\mathbf{r}^K)$ energies of the system, thus the exponential can be factorized, and (ii) the interesting quantity is the intensity $I(\mathbf{r}^K, \mathbf{q})$ (at a scattering vector $\mathbf{q}$), not depending

[38] Changing the step size should only be done as long as the algorithm has not started to generate samples, otherwise the detailed balance principle is no longer fulfilled.

on the momenta of the particles. Therefore, the remaining integrations are only over the volume of the configuration space

$$\langle I(\mathbf{q})\rangle = \frac{\int I(\mathbf{r}^K,\mathbf{q}) \exp\left(-\frac{E(\mathbf{r}^K)}{k_B T}\right) \mathrm{d}\mathbf{r}^K}{\int \exp\left(-\frac{E(\mathbf{r}^K)}{k_B T}\right) \mathrm{d}\mathbf{r}^K}, \tag{121}$$

and the unnormalized distribution $\pi(\mathbf{x})$ from the last subsection becomes

$$\pi(\mathbf{x}) = \exp\left(-\frac{E(\mathbf{r}^K)}{k_B T}\right). \tag{122}$$

Here, it is assumed (as it is commonly done, see, e.g., [114, §2]) that the potential energy of the considered system can be written as

$$E(\mathbf{r}^K) = \frac{1}{2}\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K} V(\mathbf{r}_{ij}), \tag{123}$$

where $V(\mathbf{r}_{ij})$ is the potential energy that describes the interaction of particles i and j, and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. E.g., $V(\mathbf{r}_{ij})$ can be a hard sphere potential energy that is

$$V_{\mathrm{HS}}(\mathbf{r}_{ij}) = \begin{cases} 0 & \text{particles } i \text{ and } j \text{ do not overlap} \\ \infty & \text{particles } i \text{ and } j \text{ overlap} \end{cases} \tag{124}$$

or a modified Coulomb potential energy

$$V_{\mathrm{C}}(\mathbf{r}_{ij}) = \begin{cases} c_e \frac{Q_i Q_j}{r_{ij}} & \text{particles } i \text{ and } j \text{ do not overlap} \\ \infty & \text{particles } i \text{ and } j \text{ overlap} \end{cases}, \tag{125}$$

where r_{ij} is the magnitude of $\mathbf{r}_{ij}$, Q_i and Q_j are the charges on the particles i and j, and c_e is a constant. It is considered that the number of charges on a particle scales with its surface. However, here the Coulomb potential is only used to place the particles on the vesicles, such that they have a maximum distance to each other[39], i.e., c_e is chosen so large that the particles repel each other very strongly.

Anyway, a real system will have a size distribution of vesicles, and each vesicle will have a different number of differently sized particles. Considering that there are only particles on a vesicle (no internalization) the overall intensity is given by a number of nested averages as

$$I_{cc}(q) = \iint I_c(q,R,t)\mathrm{pdf}(R)\mathrm{pdf}(t)\ \mathrm{d}R\ \mathrm{d}t, \tag{126}$$

$$I_{cc}(q,R,t) = \sum_{K=0}^{\infty} I_c(q,R,t,K)P(K,\lambda(R,t)), \tag{127}$$

$$I_{cc}(q,R,t,K) = \int I_c(q,R,t,\mathbf{R}_p^K)\mathrm{pdf}(\mathbf{R}_p^K)\ \mathrm{d}\mathbf{R}_p^K, \tag{128}$$

and

$$I_{cc}(q,R,t,\mathbf{R}_p^K) = \int \left\langle I_c(\mathbf{q},R,t,\mathbf{R}_p^K,\mathbf{r}^K)\right\rangle_{\Omega}\mathrm{pdf}(\mathbf{r}^K)\ \mathrm{d}\mathbf{r}^K. \tag{129}$$

[39] Note, the Coulomb potential cannot be the true physical potential, since the space is not homogeneous (e.g., the vesicle shell has another permittivity than water). Only numerical field calculations can determine the true potential.

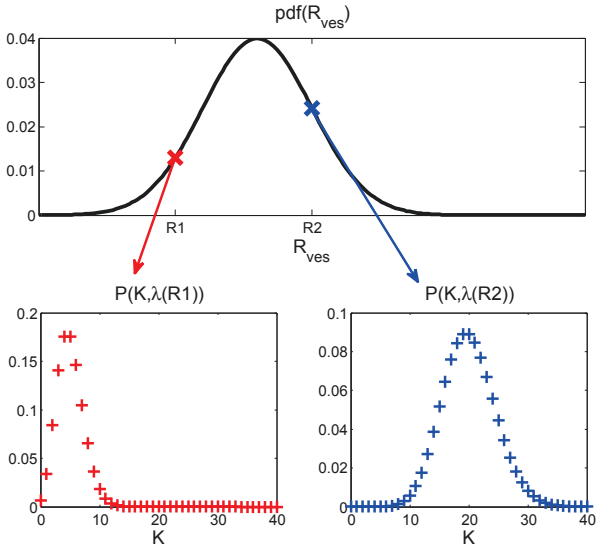


Figure 22: Schematic size distribution effect on the number distribution of particles per vesicle, if $P(K, \lambda(R))$ is assumed to be a Poisson distribution. In the upper diagram, a size distribution for the vesicle radius is plotted. In the lower left diagram, the number distribution of particles per vesicle for a vesicle radius size $R1$ is plotted. On the lower right diagram the number distribution of particles per vesicle for a vesicle radius $R2 > R1$ is plotted. $R_{\text{ves}} = R + t$.

Herein, $\text{pdf}(R)$ and $\text{pdf}(t)$ are the size distributions of the vesicle core radius R and of the vesicle shell thickness t, $P(K, \lambda(R,t))$ can be a Poisson distribution $P(K, \lambda(R,t)) = \lambda^K(R,t)/K! \cdot \exp(-\lambda(R,t))$ that gives the probability to have K particles on a vesicle, having radius R and shell thickness t, when $\lambda(R,t)$ is the mean. The mean of the distribution is assumed to change proportionally to the surface of a considered vesicle, cf. Fig. 22, according to

$$\lambda(R,t) = \lambda_{\langle R+t \rangle_{R,t}} \frac{(R+t)^2}{\langle R+t \rangle_{R,t}^2}, \tag{130}$$

where $\lambda_{\langle R+t \rangle_{R,t}}$ is average number of particles at the average (outer) vesicle radius

$$\langle R+t \rangle_{R,t} = \iint (R+t)\text{pdf}(R)\text{pdf}(t)\, \mathrm{d}R\, \mathrm{d}t. \tag{131}$$

As another option $P(K, \lambda(R,t))$ can be a Dirac distribution, i.e., all vesicles have a fixed number of particles. $\text{pdf}(\mathbf{R}_p^K) = \text{pdf}(R_p^{(1)}) \cdot \ldots \cdot \text{pdf}(R_p^{(K)})$ is the joint radii size distribution of the K particles of a vesicle, and $\left\langle I_c(\mathbf{q}, R, t, \mathbf{R}_p^K, \mathbf{r}^K) \right\rangle_\Omega$ is the isotropic intensity of an orientational average of the scattering intensity of a single vesicle with K particles. The last expectation, Eq. (129), is over the configuration space, where $\text{pdf}(\mathbf{r}^K)$ is conditioned on the cluster variables R, t and $\mathbf{R}_p^K$, i.e.,

$$\text{pdf}(\mathbf{r}^K) := \text{pdf}(\mathbf{r}^K | R, t, \mathbf{R}_p^K) = \frac{\exp\left(-\frac{E(\mathbf{r}^K, \mathbf{R}_p^K)}{k_B T}\right)}{\int \exp\left(-\frac{E(\mathbf{r}^K, \mathbf{R}_p^K)}{k_B T}\right) \mathrm{d}\mathbf{r}^K}, \tag{132}$$

since the distribution of configurations is considered for a certain complex cluster.

The nested expectation calculation given in Eq. (126)-(129) is then MC integrated by sequentially sampling values from the involved distributions given in Eq. (126)-(129). This means, that at first N samples $\{(R_n, t_n)\}_{n=1}^{N}$ are taken from pdf(R) and pdf(t) in Eq. (126), and then for each sample pair (R_n, t_n) exactly one further sample is sequentially drawn from the distributions in Eq. (127)-(129), yielding the overall sample set $\{(R_n, t_n, \mathbf{R}_p^{K_n}, \mathbf{r}^{K_n})\}_{n=1}^{N}$, and the mean (isotropic) intensity form factor

$$\overline{I}_{cc}(q) = \frac{1}{N}\sum_{n=1}^{N} \left\langle I_{cc}(\mathbf{q}, R_n, t_n, \mathbf{R}_p^{K_n}, \mathbf{r}^{K_n}) \right\rangle_{\Omega}. \tag{133}$$

The detailed formula of how the intensity is calculated is given in section 4.3.

As an extension to the decoration of vesicles by particles, it is considered in this work that each particle that was firstly placed on the vesicle may be later internalized with a specified probability P_{in}. Thus, if the n-th vesicle was created together with M_n particles on it, L_n particles may be internalized later, leaving only $K_n = M_n - L_n$ particles on the vesicle. If a particle is internalized it is considered that it receives a particle shell (encapsulation), consisting of the same material as the vesicle shell is built of and having a particle shell thickness distribution identically to the vesicle shell distribution pdf(t). There are two options for the internalization process in the program: if a particle will be internalized (i) it can get additional material for building its shell or (ii) it can take shell material from the vesicle shell in order to build its own shell. In the second case, the vesicle shell core radius needs to be adjusted (assuming that the vesicle shell thickness stays fixed) according to the loss of shell material required for building the shells of the internalized particles. Finally, the mean intensity form factor reads

$$\overline{I}_{cc}(q) = \frac{1}{N}\sum_{n=1}^{N} \left\langle I_{cc}(\mathbf{q}, R_n, t_n, \mathbf{R}_p^{K_n}, \mathbf{r}^{K_n}, \mathbf{R}_p^{L_n}, \mathbf{r}^{L_n}, \mathbf{t}_p^{L_n}) \right\rangle_{\Omega}, \tag{134}$$

where $\mathbf{r}_p^{L_n}$ are the positions of the internalized particles, $\mathbf{R}_p^{L_n}$ contains the radii of the internalized particles and $\mathbf{t}_p^{L_n}$ are corresponding particle shell thicknesses. The detailed scattering formula is given in section 4.3.

In the developed program, it is considered that the particles on a vesicle do not interact with the internalized particles. Hence, the configurations on and in a vesicle are determined independently from each other. Algorithm 4 describes the overall complex cluster parameter generation that uses the algorithm 3 (the Metropolis MCMC algorithm) as a subroutine in order to determine the configurations of particles on and in a vesicle. In the implementation of the algorithm 4 the for-loop is implemented as a parallel one in order to speed up the code execution. This is possible because in any iteration the loop body is independent of any other iteration.

Some remarks are necessary for the Metropolis MCMC algorithm (algorithm 3):

1. Whether the burn-in phase is finished or not is inferred by plotting

$$\lambda_{\text{dist.}} = \frac{1}{K_n}\sum_{k=1}^{K_n} \min_{k' \neq k} \|\mathbf{r}_k - \mathbf{r}_{k'}\|, \tag{135}$$

which is the average minimum distance between particles, over the number of MCMC-steps and then by visually determining the end of the burn-in phase, cf. Fig. 23. Practically, it is enough to evaluate this parameter every i-th MCMC iteration, in order to keep the code execution efficient.

Algorithm 3 Determination of one configuration of particles via the Metropolis-MCMC method. In this algorithm equilibration is checked by means of plotting the quantity in Eq. (135) over the number of MCMC-steps and then by deciding from visual inspection when the burn-phase is over.

initialize the system: set the particles randomly in / on the vesicle (using algorithm 5 or 6) **or** optionally, if the volume fraction of particles in the vesicle is too high, assume a larger vesicle, and then perform a vesicle shrinking
calculate the energy E of the system
$burnInPhase \leftarrow$ true
while $burnInPhase =$ true **do**
    take a particle at random and move it randomly, according to a given symmetrical proposal distribution, to a new position
    calculate the new energy E' of the system
    if $E' < E$ **then**
        accept the new particle position and set $E \leftarrow E'$
    else
        generate a random realization $u \sim U(0,1)$
        if $u < e^{-(E'-E)/(k_B T)}$ **then**
            accept the new particle position and set $E \leftarrow E'$
        else
            set the particle to its previous position
        end if
    end if
    if burn-in phase is over **then**
        $burnInPhase \leftarrow$ false
    end if
end while
return configuration of particles

Algorithm 4 Creation of the overall MC sample set.

MCSampleSet $= \{\emptyset\}$
for $n = 1$ **to** N **do**
    create samples $R_n \sim \text{pdf}(R)$ and $t_n \sim \text{pdf}(t)$
    create M_n particles according to $P(M, \lambda(R_n, t_n))$
    create M_n particle radii $\mathbf{R}_p^{M_n} \sim \text{pdf}(\mathbf{R}_p^M)$
    with probability P_{in} let each of the M_n particles be in the vesicle, hence $\mathbf{R}_p^{K_n}$ are the radii of the particles that are on the vesicle and $\mathbf{R}_p^{L_n}$ are the radii of the internalized particles
    optional: generate particle shells for the internalized particles — according to pdf(t)
    optional: recalculate the vesicle core radius R_n if internalized particles took vesicle shell
    create one configuration of particles for the particles **on** a vesicle $\mathbf{r}^{K_n} \sim \text{pdf}(\mathbf{r}^K) = \text{pdf}(\mathbf{r}^K | R_n, t_n, \mathbf{R}_p^{K_n})$ via algorithm 3
    create one configuration of particles for the particles **in** a vesicle $\mathbf{r}^{L_n} \sim \text{pdf}(\mathbf{r}^L) = \text{pdf}(\mathbf{r}^L | R_n, \mathbf{R}_p^{L_n}, \mathbf{t}_p^{L_n})$ via algorithm 3
    MCSampleSet $\leftarrow \{\text{MCSampleSet}, (R_n, t_n, \mathbf{R}_p^{K_n}, \mathbf{r}^{K_n}, \mathbf{R}_p^{L_n}, \mathbf{r}^{L_n}, \mathbf{t}_p^{L_n})\}$
end for
return MCSampleSet

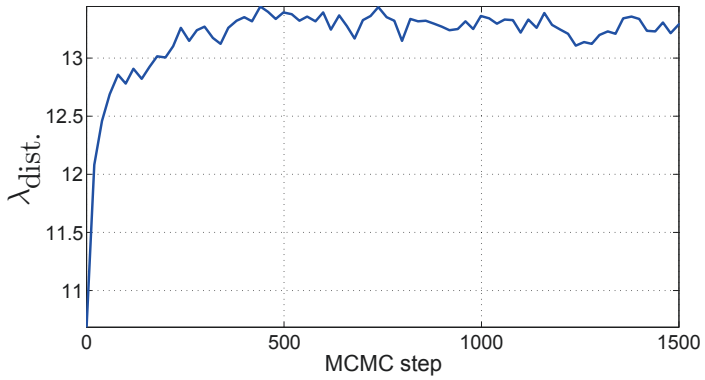


Figure 23: Exemplary evolution of the parameter $\lambda_{\text{dist.}}$ in Eq. (135) over the number of MCMC steps. The burn-in phase is approximately finished after 500 iterations.

2. In order to get initial configurations of the particles in/on a vesicle (cf. line 1 of algorithm 3), the particles can be randomly placed in/on the vesicle via (*i*) a sequential particle setting procedure described in algorithm 5 or (*ii*) a simultaneous particle setting procedure described in algorithm 6 or (*iii*) see next point.

   In comparison to the simultaneous particle setting algorithm, the sequential algorithm can be much faster in finding an initial configuration of particles if the number of particles is relatively high and if the particles occupy a relatively high volume fraction in a vesicle or a relatively high surface fraction on it.

Algorithm 5 Sequential particle setting in/on a vesicle.

```
1: while there are not set particles do
2:     take a not set particle at random
3:     maxIt ← default integer value
4:     while maxIt > 0 do
5:         if the particle shall be set in the vesicle, create a new particle position from a
   uniformly distributed probability distribution function within the vesicle, otherwise create
   a position from a uniformly distributed probability function on the vesicle
6:         if particle overlaps with any already placed particle then
7:             reject new particle position
8:             maxIt ← maxIt − 1
9:             if maxIt = 0 then
10:                set a few already placed particles as not set again
11:                jump to the first code line (perform this jump not more then x times)
12:            end if
13:        else
14:            accept the new particle position
15:            maxIt ← 0
16:        end if
17:    end while
18: end while
```

In order to create points from a uniform distribution of points in/on a vesicle, the inversion sampling method (see appendix A.2) has been applied:

Algorithm 6 Simultaneous particle setting in/on a vesicle.

```
maxIt ← default integer value
while maxIt > 0 do
    if the particles shall be set in the vesicle, create for all particles new positions from a uniformly distributed probability distribution function within the vesicle, otherwise create points from a uniformly distributed probability function on the vesicle
    if there is a particle overlap then
        reject all positions
        maxIt ← maxIt − 1
    else
        maxIt ← 0
    end if
end while
```

a) Uniformly distributed random points on a sphere: Let θ be the polar angle and ϕ the azimuthal angle. The probability of finding a point on a sphere that lies in an infinitesimal small solid angle around $\Omega = (\theta, \phi)$ is

$$\mathrm{pdf}_\Omega(\theta, \phi)\ \mathrm{d}\theta\ \mathrm{d}\phi, \tag{136}$$

where

$$\mathrm{pdf}_\Omega(\theta, \phi) = \frac{\sin(\theta)}{4\pi}. \tag{137}$$

The denominator is chosen such that

$$\int_0^{2\pi} \int_0^{\pi} \mathrm{pdf}_\Omega(\theta, \phi)\ \mathrm{d}\theta\ \mathrm{d}\phi = 1. \tag{138}$$

Thus, the two marginal density functions are

$$\mathrm{pdf}(\theta) = \int_0^{2\pi} \mathrm{pdf}_\Omega(\theta, \phi)\ \mathrm{d}\phi = \frac{\sin(\theta)}{2}, \tag{139}$$

and

$$\mathrm{pdf}(\phi) = \int_0^{\pi} \mathrm{pdf}_\Omega(\theta, \phi)\ \mathrm{d}\theta = \frac{1}{2\pi}, \tag{140}$$

and their cumulative distribution functions are

$$F_\theta(\theta_0) = \int_0^{\theta_0} \mathrm{pdf}(\theta)\ \mathrm{d}\theta = \frac{1}{2}(1 - \cos(\theta_0)), \tag{141}$$

and

$$F_\phi(\phi_0) = \int_0^{\phi_0} \mathrm{pdf}(\phi)\ \mathrm{d}\phi = \frac{\phi_0}{2\pi}. \tag{142}$$

Applying Eq. (373), it follows for two independent and uniformly distributed random variables $u_1 \sim U(0, 1)$ and $u_2 \sim U(0, 1)$ that

$$F_\theta^{-1}(u_1) = \theta_0 \quad \Rightarrow \quad \theta_0(u_1) = \arccos(1 - 2u_1) \tag{143}$$

and

$$F_\phi^{-1}(u_2) = \phi_0 \quad \Rightarrow \quad \phi_0(u_2) = 2\pi u_2. \tag{144}$$

where $\theta_0(u_1)$ and $\phi_0(u_2)$ give the appropriate random polar and azimuthal angles, such that $(\theta_0(u_1), \phi_0(u_2))$ is a uniformly distributed random point on a sphere.

b) Uniformly distributed random points in a three dimensional ball (note: a second method is given in appendix A.3): Let θ be the polar angle, ϕ the azimuthal angle and r the radial distance. Then, the probability of finding a point in a ball with radius R and lying in an infinitesimal small volume element around $M = (r, \theta, \phi)$ is

$$\mathrm{pdf}_M(r, \theta, \phi)\ \mathrm{d}r\ \mathrm{d}\theta\ \mathrm{d}\phi, \tag{145}$$

where

$$\mathrm{pdf}_M(r, \theta, \phi) = \frac{r^2 \sin(\theta)}{\frac{4\pi}{3} R^3}. \tag{146}$$

The denominator is chosen such that

$$\int_0^{2\pi} \int_0^{\pi} \int_0^{R} \mathrm{pdf}_M(r, \theta, \phi)\ \mathrm{d}r\ \mathrm{d}\theta\ \mathrm{d}\phi = 1. \tag{147}$$

Thus, there are the three marginal density functions:

$$\mathrm{pdf}(r) = \int_0^{2\pi} \int_0^{\pi} \mathrm{pdf}_M(r, \theta, \phi)\ \mathrm{d}\theta\ \mathrm{d}\phi = \frac{3r^2}{R^3}, \tag{148}$$

$$\mathrm{pdf}(\theta) = \int_0^{2\pi} \int_0^{R} \mathrm{pdf}_M(r, \theta, \phi)\ \mathrm{d}r\ \mathrm{d}\phi = \frac{\sin(\theta)}{2}, \text{ and} \tag{149}$$

$$\mathrm{pdf}(\phi) = \int_0^{\pi} \int_0^{R} \mathrm{pdf}_M(r, \theta, \phi)\ \mathrm{d}r\ \mathrm{d}\theta = \frac{1}{2\pi}, \tag{150}$$

and their appropriate cumulative distribution functions are

$$F_t(r_0) = \int_0^{r_0} \mathrm{pdf}(r)\ \mathrm{d}r = \frac{r_0^3}{R^3}, \tag{151}$$

$$F_\theta(\theta_0) = \int_0^{\theta_0} = \mathrm{pdf}(\theta)\ \mathrm{d}\theta = \frac{1}{2}(1 - \cos(\theta_0)), \text{ and} \tag{152}$$

$$F_\phi(\phi_0) = \int_0^{\phi_0} \mathrm{pdf}(\phi)\ \mathrm{d}\phi = \frac{\phi_0}{2\pi}. \tag{153}$$

Applying Eq. (373) it follows for three independent and uniformly distributed random variables $u_1 \sim U(0,1)$, $u_2 \sim U(0,1)$ and $u_3 \sim U(0,1)$ that

$$F_r^{-1}(u_1) = r_0 \quad \Rightarrow \quad r_0(u_1) = R\sqrt[3]{u_1}, \tag{154}$$

$$F_\theta^{-1}(u_2) = \theta_0 \quad \Rightarrow \quad \theta_0(u_2) = \arccos(1 - 2u_2), \text{ and} \tag{155}$$

$$F_\phi^{-1}(u_3) = \phi_0 \quad \Rightarrow \quad \phi_0(u_3) = 2\pi u_3, \tag{156}$$

where $r_0(u_1)$, $\theta_0(u_2)$ and $\phi_0(u_3)$ give the appropriate random radial distance, and random polar and azimuthal angles, such that $(r_0(u_1), \theta_0(u_2), \phi_0(u_3))$ is uniformly distributed within a ball having radius R.

3. It might be that the initial volume fraction of particles within the vesicle is relatively high, such that it is very unlikely to find an initial configuration of particles. Optionally (cf. line 1 of algorithm 3), in this case the implemented algorithm allows to put the particles within a blown up vesicle, firstly. Then, the vesicle is slowly decreasing its size, while the particles are moving around, until the original (desired) vesicle size is reached.

4. It can be that not all particles can be placed within the vesicle. In this case the implemented algorithm has a further option (instead of just producing an error message stating that a current desired configuration is not possible): since the algorithm puts the particles sequentially inside the vesicle, it can stop putting further particles in the vesicle if it becomes highly improbable.

5. According to algorithm 3 the particles have to be moved randomly to new positions. The move is described by the proposal distribution (used in Eq. (114)), and is given by the following methods:

   a) For internalized particles it is considered that the particles can move randomly in any direction within the vesicle, and that the maximum allowed displacement is $\delta R_{\max}$. Then with Eq. (154)-(156), where R is replaced by $\delta R_{\max}$ in Eq. (154), the particle shift is determined.

   b) For particles on a vesicle it is assumed that a particle can move randomly in any direction and for a random arc length. Such a particle move can be described as follows — see Fig. (24): Consider that a particle on a vesicle has the spherical coordinates (r_o, θ_o, ϕ_o) and that there is a Cartesian coordinate system (x_o, y_o, z_o) attached to it. The particle is then placed at $(r_o, 0, 0)$ within the Cartesian coordinate system (x, y, z), and randomly moved along the azimuthal direction

$$\delta\phi \sim 2\pi \cdot U(0,1) \tag{157}$$

   and for a polar angle

$$\delta\theta \sim \delta\theta_{\max} \cdot U(0,1) \tag{158}$$

   far away from $(r_o, 0, 0)$ to the spherical coordinate $(r_o, \delta\theta, \delta\phi)$, where $\delta\theta_{\max}$ is a maximal allowed angle, bounding the displacement move of the particle to a maximum arc length $b_{\max} = r_0 \delta\theta_{\max}$. The new position is in the Cartesian coordinate system (x, y, z) given as

$$\mathbf{r}_\delta = \begin{bmatrix} x_\delta &=& r_o \sin(\delta\theta)\cos(\delta\phi) \\ y_\delta &=& r_o \sin(\delta\theta)\sin(\delta\phi) \\ z_\delta &=& r_o \cos(\delta\theta) \end{bmatrix}. \tag{159}$$

   Subsequently, the Cartesian coordinate system (x, y, z) is rotated via the rotation matrices

$$\mathbf{R}_y(\alpha) = \begin{bmatrix} \cos(\alpha) & 0 & \sin(\alpha) \\ 0 & 1 & 0 \\ -\sin(\alpha) & 0 & \cos(\alpha) \end{bmatrix}, \tag{160}$$

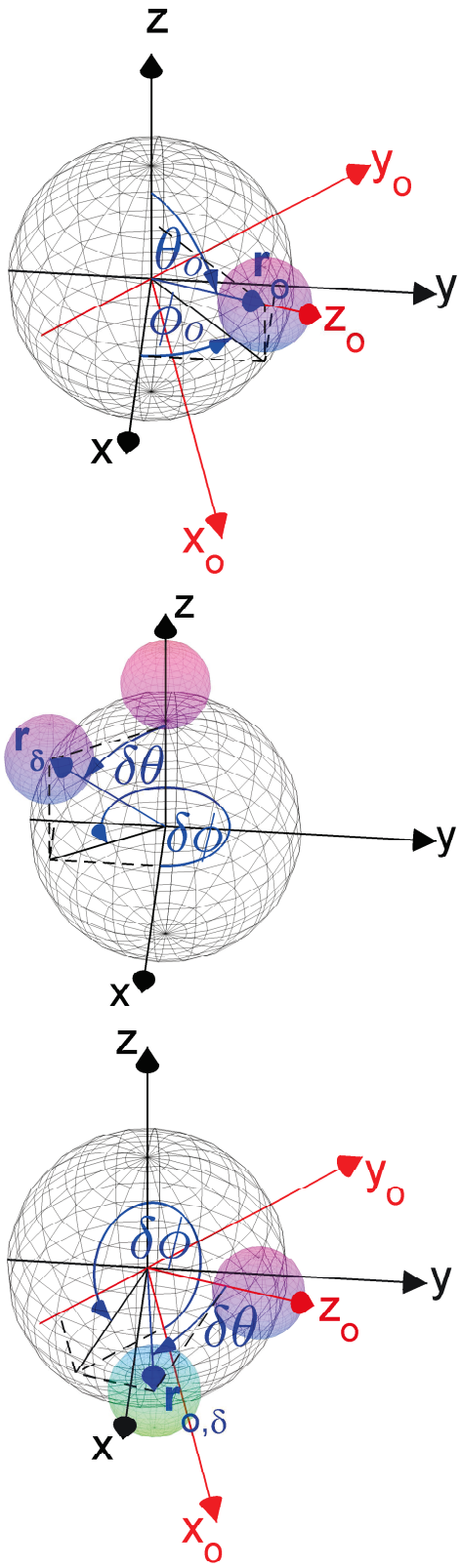
   and

$$\mathbf{R}_z(\alpha) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{161}$$

   into the coordinate system (x_o, y_o, z_o), and correspondingly also the displaced particle:

$$\mathbf{r}_{o,\delta} = \mathbf{R}_z(\phi_o)\mathbf{R}_y(\theta_o)\mathbf{r}_\delta. \tag{162}$$

6. In the implementation of the Metropolis algorithm 3 the maximum particle displacement (i.e., the width of the proposal distribution) can automatically be adjusted, such that the acceptance rate becomes optimal as discussed in section 4.1. This is done by determining



(a) The original particle is positioned at $\mathbf{r}_o$ within the Cartesian coordinate system (x, y, z) with appropriate spherical coordinates (r_0, θ_0, ϕ_0). The coordinate system (x_0, y_0, z_0) is attached to the particle, and is given by a rotation of the (x, y, z) coordinate system around the y-axis for an angle θ_0 and a subsequent rotation of the coordinate system around the z-axis for an angle ϕ_0.

(b) In order to move the original particle randomly in any direction on the sphere for a random distance, it is firstly considered that the particle is placed on the top of the vesicle. From this position the particle is then randomly moved in a direction given by $\delta\phi \sim 2\pi U(0, 1)$, and for a certain distance given by the polar angle $\delta\theta \sim \delta\theta_{\max} U(0, 1)$.

(c) In the second step, the randomly moved particle from (b) is rotated into the coordinate system (x_0, y_0, z_0) by applying Eq. (162).

Figure 24: A random particle displacement on a vesicle.

the acceptance rate for a of number of MC steps (e.g., for 100 iterations). Then, if the acceptance rate in this interval is too low (e.g., less than 0.2) the maximal particle displacement is decreased by a certain factor (e.g., 0.5), and if the acceptance rate is too high (e.g., more than 0.6) the maximal particle displacement is increased by a certain factor (e.g., 1.2).

7. The developed program allows to calculate the average form factor of a system of different complex clusters. In this case, only the geometric sizes and the number distribution of particles per vesicle have to be specified.

   However, another option allows to take into account the volume fraction of particles and the volume fraction of shell material (vesicle/particle shell) in order to calculate the overall intensity of a considered system. In this case either (*i*) the average number of particles per vesicle has to be determined from the other given quantities, or (*ii*) the mean of the vesicle distribution needs to be determined from the other given quantities. The missing parameter is determined under the constraint that there shall be no free particles as well as no remaining shell material.

   In the description below the determination of the missing parameter for the two cases is given. Therein, the following quantities are used: φ_s is the volume fraction of shell material, φ_p is the volume fraction of particles, $\mathrm{pdf}_{R_p}(R_p)$ is the particle radius size distribution, $\mathrm{pdf}(t)$ is the shell thickness size distribution. In both cases it is considered that the different size distributions $\mathrm{pdf}_x(x)$ are controlled by only two parameters, the mean μ_x of the distribution and its polydispersity (parameter) p_x, which is defined as the standard deviation of x over μ_x, hence $\mathrm{pdf}_x(x) = \mathrm{pdf}_x(x; \mu_x, p_x)$.

   Case (*i*): The average number of particles per vesicle, i.e.,

$$\langle \lambda(R+t) \rangle_{R,t} = \frac{\lambda_{\langle R+t \rangle_{R,t}}}{\langle R+t \rangle_{R,t}^2} \big\langle (R+t)^2 \big\rangle_{R,t}, \tag{163}$$

   needs to be determined (average of Eq. (130)). This is easily accomplished, since the number density of particles n_p has to be equal to the number density of vesicles n_v times the average number of particles per vesicle:

$$n_p = n_v \langle \lambda(R+t) \rangle_{R,t} \tag{164}$$

$$\Rightarrow \qquad \langle \lambda(R+t) \rangle_{R,t} = \frac{n_p}{n_v}. \tag{165}$$

   The number density of particles is

$$n_p = \frac{\varphi_p}{\langle v_p \rangle_{R_p}}, \tag{166}$$

   where

$$\langle v_p \rangle_{R_p} = \frac{4\pi}{3} \int R_p^3 \, \mathrm{pdf}_{R_p}(R_p) \, \mathrm{d}R_p \tag{167}$$

   is the average particle volume. The number density of vesicles is

$$n_v = \frac{\varphi_s}{\langle v_s \rangle_{R,t}}, \tag{168}$$

   where

$$\langle v_s \rangle_{R,t} = \frac{4\pi}{3} \iint \left((R+t)^3 - R^3 \right) \mathrm{pdf}(R) \mathrm{pdf}(t) \, \mathrm{d}R \, \mathrm{d}t \tag{169}$$

is the average shell volume associated to a vesicle/cluster. Note: If a vesicle has internalized particles, the particles may have taken a part of the shell volume of the vesicle for their encapsulation, but the amount of shell volume associated to each vesicle will not change. Hence, the formula Eq. (165), respectively Eq. (169), stays unchanged.

Case (*ii*): The mean of the vesicle distribution needs to be determined. Again, Eq. (164) has to be solved, but now $\langle\lambda(R+t)\rangle_{R,t}$ is given, while n_v is undetermined, since it depends on the unknown mean vesicle core radius μ_R, i.e., $n_v = n_v(\mu_R)$. In order to determine the mean vesicle radius that solves Eq. (164) the following optimization problem can be solved:

$$\mu_R^* = \arg\min_{\mu_R} \left(n_p - n_v(\mu_R)\,\langle\lambda(R+t)\rangle_{R,t}\right)^2 \qquad \text{s.t.} \quad \mu_R > 0. \tag{170}$$

However, if only Gaussian size distributions

$$\mathrm{pdf}(x;\mu,p) = \frac{1}{p\mu\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{(p\mu)^2}\right) \tag{171}$$

are considered, where μ is the mean and p the polydispersity of the distribution, an analytical solution can be found: At first, it is noted that the second moment $\langle x^2\rangle_x$ and the third moment $\langle x^3\rangle_x$ of the Gaussian distribution[40] can be written as

$$\langle x^2\rangle_x = \mu^2(1+p^2), \text{ and} \tag{172}$$
$$\langle x^3\rangle_x = \mu^3 + 3\mu(\mu p)^2. \tag{173}$$

This can then be used to analytically solve Eq. (164) with respect to μ_R:

$$\begin{aligned}
&\text{(Eq. (168) } \rightarrow \text{ Eq. (164))} \qquad && n_p = \frac{\varphi_s}{\langle v_s\rangle_{R,t}}\,\langle\lambda(R+t)\rangle_{R,t} \\
&\Leftrightarrow && \langle v_s\rangle_{R,t} = \frac{\varphi_s}{n_p}\langle\lambda(R+t)\rangle_{R,t} \\
&\Leftrightarrow && \langle 3R^2t + 3Rt^2 + t^3\rangle_{R,t} = \frac{3}{4\pi}\frac{\varphi_s}{n_p}\langle\lambda(R+t)\rangle_{R,t} \\
&\Leftrightarrow && \langle R^2\rangle_R\langle t\rangle_t + \langle R\rangle_R\langle t^2\rangle_t + \frac{\langle t^3\rangle_t}{3} = \frac{1}{4\pi}\frac{\varphi_s}{n_p}\langle\lambda(R+t)\rangle_{R,t} \\
&\Leftrightarrow && \mu_R^2(1+p_R^2) + \mu_R\frac{\langle t^2\rangle_t}{\langle t\rangle_t} + \frac{\langle t^3\rangle_t}{3\langle t\rangle_t} = \frac{1}{4\pi}\frac{\varphi_s}{n_p\langle t\rangle_t}\langle\lambda(R+t)\rangle_{R,t} \\
&\Leftrightarrow && \mu_R^2 + \mu_R P + Q = 0
\end{aligned} \tag{174}$$

where

$$P = \frac{\langle t^2\rangle_t}{\langle t\rangle_t(1+p_R^2)} \tag{175}$$

and

$$Q = \frac{\langle t^3\rangle_t}{3\langle t\rangle_t(1+p_R^2)} - \frac{1}{4\pi}\frac{\varphi_s}{n_p\langle t\rangle_t(1+p_R^2)}\langle\lambda(R+t)\rangle_{R,t}. \tag{176}$$

[40] Note, the Gaussian distribution is positive for negative arguments, which is physically not possible for a geometric quantity (thickness or radius). Nevertheless, for low polydispersities ($p \lesssim 0.25$) the probability of producing a negative value is usually rather small — meaning that in MC simulations as done here, negative values usually do not realize (there are only a few hundreds random parameter realizations). However, if a negative value occurs the simulation will be canceled.

Then, the physical solution for the mean vesicle radius is

$$\mu_R = -\frac{P}{2} + \sqrt{\left(\frac{P}{2}\right)^2 - Q}. \tag{177}$$

8. In case that a hard sphere interaction potential is used, one may wonder if it is feasible to let the algorithm 3 stop directly after line 1. In case that the simultaneous particle setting algorithm 6 is used, this is admissible. However, it is not anymore admissible in case of the sequential particle setting algorithm 5: The sequential particle setting procedure does not give configurations that are drawn from the desired equilibrium distribution. There is an illustrative example in appendix B.1 explaining this issue. The sequential particle setting algorithm may find a configuration faster than the simultaneous particle setting algorithm, but it requires additional time to perform equilibration, and thus the simultaneous algorithm may be faster nevertheless. However, in practice, when there are a number of particles that occupy a relatively large volume/surface fraction, the sequential algorithm usually outperforms the simultaneous particle setting algorithm.

4.3. Calculating the Scattering Intensity of Complex Clusters

Using the methods described in section 2.1 the scattering amplitude of a complex cluster (i.e., a decorated vesicle with internalized particles) as depicted in Fig. 20 can be written as the 3-dimensional Fourier transform of the excess scattering length density

$$A_{\mathrm{cc}}(\mathbf{q}) = \int\limits_{\dot{\bigcup}\limits_{\{o_m\}_{m=1}^M} v_{o_m}} \big(\mathrm{SLD}(\mathbf{r}) - \mathrm{SLD}_{\mathrm{matrix}}\big) e^{-i\mathbf{q}^T\mathbf{r}}\, \mathrm{d}\mathbf{r}, \tag{178}$$

where $\dot{\bigcup}\limits_{\{o_m\}_{m=1}^M} v_{o_k}$ indicates the disjoint union of the volumes of the M objects o_m that constitute the complex cluster, and that are embedded in the matrix. Objects of a cluster are the vesicle shell, the particles on the outside of the vesicle shell, and the internalized particles with shell. In this work, the particles and the vesicle shell are considered to be spherical. Without loss of generality, the vesicle shell may be centered at the origin, thus Eq. (178) becomes

$$\begin{aligned} A_{\mathrm{cc}}(\mathbf{q}) = A_{\mathrm{shell}}(q,R,t) &+ \sum_{k=1}^{\#(\text{particles on shell})} A_{\mathrm{sphere}}(q,R_k) e^{-i\mathbf{q}^T\mathbf{r}_k} \\ &+ \sum_{l=1}^{\#(\text{particles in core})} A_{\text{core-shell}}(q,R_l,t_l) e^{-i\mathbf{q}^T\mathbf{r}_l}. \end{aligned} \tag{179}$$

$A_{\mathrm{shell}}(q,R,t)$ gives the contribution of the vesicle shell, where R is the vesicle radius and t the thickness of the vesicle shell. The formula of it reads as follows

$$\begin{aligned} A_{\mathrm{shell}}(q,R,t) &= \int_{v_{\mathrm{shell}}} (\mathrm{SLD}_{\mathrm{shell}} - \mathrm{SLD}_{\mathrm{matrix}}) e^{-i\mathbf{q}^T\mathbf{r}}\, \mathrm{d}\mathbf{r} \\ &= 4\pi \int_R^{R+t} (\mathrm{SLD}_{\mathrm{shell}} - \mathrm{SLD}_{\mathrm{matrix}}) r^2 \frac{\sin(qr)}{qr}\, \mathrm{d}r \\ &= 4\pi(\mathrm{SLD}_{\mathrm{shell}} - \mathrm{SLD}_{\mathrm{matrix}}) \\ &\times \left(\frac{\sin(q(R+t)) - q(R+t)\cos(q(R+t))}{q^3} - \frac{\sin(qR) - qR\cos(qR))}{q^3} \right), \end{aligned} \tag{180}$$

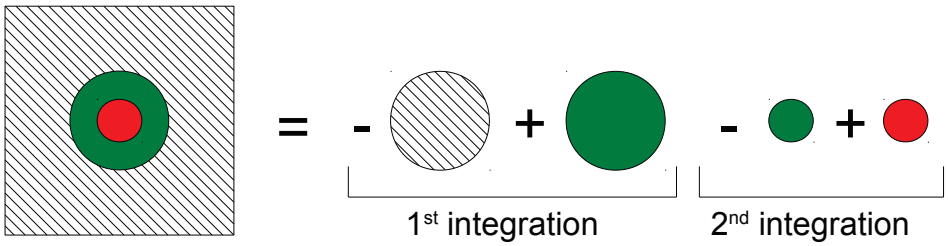


Figure 25: Schematic procedure for determining a core-shell amplitude form factor.

and the forward amplitude is

$$\lim_{q\to 0} A_{\text{shell}}(q,R,t) = (\text{SLD}_{\text{shell}} - \text{SLD}_{\text{matrix}})\left[v_{\text{sphere}}(R+t) - v_{\text{sphere}}(R)\right], \tag{181}$$

where $v_{\text{sphere}}(R)$ is the volume of a sphere with radius R, Eq. (9). The second term on the r.h.s. of Eq. (179) gives the amplitude contribution of the particles on the shell. Herein, the k-th particle contribution is given by an amplitude form factor of a sphere, Eq. (8), with radius R_k that is situated at the origin, and that is then moved to the position $\mathbf{r}_k$. Such a translation is described mathematically by a multiplication of the sphere amplitude form factor with $e^{-i\mathbf{q}^T\mathbf{r}_k}$, cf. Eq. (13). The last term on the r.h.s. of Eq. (179) is the contribution of the particles internalized in the vesicle, and analogously described as the previous term, but a core-shell amplitude form factor is taken into account for the core-shell particles. This amplitude form factor is given by

$$\begin{aligned} A_{\text{core-shell}}(q,R_l,t_l) &= \int_{v_{\text{total}}} (\text{SLD}_{\text{shell}} - \text{SLD}_{\text{matrix}}) e^{-i\mathbf{q}^T\mathbf{r}}\,\mathrm{d}\mathbf{r} \\ &\quad + \int_{v_{\text{core}}} (\text{SLD}_{\text{core}} - \text{SLD}_{\text{shell}}) e^{-i\mathbf{q}^T\mathbf{r}}\,\mathrm{d}\mathbf{r} \\ &= 4\pi(\text{SLD}_{\text{shell}} - \text{SLD}_{\text{matrix}}) \frac{\sin(q(R_l+t_l)) - q(R_l+t_l)\cos(q(R_l+t_l))}{q^3} \\ &\quad + 4\pi(\text{SLD}_{\text{core}} - \text{SLD}_{\text{shell}}) \frac{\sin(qR_l) - qR_l\cos(qR_l)}{q^3}, \end{aligned} \tag{182}$$

and the forward amplitude is

$$\begin{aligned} \lim_{q\to 0} A_{\text{core-shell}}(q,R_l,t_l) &= (\text{SLD}_{\text{shell}} - \text{SLD}_{\text{matrix}}) v_{\text{sphere}}(R_l+t_l) \\ &\quad + (\text{SLD}_{\text{core}} - \text{SLD}_{\text{shell}}) v_{\text{sphere}}(R_l). \end{aligned} \tag{183}$$

The first integral in Eq. (182) is an integration over the complete particle volume, but since the shell material does not exist in the core of the particle, it is removed by the second integration over the core volume. This integration scheme can be nicely illustrated as done in Fig. 25, and is already used in Eq. (7).

The intensity of the decorated vesicle with internalized particles is obtained by substituting

Eq. (178) into Eq. (4) yielding

$$\begin{aligned}
I_{\mathrm{cc}}(\mathbf{q}) &= A_{\mathrm{cc}}(\mathbf{q}) \cdot A_{\mathrm{cc}}^{*}(\mathbf{q}) \\
&= A_{\mathrm{shell}}^{2}(q,R,t) \\
&\quad + \sum_{k}\sum_{k'} A_{\text{sphere}}(q,R_k) A_{\text{sphere}}(q,R_{k'}) e^{-i\mathbf{q}^T\mathbf{r}_{kk'}} \\
&\quad + \sum_{l}\sum_{l'} A_{\text{core-shell}}(q,R_l,t_l) A_{\text{core-shell}}(q,R_{l'},t_{l'}) e^{-i\mathbf{q}^T\mathbf{r}_{ll'}} \\
&\quad + A_{\mathrm{shell}}(q,R,t) \sum_{k} A_{\text{sphere}}(q,R_k)\left(e^{i\mathbf{q}^T\mathbf{r}_k} + e^{-i\mathbf{q}^T\mathbf{r}_k}\right) \\
&\quad + A_{\mathrm{shell}}(q,R,t) \sum_{l} A_{\text{core-shell}}(q,R_l,t_l)\left(e^{i\mathbf{q}^T\mathbf{r}_l} + e^{-i\mathbf{q}^T\mathbf{r}_l}\right) \\
&\quad + \sum_{k}\sum_{l} A_{\text{sphere}}(q,R_k) A_{\text{core-shell}}(q,R_l,t_l)\left(e^{i\mathbf{q}^T\mathbf{r}_{kl}} + e^{-i\mathbf{q}^T\mathbf{r}_{kl}}\right),
\end{aligned} \tag{184}$$

where $\mathbf{r}_{kk'} = \mathbf{r}_k - \mathbf{r}_{k'}$ and analogously for the other double indexed $\mathbf{r}$'s. A scattering experiment captures a time average of the freely moving complex cluster, or equivalently an orientational ensemble average of it. Since $\left\langle e^{-i\mathbf{q}^T\mathbf{r}}\right\rangle_{\Omega} = \sin(qr)/(qr)$, cf. Eq. (18), it follows for the orientationally averaged intensity form factor of a single complex cluster

$$\begin{aligned}
I_{\mathrm{cc}}(q) &= \left\langle I_{\mathrm{cc}}(\mathbf{q})\right\rangle_{\Omega} \\
&= A_{\mathrm{shell}}^{2}(q,R,t) \\
&\quad + \sum_{k}\sum_{k'} A_{\text{sphere}}(q,R_k) A_{\text{sphere}}(q,R_{k'}) \frac{\sin(qr_{kk'})}{qr_{kk'}} \\
&\quad + \sum_{l}\sum_{l'} A_{\text{core-shell}}(q,R_l,t_l) A_{\text{core-shell}}(q,R_{l'},t_{l'}) \frac{\sin(qr_{ll'})}{qr_{ll'}} \\
&\quad + 2A_{\mathrm{shell}}(q,R,t) \sum_{k} A_{\text{sphere}}(q,R_k) \frac{\sin(qr_k)}{qr_k} \\
&\quad + 2A_{\mathrm{shell}}(q,R,t) \sum_{l} A_{\text{core-shell}}(q,R_l,t_l) \frac{\sin(qr_k)}{qr_l} \\
&\quad + 2\sum_{k}\sum_{l} A_{\text{sphere}}(q,R_k) A_{\text{core-shell}}(q,R_l,t_l) \frac{\sin(qr_{kl})}{qr_{kl}},
\end{aligned} \tag{185}$$

which describes isotropic scattering. For a number of N clusters the mean isotropic intensity form factor reads

$$\overline{I}_{\mathrm{cc}}(q) = \frac{1}{N}\sum_{n=1}^{N} I_{\mathrm{cc},n}(q), \tag{186}$$

where $I_{\mathrm{cc},n}(q)$ is the formula given in Eq. (185), but applied to the n-th cluster. Eq. (186) is identical to Eq. (134), but here the additional parameter symbols are suppressed in order to keep the notation uncluttered.

In the program code it is important to check whether the formula is correctly implemented or not. This has been done by comparing the implemented scattering formula for $q \to 0$ with the forward intensity (cf. section 2.1) of Eq. (186), i.e., determining $\lim_{q\to 0} \overline{I}_{\mathrm{cc}}(q)$, which is easily accomplished by applying the results from Eq. (11), (181), and (183) in Eq. (186).

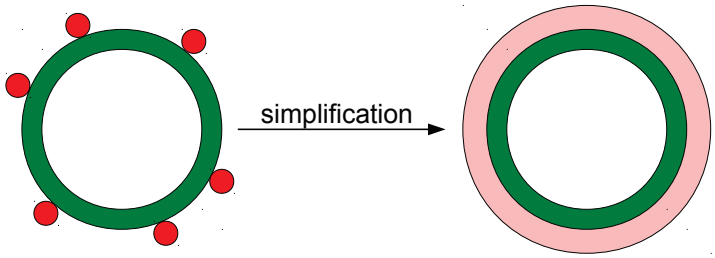


Figure 26: Model simplification. As a model simplification the decorated vesicle can be considered to be a vesicle which is surrounded with a second shell. The second shell might be described by a constant SLD profile as well as by a non constant profile — see text for further explanations.

4.3.1. Analytical Models

As a special case of a complex cluster system only decorated vesicles are considered in the following. In this case the complex cluster intensity Eq. (185), respectively Eq. (186), simplifies, since all core-shell amplitude terms are equal to zero.

Since MC simulations can be quite compute-intensive the question arises if a MC simulated decorated vesicle model can be described by a simpler, analytical model as well. Following models are considered:

Vesicle-(homogeneous particle shell) model: It is considered that the vesicles are monodisperse and that the particles on the vesicle shell can be described by a second shell, a 'homogeneous particle' shell, having a constant SLD profile — see Fig. 26 and 27a. It is considered that the second shell has the thickness $2R_p$ and consists of particle material of K particles that is homogeneously distributed in the hydrated shell. The intensity form factor of this model reads

$$\begin{aligned} I_{\text{vesicle + homog. shell}}(q) = \big[&A_{\text{shell}}(q, R, t; \text{SLD}_{\text{vesicle shell}}, \text{SLD}_{\text{matrix}}) \\ &+ A_{\text{shell}}(q, R+t, 2R_p; \text{SLD}_{\text{particle shell}}, \text{SLD}_{\text{matrix}})\big]^2, \end{aligned} \tag{187}$$

where the used amplitude form factors are given by the shell form factor in Eq. (180), but here the SLDs are explicitly written: $\text{SLD}_{\text{matrix}}$ is the SLD of the matrix and $\text{SLD}_{\text{vesicle shell}}$ is the SLD of the vesicle shell that has an inner radius R and an outer radius $R+t$. The only unknown parameter is the SLD of the particle shell, $\text{SLD}_{\text{particle shell}}$. It can be calculated as follows: $v_{ps} = 4\pi/3\big[(R+t+2R_p)^3 - (R+t)^3\big]$ is the volume of the homogeneous particle shell, and the total particle volume in this shell is $v_p = Kv_{p,1}$, where $v_{p,1} = 4\pi R_p^3/3$, and therefore the volume of hydration is $v_h = v_{ps} - v_p$. Consequently, the SLD of the homogeneous particle shell reads

$$\text{SLD}_{\text{particle shell}} = \frac{\text{SLD}_{\text{particle}} \cdot v_p + \text{SLD}_{\text{matrix}} \cdot v_h}{v_p + v_h}. \tag{188}$$

Vesicle-(particle profile shell) model: Similar to the previous model this model consists of a particle shell on the vesicle, but this time the shell has a non constant SLD profile that aims at reproducing the radial average SLD corresponding to monodisperse spheres laying on a sphere — see Fig. 26 and 27b. The SLD profile of this shell is defined for $r \in [R_{\text{ves}}, R_{\text{ves}} + 2R_p]$, where $R_{\text{ves}} = R + t$, and at position r the SLD depends on the

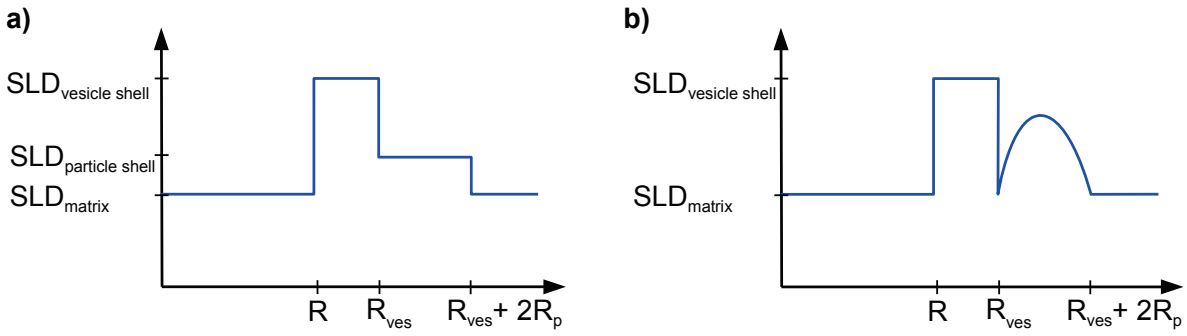


Figure 27: a) SLD profile of the vesicle-(homogeneous particle shell) model. b) SLD profile of the vesicle-(particle profile shell) model. $R_{\text{ves}} = R + t$.

amount of particle surface that is given by the intersection of a sphere having radius r with the particles, cf. Fig. 28.

The volume of one particle, which is situated at the top of a vesicle (see Fig. 28), can be written as

$$\begin{aligned} v_{p,1} &= \int_{R_{\text{ves}}}^{R_{\text{ves}}+2R_p} \int_0^{2\pi} \int_0^{\vartheta(r)} r^2 \sin(\vartheta)\, \mathrm{d}\vartheta\, \mathrm{d}\varphi\, \mathrm{d}r \\ &= \int_{R_{\text{ves}}}^{R_{\text{ves}}+2R_p} 2\pi r^2 \big(1 - \cos(\vartheta(r))\big)\, \mathrm{d}r \\ &= \int_{R_{\text{ves}}}^{R_{\text{ves}}+2R_p} 2\pi r^2 \left(1 - \frac{R_p^2 - (R_{\text{ves}} + R_p)^2 - r^2}{-2r(R_{\text{ves}} + R_p)}\right) \mathrm{d}r = \frac{4\pi}{3} R_p^3, \end{aligned} \tag{189}$$

where the law of cosines, $\cos(\vartheta(r)) = \big[R_p^2 - (R_{\text{ves}} + R_p)^2 - r^2\big] / \big[-2r(R_{\text{ves}} + R_p)\big]$, is used. The integrand in the last line of Eq. (189)

$$s_{p,1}(r) = 2\pi r^2 \left(1 + \frac{R_p^2 - (R_{\text{ves}} + R_p)^2 - r^2}{2r(R_{\text{ves}} + R_p)}\right) \tag{190}$$

is the intersection of a sphere with radius r with the particle volume, i.e., it defines a sphere segment within the particle. Accordingly, if the shell consists of K identical particles and the remaining shell volume consists of matrix, the SLD profile of the particle shell changes according to

$$\text{SLD}_{\text{particle shell}}(r) = \frac{\text{SLD}_{\text{particle}} \cdot K s_{p,1}(r) + \text{SLD}_{\text{matrix}} \cdot (4\pi r^2 - K s_{p,1}(r))}{4\pi r^2}. \tag{191}$$

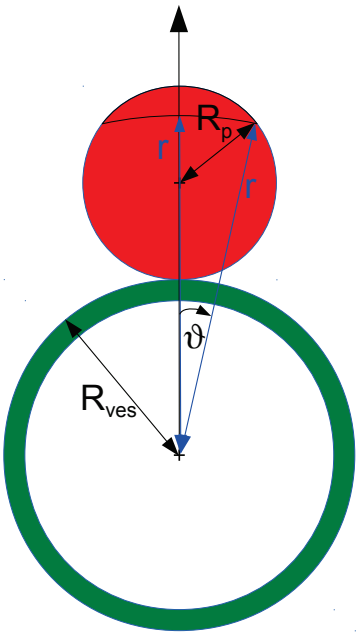


Figure 28: A particle situated on the top of a vesicle. ϑ is the polar angle. As depicted, the radius r defines a sphere segment within the particle.

Hence, the amplitude form factor is

$$\begin{aligned} A_{\text{profile shell}}(q, R_{\text{ves}}, R_p, K) &= 4\pi \int_{R_{\text{ves}}}^{R_{\text{ves}}+2R_p} (\text{SLD}_{\text{particle shell}}(r) - \text{SLD}_{\text{matrix}}) \\ &\qquad \times r^2 \frac{\sin(qr)}{qr}\, \mathrm{d}r \\ &= (\text{SLD}_{\text{particle}} - \text{SLD}_{\text{matrix}}) \int_{R_{\text{ves}}}^{R_{\text{ves}}+2R_p} K s_{p,1}(r) \frac{\sin(qr)}{qr}\, \mathrm{d}r \\ &= K(\text{SLD}_{\text{particle}} - \text{SLD}_{\text{matrix}}) \frac{-4\pi}{q^4(R_{\text{ves}} + R_p)} \\ &\quad \times \Big(qR_p \sin(qR_{\text{ves}}) \cos^2(qR_p) \\ &\quad + qR_p \cos(qR_{\text{ves}}) \sin(qR_p) \cos(qR_p) \\ &\quad + \cos(qR_{\text{ves}}) \cos^2(qR_p) - \cos(qR_{\text{ves}}) \\ &\quad - \sin(qR_{\text{ves}}) \sin(qR_p) \cos(qR_p) \Big), \end{aligned} \tag{192}$$

and the forward amplitude is $A_{\text{profile shell}}(0) = K(\text{SLD}_{\text{particle}} - \text{SLD}_{\text{matrix}}) \frac{4\pi}{3} R_p^3$.

Finally, the intensity of the vesicle decorated with a SLD particle profile shell is

$$\begin{aligned} I_{\text{vesicle + shell}}(q) = \big[& A_{\text{shell}}(q, R, t; \text{SLD}_{\text{vesicle shell}}, \text{SLD}_{\text{matrix}}) \\ & + A_{\text{profile shell}}(q, R + t, R_p, K; \text{SLD}_{\text{particle}}, \text{SLD}_{\text{matrix}}) \big]^2. \end{aligned} \tag{193}$$

Vesicle-(two particles correlation) model: The previous two models are not capable to capture any particle correlations, whereas in this model the interaction of two particles is taken into account. The contribution of the correlation of two particles can be obtained

by placing one particle at the top of a vesicle while a second particle is considered to be randomly distributed over the vesicle, see Fig. 29. Consider K identical particles on a vesicle, then Eq. (185) gives

$$\begin{aligned} I_{\text{vesicle, 2-particle corr.}} &= A^2_{\text{shell}}(q,R,t) \\ &\quad + KA^2_{\text{sphere}}(q,R_p) + A^2_{\text{sphere}}(q,R_p)\sum_{k}^{K}\sum_{k'\neq k}^{K}\left\langle \frac{\sin(qr_{kk'})}{qr_{kk'}}\right\rangle \\ &\quad + 2KA_{\text{shell}}(q,R,t)A_{\text{sphere}}(q,R_p)\frac{\sin(q(R_{\text{ves}}+R_p))}{q(R_{\text{ves}}+R_p)}, \end{aligned} \tag{194}$$

where the bracket indicates a distance average between particle k and particle k', and $R_{\text{ves}} = R+t$. Using spherical coordinates, the distance $r_{kk'}$ can be written as a function of the polar angle ϑ, i.e., $r_{kk'} = r_{kk'}(\vartheta)$. Let φ the azimuthal angle, then

$$\begin{aligned} r_{kk'}(\vartheta) &= \left\| \begin{pmatrix} 0 \\ 0 \\ R_{\text{ves}}+R_p \end{pmatrix} - \begin{pmatrix} (R_{\text{ves}}+R_p)\sin(\vartheta)\cos(\varphi) \\ (R_{\text{ves}}+R_p)\sin(\vartheta)\sin(\varphi) \\ (R_{\text{ves}}+R_p)\cos(\vartheta) \end{pmatrix} \right\| \\ &= \sqrt{(R_{\text{ves}}+R_p)^2\sin^2(\vartheta) + (R_{\text{ves}}+R_p)^2(1-\cos(\vartheta))^2} \\ &= \sqrt{2(R_{\text{ves}}+R_p)^2(1-\cos(\vartheta))} = 2(R_{\text{ves}}+R_p)\sin\left(\frac{\vartheta}{2}\right). \end{aligned} \tag{195}$$

Hence, a single two particles correlation is

$$\begin{aligned} \left\langle \frac{\sin(qr_{kk'})}{qr_{kk'}}\right\rangle &= \int_0^{\pi} \frac{\sin(qr_{kk'}(\vartheta))}{qr_{kk'}(\vartheta)}\frac{\sin(\vartheta)}{2}\,\mathrm{d}\vartheta \\ &= \frac{1-\cos^2(q(R_{\text{ves}}+R_p))}{q^2(R_{\text{ves}}+R_p)^2} \\ &= \frac{\sin^2(q(R_{\text{ves}}+R_p))}{q^2(R_{\text{ves}}+R_p)^2}, \end{aligned} \tag{196}$$

and is used in Eq. (194) in order to calculate the correlations, i.e., the second term in the second line becomes

$$A^2_{\text{sphere}}(q,R_p)\sum_{k}^{K}\sum_{k'\neq k}^{K}\left\langle \frac{\sin(qr_{kk'})}{qr_{kk'}}\right\rangle = K(K-1)A^2_{\text{sphere}}(q,R_p)\frac{\sin^2(q(R_{\text{ves}}+R_p))}{q^2(R_{\text{ves}}+R_p)^2}. \tag{197}$$

4.4. Evaluations

In the next subsection, a comparison between the different analytical models given in section 4.3.1 with MC simulated decorated vesicle models is given. In the succeeding subsection, scattering simulations are analyzed for the considered decorated vesicle systems (DPPC vesicles, decorated with silica nanoparticles). The final subsection analyzes expected scattering intensities of more complex systems consisting of DOPC vesicles that are decorated and/or contain internalized silica nanoparticles.

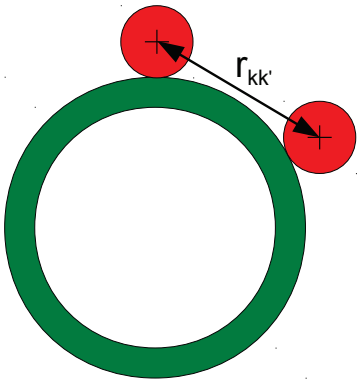


Figure 29: Two particles correlation. It is considered that the k-th particle is situated at the top of the vesicle and the other particle k' is randomly placed on the vesicle.

4.4.1. Analytical Models vs. MC Simulations

A model system is considered that comprises vesicles consisting of a DPPC bilayer decorated with silica (nano-)particles. The vesicles are monodisperse as well as the particles, and the number of particles per vesicle is fixed. The vesicle core radius is $R = 37.9\,\mathrm{nm}$ and the shell thickness is $t = 4.4\,\mathrm{nm}$, the particle radius is $R_p = 8.36\,\mathrm{nm}$. The number of particles per vesicle is $K \in \{2, 5, 10, 20, 30, 60\}$. The intensity is calculated for the case of X-ray scattering (CuKα characteristic radiation), i.e., $\mathrm{SLD_{DPPC}} = 10.1 \cdot 10^{-4}\,\mathrm{nm^{-2}}$, and $\mathrm{SLD_{SiO_2}} = 18.9 \cdot 10^{-4}\,\mathrm{nm^{-2}}$. The matrix is H_2O, having $\mathrm{SLD_{H_2O}} = 9.46 \cdot 10^{-4}\,\mathrm{nm^{-2}}$, i.e., the vesicle shell is nearly contrast matched.

In the following the surface fraction of nanoparticles on a vesicle φ'_{pv} is defined as the total particle profile area per vesicle surface area:

$$\varphi'_{pv} = \frac{K\pi R_p^2}{4\pi(R+t)^2}. \tag{198}$$

For different K the scattering intensities of the three different analytical models of decorated vesicles, given in section 4.3.1, are plotted together with the scattering intensities of MC simulations (where a hard sphere potential energy is used, Eq. (124)) in Fig. 30. Clearly for $K \lesssim 10$ ($\varphi'_{pv} \lesssim 10\,\%$), the vesicle-(two particles correlation) model gives an excellent match with the MC simulation one. The two different vesicle-shell models do not give a good match for these low surface fractions, except for the low q-regime ($q \lesssim 0.05\,\mathrm{nm^{-1}}$). However, for higher surface fractions the vesicle-(two particles correlation) model does not match well the intensity of the MC simulated one in the mid q-regime ($0.05\,\mathrm{nm^{-1}} \lesssim q \lesssim 0.3\,\mathrm{nm^{-1}}$), whereas the quality of the match of the intensity of the vesicle-shell models with the MC simulated one increases. Of course, the two different vesicle-shell models can never describe properly the high q-regime, since the intensity in this regime is described by the intensity form factor of the particles (spheres). The differences between the two vesicle-shell models are negligible in the mid q-regime.

In Fig. 31 results of MC simulations are depicted, where the intensity of decorated vesicles having a fixed number of particles is compared with the intensity of decorated vesicles having the same average number of particles, but the number of particles is described by a Poisson distribution. Therein, the average surface fraction of nanoparticles per vesicle is defined as

$$\langle \varphi'_{pv} \rangle = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K_n} \frac{K_n \pi R_{p,k}^2}{4\pi(R_n + t_n)^2}, \tag{199}$$

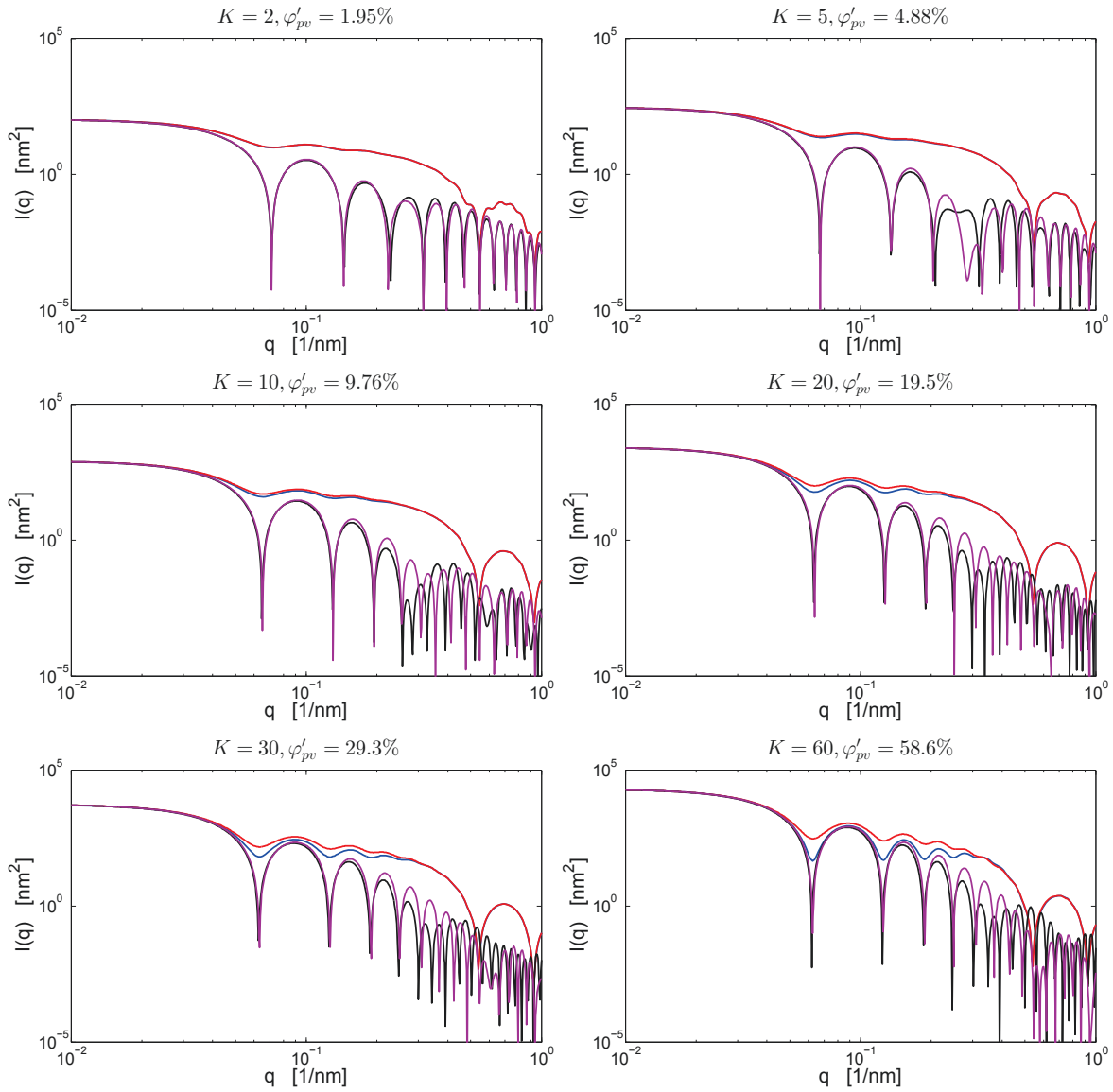


Figure 30: Intensity form factors of models of decorated vesicles. K is the number of particles on a vesicle, and φ'_{pv} is the surface fraction of particles on a vesicle. Blue intensity curves: MC simulation of decorated vesicles. Red intensity curves: vesicle-(two particles correlation) model. Black intensity curves: vesicle-(homogeneous particle shell) model. Magenta intensity curves: vesicle-(particle profile shell) model. The blue curve is covered by the red curve in the diagrams of the first row.

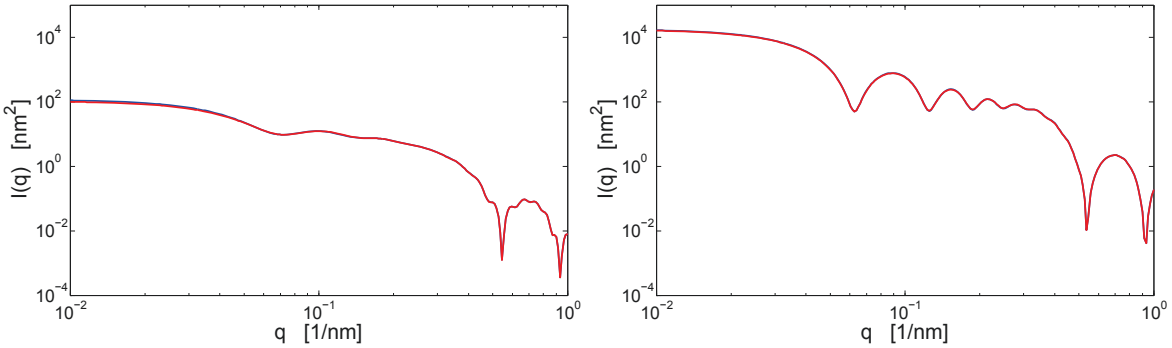


Figure 31: Intensities of MC simulations of decorated vesicles. Blue intensity curves: the vesicles are decorated by particles, where the number of particles is Poisson distributed. Red intensity curves: the vesicles are decorated with a fixed number of particles. Left diagram: there are two particles per vesicle in average (i.e., $\varphi'_{pv} = \langle \varphi'_{pv} \rangle = 1.96\,\%$). Right diagram: there are 55 particles per vesicle in average (i.e., $\varphi'_{pv} = \langle \varphi'_{pv} \rangle = 53.8\,\%$). The blue curves are nearly perfectly matched by the red curves, except that the low q-intensity is slightly higher for the blue curves.

where this formula also takes a possible size distribution of nanoparticles into account, which is used in the next subsection. Two limiting cases are considered: (i) 2 particles in average per vesicle and (ii) 55 particles in average per vesicle[41]. As depicted in Fig. 31 the intensity differences are negligible, only the intensity in the low q-regime is slightly higher for the case where the number of particles per vesicle is Poisson distributed compared to the case of having a fixed number of particles. Considering that the vesicle shell is contrast matched, this effect can be mathematically described by applying Eq. (185),(186), which then becomes for $q \to 0$

$$\begin{aligned} \lim_{q\to 0} \overline{I}_{cc}(q) &= \lim_{q\to 0} \left(A^2_{\text{sphere}}(q, R_p) \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K_n} \sum_{k'=1}^{K_n} \frac{\sin(q r_{kk',n})}{q r_{kk',n}} \right) \\ &= A^2_{\text{sphere}}(0, R_p) \frac{1}{N} \sum_{n=1}^{N} K_n^2 \\ &= (\text{SLD}_{\text{particle}} - \text{SLD}_{\text{matrix}})^2 v^2_{\text{sphere}}(R_p) \langle K_n^2 \rangle_n, \end{aligned} \tag{200}$$

where $\langle K_n^2 \rangle_n := \frac{1}{N} \sum_{n=1}^{N} K_n^2$ is the second moment, which can be expressed by the average $\langle K_n \rangle_n$ and the variance $\langle (K_n - \langle K_n \rangle_n)^2 \rangle_n \geq 0$ as $\langle K_n^2 \rangle_n = \langle K_n \rangle_n^2 + \langle (K_n - \langle K_n \rangle_n)^2 \rangle_n$, and thus showing that if the average number of particles per vesicle stays fixed, any particle number distribution having a variance larger than zero will increase the intensity compared to the case of having a Dirac distribution with the number of particles per vesicle equal to $\langle K_n \rangle_n$.

4.4.1.1. Results

The evaluations show that the considered decorated (silica nanoparticles) DPPC vesicle system is well described by the vesicle-(two particles correlation) model for surface fractions of $\varphi'_{pv} \lesssim 10\,\%$, and considering X-ray contrast conditions. For $\varphi'_{pv} \gtrsim 60\%$ the vesicle-(homogeneous particle shell) model and the vesicle-(particle profile shell) one give a better

[41] Less than 60 particles in average per vesicle are taken, otherwise problems would occur with placing the particles on a vesicle (too high surface fraction of particles on a vesicle).

model description in the mid q regime. Last but not least, the effect of having a Poisson distribution of particles instead of taking a Dirac distribution only produces a minor intensity difference at the low q-regime.

		DP_0	$DP_{0.7}$	$DP_{0.85}$	$DP_{1.0}$	$DP_{1.5}$	$DP_{3.0}$
φ_p		0	$3.18 \cdot 10^{-4}$	$3.87 \cdot 10^{-4}$	$4.55 \cdot 10^{-4}$	$6.82 \cdot 10^{-4}$	$13.7 \cdot 10^{-4}$
HS	$\langle$#NP/vesicle$\rangle$	0	12.1	15.1	17.9	26	52.7
	$\langle \varphi'_{pv} \rangle$ /%	0	11.3	14.1	16.5	24.8	50.2
C	$\langle$#NP/vesicle$\rangle$	0	12.5	15.1	17.7	27.1	53.5
	$\langle \varphi'_{pv} \rangle$ /%	0	12	14.4	16.9	25.1	50.1

Table 5: Values of φ_p (volume fraction of nanoparticles) used to create decorated DPPC vesicles. DP_x defines the simulation set names (x indicates the weight percentage that ranges from 0-3 wt%). 'HS' means hard sphere potential energy, see Eq. (124), and 'C' means strong repulsive Coulomb potential energy, see Eq. (125) and description in the text thereafter. 500 decorated vesicles are generated and corresponding scattering intensities are calculated (see Fig. 32) — according to algorithm 2. From the simulations the following two quantities are obtained: (*i*) $\langle$#NP/vesicle$\rangle$: the average number of nanoparticles per vesicle, (*ii*) $\langle \varphi'_{pv} \rangle$: the average of the surface fraction (here in percentages) of nanoparticles on a vesicle, cf. Eq. (199). Simulating more decorated vesicles does not change the scattering intensity curves significantly.

4.4.2. MC Simulations of Decorated DPPC Vesicles

In this section MC simulations are presented for systems consisting of DPPC vesicles that are decorated by silica nanoparticles (NP). These systems have already been analyzed in some depth in [104, 107] (via SANS, cryo-TEM, dynamic and static light scattering, differential scanning calorimetry, zeta potential measurements). Here now, MC simulations are performed according to algorithm 2 in order to get detailed insights into the scattering intensity behavior of such systems. The simulations have been performed with a hard sphere potential energy, Eq. (124), as well as with a strong repulsive Coulomb potential energy, cf. Eq. (125) and text thereafter. Moreover, X-ray contrast conditions are assumed, hence the scattering intensity reflects more strongly (compared to SANS) the form factor of the particles and their interactions, i.e., their structure factor.

According to [104], it is considered that the vesicle core radius distribution has a mean $\langle R \rangle = 37.9\,\text{nm}$ and a polydispersity $p_R = 0.25$, and here the distribution is assumed to be a Gaussian one. The vesicle shell thickness distribution is considered to have a mean $\langle t \rangle = 4.4\,\text{nm}$ and a polydispersity $p_t = 0.1$ and is assumed to be Gaussian, too. The radius distribution of the particles is also assumed to be Gaussian with mean radius $\langle R_p \rangle = 8.36\,\text{nm}$ and polydispersity $p_{R_p} = 0.14$. The total volume fraction of shell material is taken as $\varphi_{\text{DPPC}} = 9.39 \cdot 10^{-4}$, and different volume fractions of nanoparticles φ_p are considered, which are listed in table 5. It is assumed that the number of particles on a vesicle is Poisson distributed, and changing φ_p changes the average number of particles per vesicle, see Eq. (165),(166). The SLDs of the considered substances are already given in section 4.4.1.

MC simulations have been done according to table 5, which additionally contains parameter results obtained from the simulations (the average number of nanoparticles per vesicle, and the average surface fraction of nanoparticles on a vesicle). Fig. 32 shows the results of the MC simulations, and Fig. 33 shows a plot of all contributions to the overall intensity of the sample system $DP_{3.0}$ having strong repulsive interactions. From these simulations it is clearly seen that there is an oscillation in the mid q-regime ($q \approx 0.05 - 0.07\,\text{nm}^{-1}$), which becomes more significant with (*i*) the increase of the volume fraction of nanoparticles φ_p as well as with (*ii*) the presence of repulsive forces. Moreover, the presence of (strongly) repulsive forces create a bump at $q \approx 0.13 - 0.3\,\text{nm}^{-1}$. This bump becomes more significant with increasing

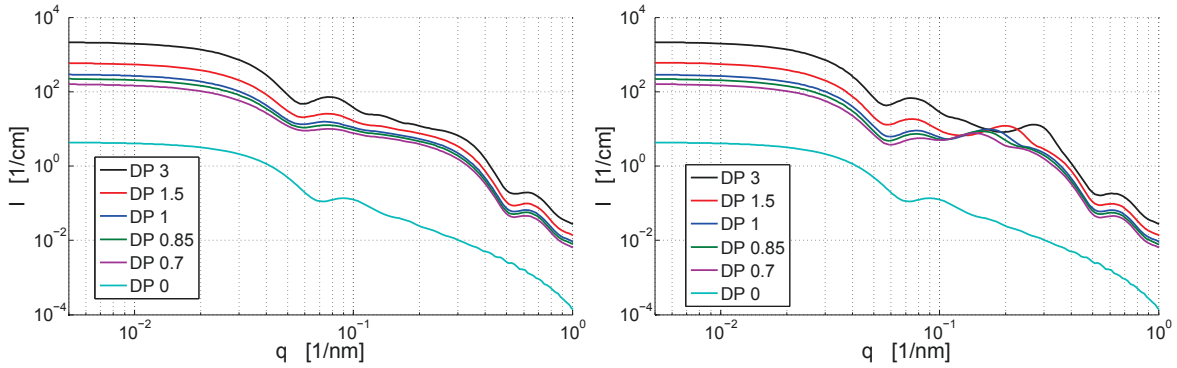


Figure 32: Simulation results of decorated vesicles according to table 5 and description in the text. Left diagram: hard sphere potential is used. Right diagram: a strong repulsive Coulomb potential is used. Clearly, an increase of the volume fraction of nanoparticles φ_p lifts the scattering curve, especially it increases the forward intensity. The presence of strongly repulsive forces produces a bump at $q \approx 0.13 - 0.3\,\mathrm{nm}^{-1}$, which becomes more significant with an increase of φ_p.

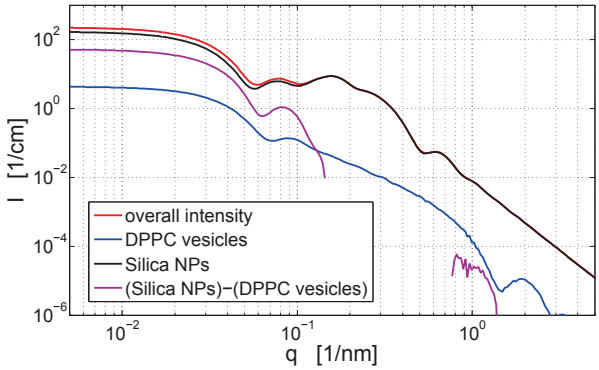


Figure 33: Plot of all amplitude-amplitude contributions to the overall scattering intensity of sample system $\mathrm{DP}_{3.0}$ having strong repulsive interactions between the particles (i.e., a strong repulsive Coulomb potential energy is used for the simulation). The red curve is the overall intensity of all amplitude-amplitude contributions in Eq. (185),(186). Note, the cross amplitude form factor is partially negative.

particle volume fraction φ_p, and moves from $0.13\,\mathrm{nm}^{-1}$ to $0.3\,\mathrm{nm}^{-1}$ as φ_p increases from 0 up to $13.7 \cdot 10^{-4}$. This shift to higher q-values is due to the fact that the interparticle distances of the particles on a vesicle decrease as φ_p increases. Last but not least, the forward scattering increases with φ_p — cf. Eq. (24) and text thereafter.

4.4.2.1. Results

For a possible SAXS data evaluation of the considered systems (DPPC vesicles decorated by silica nanoparticles), the average number of particles per vesicle can be deduced from the forward scattering[42]. Evidence of repulsive forces between particles are mainly encoded in a bump at $q \approx 0.13 - 0.3\,\mathrm{nm}^{-1}$, and are also slightly encoded in an oscillation at $q \approx 0.05 - 0.07\,\mathrm{nm}^{-1}$.

[42] E.g., the (average) number of particles can be estimated by fitting the formula in Eq. (194) or in Eq. (200) to the experimentally observed forward intensity.

4.4.3. MC Simulations of Decorated DOPC Vesicles with Internalized Particles

In this section different MC simulations are presented of systems consisting of DOPC vesicles that are decorated and/or contain internalized silica nanoparticles. Such systems have already been analyzed in some depth in [104, 9, 106] (via SANS, and cryo-TEM, dynamic and static light scattering, fluorescence correlation spectroscopy). Here now, MC simulations are performed in order to get detailed insights into the scattering intensity behavior of such systems. X-ray contrast conditions are assumed, hence the scattering intensity reflects more strongly (compared to SANS) the form factor of the particles and their interactions, i.e., their structure factor.

Two different systems are considered, which consist of

(A) a DOPC volume fraction of $\varphi_{\mathrm{DOPC}} = 9.983 \cdot 10^{-3}$ and a silica nanoparticle volume fraction of $\varphi_p = 1.093 \cdot 10^{-3}$,

(B) a DOPC volume fraction of $\varphi_{\mathrm{DOPC}} = 9.984 \cdot 10^{-3}$ and a silica nanoparticle volume fraction of $\varphi_p = 2.194 \cdot 10^{-3}$.

In [104] it was inferred that after mixing the above given quantities of DOPC and silica nanoparticles, silica nanoparticles attach to the surface of the DOPC vesicles, which prevent the vesicles from coalescence. Then, particles are gradually internalized by the vesicles whereby the particles remove a part of the vesicle shell in order to encapsulate themselves with it forming a supported lipid bilayer [91, 106] — cf. Fig. 20. As a consequence of the reduced amount of particles on the vesicles, the vesicles coalesce and grow in size. According to [104] it is considered that the initial vesicle core radius distribution has a mean $\langle R \rangle = 43.4\,\mathrm{nm}$ and a polydispersity of $p_R = 0.25$, and here the distribution is assumed to be a Gaussian one. Furthermore, the vesicle shell thickness distribution is considered to be Gaussian with mean $\langle t \rangle = 4.4\,\mathrm{nm}$ and polydispersity $p_t = 0.1$. The radius distribution of the particles is Gaussian too with mean radius $\langle R_p \rangle = 8.36\,\mathrm{nm}$ and polydispersity $p_{R_p} = 0.14$ [104].

The SLDs of water and of the silica nanoparticles are given in section 4.4.1, and $\mathrm{SLD}_{\mathrm{DOPC}} = 9.4 \cdot 10^{-4}\,\mathrm{nm}^{-2}$.

MC simulations have been done with a hard sphere potential as well as with a strong repulsive Coulomb potential according to tables 6 and 7 and the quantities given above. Fig. 34 shows the simulation results of systems (A) and (B), and obtained simulation parameters are listed in tables 6 and 7, too.

System (A): If a hard sphere potential is used, there is only a slight structure factor peak (at ca. $0.07\,\mathrm{nm}^{-1}$) if all particles are on the surface (sample '0 days, 0 % int.'). All other simulations of the same system do not show an oscillation. However, if a strong repulsive Coulomb potential is used (only for the particles on the vesicles; the particles in the vesicles are encapsulated with shell material, hence screened, and therefore only a hard sphere interaction is considered for them), a deep oscillation is produced for sample '0 days, 0 % int.' (at ca. $0.05\,\mathrm{nm}^{-1}$) and the structure factor peak is amplified and moved[43] to ca. $0.08\,\mathrm{nm}^{-1}$. The sample curves '5 days, 30 % int.' and '8 days, 70 % int.' show only a slight 'wobbling' at $q \approx 0.04 - 0.1\,\mathrm{nm}^{-1}$.

System (B): The simulations show the same effects as discussed for system (A), but here, the features in the curves are amplified due to the fact that there are approximately two times the number of nanoparticles on the vesicles. Moreover, the structure factor peaks move to higher

[43] The peak moved to a higher value (compared to the hard sphere potential energy case), whereas the average particle distance increased (hence the peak position should be shifted to a lower q value), as a result of the influence of the adjacent deep oscillation at ca. $0.05\,\mathrm{nm}^{-1}$.

q values, since the interparticle distances decrease with an increase of the particle volume fraction. E.g., considering the sample '0 days, 0 % int.' for the strong Coulomb potential, the structure factor peak moves from $q \approx 0.08\,\mathrm{nm}^{-1}$ for $\varphi_p = 1.093 \cdot 10^{-3}$ to $q \approx 0.12\,\mathrm{nm}^{-1}$ for $\varphi_p = 2.194 \cdot 10^{-3}$. Last but not least, the forward scattering increases with φ_p (cf. Eq. (24) and text thereafter).

		0 days, 0 % int.	5 days, 30 % int.	8 days, 70 % int.	12 days, 100 % int.
	initial μ_R /nm	43.4	43.4	110	125
HS	$\langle\varphi_{pv}\rangle$ /%	0	4.65	3.16	4.47
	$\langle\varphi'_{pv}\rangle$ /%	3.15	2.37	1.14	0
	$\langle\#(\mathrm{NP})_{\mathrm{in}}/\mathrm{vesicle}\rangle$	0	1.25	17.7	32.4
	$\langle\#(\mathrm{NP})_{\mathrm{on}}/\mathrm{vesicle}\rangle$	4.19	2.93	7.54	0
	$\langle\#(\mathrm{NP})_{\mathrm{in}}/\#(\mathrm{NP})_{\mathrm{tot.}}\rangle$ /%	0	29.8	70.2	100
	determined μ_R /nm	43.6	42.4	102	111
C	$\langle\varphi_{pv}\rangle$ /%	0	4.91	3.2	identical to HS case
	$\langle\varphi'_{pv}\rangle$ /%	3.12	2.36	1.16	
	$\langle\#(\mathrm{NP})_{\mathrm{in}}/\mathrm{vesicle}\rangle$	0	1.26	17.8	
	$\langle\#(\mathrm{NP})_{\mathrm{on}}/\mathrm{vesicle}\rangle$	4.1	2.91	7.57	
	$\langle\#(\mathrm{NP})_{\mathrm{in}}/\#(\mathrm{NP})_{\mathrm{tot.}}\rangle$ /%	0	30.3	70.2	
	determined μ_R /nm	43.4	42.2	101	

Table 6: Simulation results of system (A) — DOPC vesicles ($\varphi_{\mathrm{DOPC}} = 9.983 \cdot 10^{-3}$) that are decorated and/or contain internalized nanoparticles ($\varphi_p = 1.093 \cdot 10^{-3}$). 2000 complex clusters are generated and corresponding scattering intensities are calculated (see Fig. 34) — according to algorithm 2. The top row defines the particle percentages that are internalized after a certain delay, as inferred in [104]. The second row contains the initially expected mean vesicle core radii μ_R if particles would be internalized without removing a part of the vesicle shell. However, it is considered that the particles remove a part of the vesicle shell to build their one shell encapsulation if they are internalized. Hence, the final vesicle core radii, which are found by the simulations, are smaller. These radii are given in the rows 'determined μ_R /nm'. The other rows give parameter values determined from the simulations. 'HS' means hard sphere potential energy, see Eq. (124), and 'C' means Coulomb potential energy, see Eq. (125). The Coulomb potential energy is only used for the particles on the surface of a vesicle in order to place them in such a way that they have a maximum distance to each other. The particles in the vesicles are considered to be encapsulated with vesicle shell material, therefore only a hard sphere interaction is considered for them. $\langle\varphi_{pv}\rangle$: average of the volume fraction (in percentages) of nanoparticles internalized in a vesicle. $\langle\varphi'_{pv}\rangle$: average of the surface fraction (in percentages) of the nanoparticles on a vesicle. $\langle\#(\mathrm{NP})_{\mathrm{in}}/\mathrm{vesicle}\rangle$: average number of internalized nanoparticles per vesicle. $\langle\#(\mathrm{NP})_{\mathrm{on}}/\mathrm{vesicle}\rangle$: average number of nanoparticles on a vesicle. $\langle\#(\mathrm{NP})_{\mathrm{in}}/\#(\mathrm{NP})_{\mathrm{tot.}}\rangle$: average of the fraction (in percentages) of internalized particles ($\#(\mathrm{NP})_{\mathrm{tot.}}$: total number of particles of a vesicle).

		0 days, 0 % int.	5 days, 30 % int.	8 days, 70 % int.	12 days, 100 % int.
	initial μ_R /nm	43.4	43.4	110	125
HS	$\langle\varphi_{pv}\rangle$ /%	0	8.64	8.45	14.1
	$\langle\varphi'_{pv}\rangle$ /%	6.48	5.34	2.81	0
	$\langle\#(\mathrm{NP})_{\mathrm{in}}/\mathrm{vesicle}\rangle$	0	2.42	36.8	68
	$\langle\#(\mathrm{NP})_{\mathrm{on}}/\mathrm{vesicle}\rangle$	8.37	6.02	15.6	0.02
	$\langle\#(\mathrm{NP})_{\mathrm{in}}/\#(\mathrm{NP})_{\mathrm{tot.}}\rangle$ /%	0	28.7	70.2	100
	determined μ_R /nm	43.3	40.5	93.5	97.1
C	$\langle\varphi_{pv}\rangle$ /%	0	8.32	8.75	identical to HS case
	$\langle\varphi'_{pv}\rangle$ /%	6.34	5.17	2.83	
	$\langle\#(\mathrm{NP})_{\mathrm{in}}/\mathrm{vesicle}\rangle$	0	2.41	35.5	
	$\langle\#(\mathrm{NP})_{\mathrm{on}}/\mathrm{vesicle}\rangle$	8.46	5.8	15.2	
	$\langle\#(\mathrm{NP})_{\mathrm{in}}/\#(\mathrm{NP})_{\mathrm{tot.}}\rangle$ /%	0	29.3	70.1	
	determined μ_R /nm	43.4	40.4	92.2	

Table 7: Simulation results of system (A) — DOPC vesicles ($\varphi_{\mathrm{DOPC}} = 9.983 \cdot 10^{-3}$) that are decorated and/or contain internalized nanoparticles ($\varphi_p = 2.194 \cdot 10^{-3}$). 2000 complex clusters are generated and corresponding scattering intensities are calculated (see Fig. 34) — according to algorithm 2. The top row defines the particle percentages that are internalized after a certain delay, as inferred in [104]. The second row contains the initially expected mean vesicle core radii μ_R if particles would be internalized without removing a part of the vesicle shell. However, it is considered that the particles remove a part of the vesicle shell to build their one shell encapsulation if they are internalized. Hence, the final vesicle core radii, which are found by the simulations, are smaller. These radii are given in the rows 'determined μ_R /nm'. The other rows give parameter values determined from the simulations. 'HS' means hard sphere potential energy, see Eq. (124), and 'C' means Coulomb potential energy, see Eq. (125). The Coulomb potential energy is only used for the particles on the surface of a vesicle in order to place them in such a way that they have a maximum distance to each other. The particles in the vesicles are considered to be encapsulated with vesicle shell material, therefore only a hard sphere interaction is considered for them. $\langle\varphi_{pv}\rangle$: average of the volume fraction (in percentages) of nanoparticles internalized in a vesicle. $\langle\varphi'_{pv}\rangle$: average of the surface fraction (in percentages) of the nanoparticles on a vesicle. $\langle\#(\mathrm{NP})_{\mathrm{in}}/\mathrm{vesicle}\rangle$: average number of internalized nanoparticles per vesicle. $\langle\#(\mathrm{NP})_{\mathrm{on}}/\mathrm{vesicle}\rangle$: average number of nanoparticles on a vesicle. $\langle\#(\mathrm{NP})_{\mathrm{in}}/\#(\mathrm{NP})_{\mathrm{tot.}}\rangle$: average of the fraction (in percentages) of internalized particles ($\#(\mathrm{NP})_{\mathrm{tot.}}$: total number of particles of a vesicle).

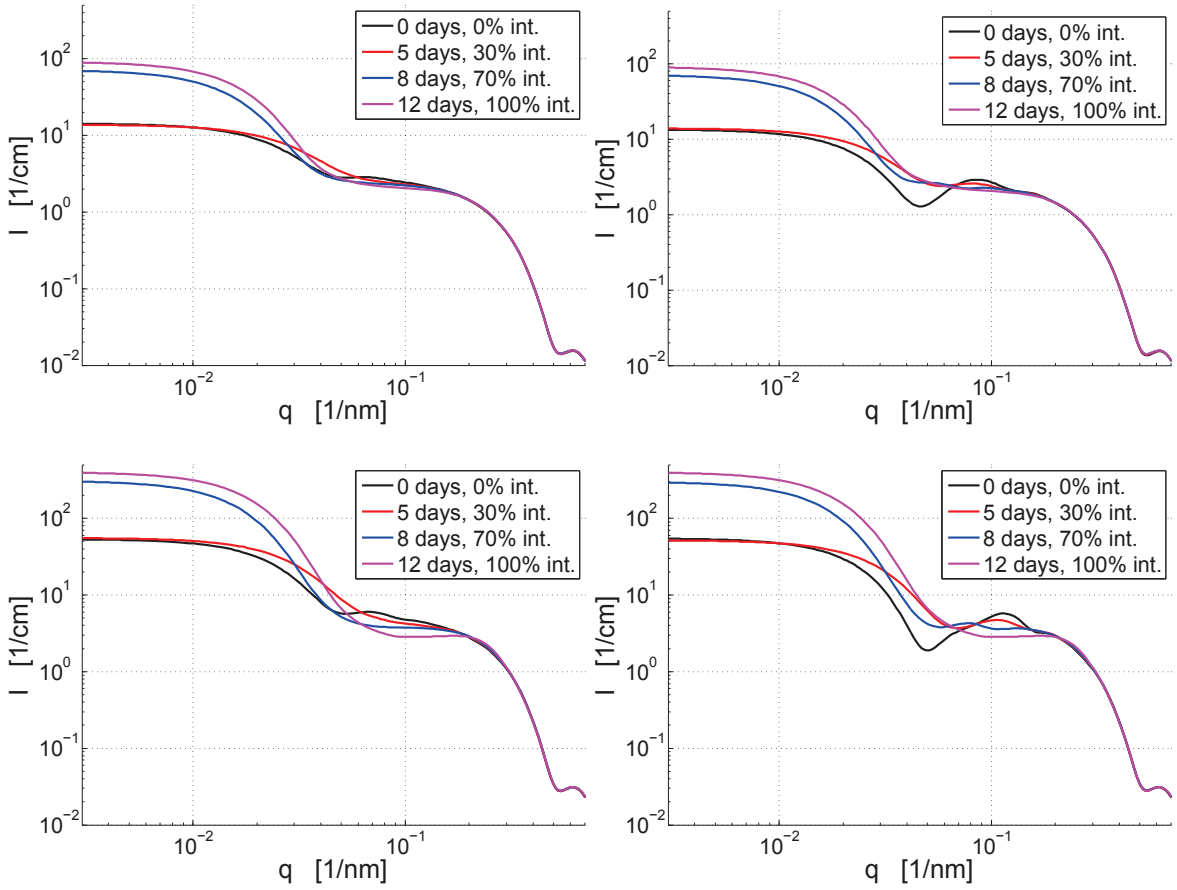


Figure 34: Simulation results of decorated vesicles with internalized particles according to tables 6 and 7 and description in the text. Top row: system (A), i.e., $\varphi_p = 1.09 \cdot 10^{-4}$. Bottom row: system (B), i.e., $\varphi_p = 2.19 \cdot 10^{-4}$. Left column: hard sphere potential energy is used, right column: strong repulsive Coulomb potential energy is used. The scattering intensity of only DOPC vesicles is not shown in the plots, since the forward intensity is already less than $0.05\,\mathrm{cm}^{-1}$. The increase of intensity at low q for curves '8 days, 70% int.' and '12 days, 100% int.' is a result of the increasing scattering mass of the complex clusters — resulting from the coalescence of the complex clusters, see text.

4.4.3.1. Results

For a possible SAXS data evaluation of the considered systems (DOPC vesicles decorated with nanoparticles and/or contain internalized nanoparticles), the average number of particles per vesicle can mainly be deduced from the forward scattering[44]. Shortly after mixing the systems, evidence of repulsive forces between particles on the vesicles is encoded in an oscillation at $q \approx 0.05\,\mathrm{nm}^{-1}$, and in a structure factor peak at $q \approx 0.07\,\mathrm{nm}^{-1}$ for $\varphi_p = 1.093 \cdot 10^{-3}$ and at $q \approx 0.12\,\mathrm{nm}^{-1}$ for $\varphi_p = 2.194 \cdot 10^{-3}$. At later times, evidence of strongly repulsive forces between the particles on the vesicles is not so strongly encoded anymore in the scattering intensities.

[44] E.g., the (average) number of particles per vesicle can be estimated by fitting the formula in Eq. (194) or in Eq. (200) to the experimentally observed forward intensity, if additionally the number of clusters is taken into account in the formulas.

4.5. Conclusion

Monte-Carlo simulations were successfully implemented that produce quickly (from seconds up to minutes for the more demanding cases) scattering patterns of rather complex hierarchical structures, harvesting the multiple cores of today's personal computers.

The cases of vesicles decorated with particles, encapsulated particles internalized by vesicles, and intermediate situations have been examined. Simulations were performed using physically realistic parameters, aiming to clarify the structure of systems already analyzed in [105] by probing them via SAXS, but because of technical reasons experimental SAXS data have yet to be acquired.

Simulations have shown which parameters can be deduced from features in the scattering intensities. In general, the two important features for vesicle-particles interactions are the intensities at low q (showing the increase in mass of the scatterer), and a structure factor resulting from the particles in a confined space. Moreover, it has been found that the considered decorated vesicle system can be well described by an analytical 'two particles-correlation' model for surface fractions of particles on a vesicle up to ca. 10 %.

5. Spontaneous Formation of Vesicles

The following coalescence study is an example of how the available information in a comprehensive scattering series is related to a few physical parameters. Moreover, this study shows which physical parameters can be determined from the scattering intensities reliably.

Vesicles are very interesting colloidal objects, since they are relevant as delivery systems [90, 29], and as models for biological membranes [12, 151]. Hence, the formation process of vesicles is very interesting, since this process defines the vesicular structures, e.g., the size and polydispersity of the vesicles.

Experiments show that vesicles can assemble spontaneously in diverse amphiphilic systems, e.g., in mixtures of cationic and anionic (catanionic) surfactants [81, 82, 8], in mixtures of mesoionic and anionic (mesoanionic) surfactants [175, 181], by means of addition of cosurfactant to surfactant solutions [44, 53], etc. Certain mixtures in particular catanionic and mesoanionic systems have been shown to often yield rather well defined unilamellar vesicles [81, 130, 23]. However, only in few cases the vesicles have been observed to be thermodynamically stable, while in the other cases the vesicles are metastable and age more or less quickly with time [73].

Understanding the pathway of the vesicle building process does not only rationalize the experimental observations but also allows to modify it if one wishes to control the vesicular structures. Therefore, many studies have been undertaken to follow the formation of vesicles as they are formed by mixing the corresponding surfactant solutions. Such experiments have for instance been done by means of a stopped-flow apparatus where the surfactants are mixed in a time of 1-5 ms and subsequently the structural evolution of the surfactant mixture can be followed by methods such as turbidimetry [110], static and dynamic light scattering [120, 93], SAXS [136] or SANS [57]. For the case of mixing anionic with cationic or mesoionic surfactants (concentrations above the critical micelle concentration (cmc)) these experiments have shown that typically as intermediate structures disk-like micelles are formed that grow in size until reaching a certain critical size at which they close to form unilamellar vesicles (vesiculation). This experimental finding and the growth of the initially present disk-like micelles have already some while ago been explained by means of a coalescence model, in which disks coalesce after collision processes [93, 178, 140, 92]. Beyond a certain size they become unstable due to the fact that their unfavorable rim energy, described by the line tension, becomes larger compared to the bending energy required to close the disk, thereby eliminating the rim by vesicle formation [140, 47, 95, 79]

In this work now, the experimental parameters of the considered system are varied in order to understand how the growth process of the disks is influenced, and as a consequence, how the size distribution of the initially formed vesicles depends on it. Compared to previously published analyses of the growth and vesiculation process, based only on the forward scattering, here, the entire q-range is modeled that contains further information on the size distribution (which is encoded in the oscillations in the scattering pattern). This knowledge is important for building well-defined vesicles.

For this purpose diffusion and reaction limited models of disks coalescence are employed as similarly outlined in previous works [140, 92], but here the intention is to obtain a detailed and systematic insight into how certain parameters of this growth process affect experimentally observable structural properties of the disks, such as size distribution, and in particular how they influence the size distribution of the finally formed vesicles.

Physical quantities and parameters that are varied in this study are

1. the initial size distribution of the disks,

2. the critical radius at which vesiculation occurs, and

3. the effect of an activation energy which controls the speed of the coalescence process.

Moreover, the appropriate scattering intensities for the different cases are calculated in order to get an insight into the information content of time resolved scattering experiments. Scattering experiments are especially interesting, since they allow to observe colloidal systems in situ and time resolved.

A refined understanding of the vesicle formation process will be important for future developments in the formation of vesicles with tailor-made structural properties.

5.1. Modeling the Disk Growth Pathway

The considered dynamical growth model is as follows:

1. It is assumed that the initial coalescence state consists of only disk-like micelles. This is considered as experimental observations have been shown that very quickly after mixing of a zwitterionic or a cationic surfactant with an anionic surfactant, these disk-like micelles are build. Moreover, this time is close to the dead-time of the stopped-flow experiment, hence experimentally not observable.

2. In a second step, the initially formed disks grow by means of coalescence. The rate of fusion may depend (*i*) only on the speed of diffusion, where each collision of particles lead to a fusion of them, hence the process is called diffusion limited, or (*ii*) additionally on an energy barrier of fusion, hence not all particle collisions result in a fusion of the particles. In this case the process is called reaction limited.

3. Disks that reach a certain, critical size are assumed to become unstable and close to form vesicles.

4. The rate of coalescence of vesicles is much smaller than the rate of coalescence of disks. Moreover, it is assumed that there is no inter-species (disk-vesicle) coalescence. Hence, the pathway of the coalescence of disks can be independently analyzed from the one of the vesicles.

For a mathematical description of the coalescence mechanism the formula of aggregation of M. von Smoluchowski [165, 166] is employed, which was already used in [140, 92].

5.2. Von Smoluchowski Differential Equation

The following discussion is based on the von Smoluchowski differential equation[45] (DEQ) [165, 166]

$$\frac{\mathrm{d}n_m(t)}{\mathrm{d}t} = \frac{1}{2}\sum_{j=1}^{m-1} K_{m-j,j}(t)n_{m-j}(t)n_j(t) - n_m(t)\sum_{j=1}^{\infty} K_{m,j}(t)n_j(t). \tag{201}$$

[45] For 'continuously' sized particles Eq. (201) can be formulated with integrals instead of sums as an integro-differential equation, which can be justified due the relatively large number of surfactant molecules contained already in the smallest sized disk-like micelles (> 200). However, an analytical solution is only available for a few simple kernels [118] (e.g., a constant kernel with a monodisperse starting distribution [165]). Therefore, for more realistic cases, numerical methods must be employed, requiring a discretization of the integrals, hence resulting into Eq. (201) (possibly with additional integration weights within the sums).

It describes irreversible particle coalescence under the assumption that only two particles can fuse with each other. Thus, coalescence of three and more particles is not taken into account, which is a reasonable assumption for diluted systems. In this equation $n_m(t)$ is the number density of colloidal particles consisting of m unimers (m-mers) at time t, $\mathrm{d}n_m(t)/\mathrm{d}t$ is its rate of change and $K_{m,j}(t)$ gives the rate coefficient for collisions of m-mers with j-mers at time t. In general the rate coefficient can be time-dependent, since the properties of the particles might change with time, e.g., due to a rearrangement of charges between different particles. However, here it is considered that properties are time-invariant and therefore the evolution of the system is solely determined by a time independent kernel matrix $\mathbf{K}$, where $[\mathbf{K}]_{m,j} = K_{m,j}$. The first term on the r.h.s. of Eq. (201) describes the increase of concentration of m-mers resulting from the coalescence of j-mers with $(m-j)$-mers. The factor $1/2$ takes into account that each particle can coalesce only once, whereas the sum takes it into account twice. The second term on the r.h.s. of Eq. (201) describes the decrease of concentration of m-mers resulting from fusions of m-mers with j-mers. The colloidal particles are considered to stay finite in size. Therefore the second sum will only give a contribution up to a certain maximum index $j_{\max}$, hence $n_j = 0$ for $j > j_{\max}$.

Since it is considered that the disks close and become vesicles at a certain critical disk size, the kernel matrix can be decomposed into four submatrices

$$\mathbf{K} = \left(\begin{array}{c|c} \mathbf{K}^{dd} & \mathbf{K}^{dv} \\ \hline \mathbf{K}^{vd} & \mathbf{K}^{vv} \end{array} \right), \tag{202}$$

where $\mathbf{K}^{dd}$ is the kernel describing the disk-disk coalescence, $\mathbf{K}^{vv}$ is the kernel for vesicle-vesicle coalescence, and $\mathbf{K}^{dv}$ and $\mathbf{K}^{dv}$ is the kernel for inter-species coalescence. Here, only $\mathbf{K}^{dd}$ is a non-zero matrix, since it is assumed that the disk-disk coalescence process is much faster than the vesicle-vesicle coalescence process, and the interesting size is the initially formed vesicle size distribution; moreover it is assumed that there is no inter-species coalescence. Considering that vesiculation occurs at a size index $j_{\mathrm{cr.}} + 1$, it follows $j_{\max} = 2j_{\mathrm{cr.}}$, since a maximum sized vesicle is build from two maximum sized disks that coalesce. The aggregate volume associated to index j is $\pi R_j^2 t_m$, corresponding to a disk with radius R_j and thickness t_m if $j \leq j_{\mathrm{cr.}}$, otherwise to a vesicle — see next section. Since a j-mer consists of j unimers with disk radius R_1, the disk radius of the j-mer is

$$\pi R_j^2 t_m = j \pi R_1^2 t_m \tag{203}$$

$$\Rightarrow \qquad R_j = \sqrt{j} R_1 \tag{204}$$

for $j \leq j_{\mathrm{cr.}}$.

5.2.1. Initial Conditions

Since the disk-like micelles and the vesicles are build of bilayer membranes, it is assumed in the following that the thickness (height) of the disk-like micelles is fixed, as well as the thickness of the vesicles is fixed, and both thicknesses are equal to the thickness of the bilayer membrane t_m. For $j > j_{\mathrm{cr.}}$ in the kernel matrix the appropriate vesicle core radius R_j^v can be determined by equating the volume $\pi R_j^2 t_m$ of a disk with radius $R_j = \sqrt{j} R_1$ with the volume $\frac{4\pi}{3}[(R_j^v + t_m)^3 - (R_j^v)^3]$ of a vesicle with core radius R_j^v:

$$R_j^v = -\frac{t_m}{2} + \frac{1}{6}\sqrt{9R_j^2 - 3t_m^2}. \tag{205}$$

The initial disk radius size distribution is considered to be Gaussian, i.e.,

$$R \sim \mathcal{N}\big(R|\mu_R, \sigma_R^2\big), \tag{206}$$

where μ_R is the mean radius and σ_R^2 is the variance of the distribution. Hence the initial continuous number density distribution reads

$$n(R, t = 0) = n_0 \cdot \mathcal{N}\big(R|\mu_R, \sigma_R^2\big), \tag{207}$$

where n_0 is the initial total number density of disk-like micelles, i.e.,

$$n_0 = \frac{\varphi}{\langle v_D(R)\rangle_R}, \tag{208}$$

where φ is the volume fraction of bilayer material, and $\langle v_D(R)\rangle_R$ is the average volume of the initial disk-like micelles. In the implemented program, the initial disk distribution is discretized as

$$n_j(t) := n(R_j, t)\Delta R_j \tag{209}$$

by means of the midpoint integration formula, where the integration support points are taken as $R_j = \sqrt{j}R_1$ (Eq. (204)), the j-th integration interval is $[R_j - \frac{1}{2}(R_j - R_{j-1});\ R_j + \frac{1}{2}(R_{j+1} - R_j)\big]$, and

$$\Delta R_j = \frac{R_{j+1} - R_{j-1}}{2}. \tag{210}$$

The radius of the 'unimer' is chosen according to $R_1 = (\mu_R - 5\sigma_R)/\sqrt{j_{5\sigma_R}}$, hence allowing to control the initial discretization error of the Gaussian distribution by changing $j_{5\sigma_R}$. In order to determine an error of the initial discretization, the relative error

$$\left| \frac{n_0 - \sum_j n_0 \cdot \mathcal{N}(R_j|\mu_R, \sigma_R^2)\frac{R_{j+1}-R_{j-1}}{2}}{n_0} \right| \tag{211}$$

is used. For all evaluations this error was found to be not more than 0.1 %.

5.2.2. Solving the v. Smoluchowski DEQ

There are only a few known solutions to the v. Smoluchowski DEQ, Eq. (201), considering special cases of kernels and a monodisperse initial particle distribution (see, e.g., [118]). Here, the initial particle distributions are considered to be polydisperse and the kernel is another one, hence in this work numerical approaches are employed, and solutions are determined via DEQ solvers provided by MATLAB [153]. E.g., by means of the *ode45* solver[46] using a Runge-Kutta $(4, 5)$ formula or by means of the *ode113* solver[47] using the Adams-Bashforth-Moulton PECE algorithm. Different DEQ solver algorithms have been tested, and all solvers can yield identical solutions if only the specified tolerances are chosen small enough, but the execution time can be significantly different. It has been found that especially the Adams-Bashforth-Moulton PECE solver was relatively fast compared to the other solvers, hence in agreement with the description of the MATLAB help [152]. The DEQ solvers require the kernel matrix $\mathbf{K}$ as well as the discretized initial number density distribution, and they return the time evolution of the number density distribution.

[46] "*ode45* is based on an explicit Runge-Kutta (4,5) formula, the Dormand-Prince pair. It is a one-step solver — in computing $y(t_n)$, it needs only the solution at the immediately preceding time point, $y(t_n - 1)$. In general, *ode45* is the best function to apply as a first try for most problems" [152].

[47] "*ode113* is a variable order Adams-Bashforth-Moulton PECE solver. It may be more efficient than *ode45* at stringent tolerances and when the ODE file function is particularly expensive to evaluate. *ode113* is a multistep solver — it normally needs the solutions at several preceding time points to compute the current solution" [152].

As a check if an algorithm did not diverged the total volume of material at the end of the simulation was checked against the initial volume of material at the beginning (the volume of material is a conserved quantity).

In order to characterize the transient coalescence process following quantities are used:

1. The mean disk radius
$$\mu_R(t) = \frac{1}{\sum_{j=1}^{j_{\text{cr.}}} n_j(t)} \sum_{j=1}^{j_{\text{cr.}}} R_j n_j(t). \tag{212}$$

2. The polydispersity of the radius size distribution
$$p_R(t) = \frac{\sigma_R(t)}{\mu_R(t)}, \tag{213}$$
where
$$\sigma_R(t) = \sqrt{\frac{1}{\sum_{j=1}^{j_{\text{cr.}}} n_j(t)} \sum_{j=1}^{j_{\text{cr.}}} \big(R_j - \mu_R(t)\big)^2 n_j(t)}. \tag{214}$$

3. The z-averaged gyration radius [133, §5.2.4.2]
$$\mu_{R_g}(t) = \sqrt{\frac{\sum_{j=1}^{j_{\max}} n_j(t) v_j^2 R_{g,j}^2}{\sum_{j=1}^{j_{\max}} n_j(t) v_j^2}}, \tag{215}$$
where v_j is the volume of a disk with radius R_j if $1 \leq j \leq j_{\text{cr.}}$, otherwise the volume of a vesicle with core radius R_j^v, cf. Eq. (205). $R_{g,j}$ is the gyration radius of a disk if $1 \leq j \leq j_{\text{cr.}}$, and reads as follows [133, §5.2.1]
$$R_{g,j} = R_g^{\text{disk}}(R_j, t_m) := \sqrt{\frac{R_j^2}{2} + \frac{t_m^2}{12}} \qquad 1 \leq j \leq j_{\text{cr.}}, \tag{216}$$
otherwise the gyration radius of a vesicle, which reads [41, table 3.1]
$$R_{g,j} = R_g^{\text{shell}}(R_j^v, t_m) := \sqrt{\frac{3\big[(R_j^v + t_m)^5 - (R_j^v)^5\big]}{5\big[(R_j^v + t_m)^3 - (R_j^v)^3\big]}} \qquad j > j_{\text{cr.}}. \tag{217}$$

4. The normalized scattering intensity
$$I^n(q,t) = \sum_{j=1}^{j_{\text{cr.}}} n_j(t) I_{\text{disk}}(q, R_j, t_m) + \sum_{j=j_{\text{cr.}}+1}^{j_{\max}} n_j(t) A_{\text{shell}}^2(q, R_j^v, t_m), \tag{218}$$
where $A_{\text{shell}}(q, R_j^v, t_m)$ is the shell (vesicle) amplitude form factor already defined in Eq. (180) with core radius R_j^v and shell (membrane) thickness t_m, and $I_{\text{disk}}(q, R_j, t_m)$ is the intensity form factor of an orientationally averaged disk with radius R_j and thickness t_m given in Eq. (472) in the appendix D.3.

5. The evolution of the volume fraction of disks over time $\varphi(t)$.

6. The characteristic time τ_φ of volume fraction of disks, which describes the point in time at which there is only half of the original volume fraction of disks left, i.e., $\frac{1}{2} = \frac{\varphi(\tau_\varphi)}{\varphi}$.

7. The mean vesicle core radius R_v and the polydispersity p_v of the vesicle core radius distribution at simulation end. These quantities are calculated analogously to Eq. (212)-(214), but here $j > j_{\text{cr.}}$ and the vesicle core radius is used, Eq. (205).

8. The probability density function $\text{pdf}(R_v)$ of the vesicle core radius size distribution is plotted[48]. $\text{pdf}(R_v)$ is obtained as follows: First, the discrete number density distribution $n(j)$ is made continuously in $[j_{\text{cr.}}, j_{\max}]$ by applying a spline interpolation (using MATLAB's `interp1()` function) and then, this new continuous distribution is normalized, yielding $\text{pdf}(j)$. Next, the probability density transformation from appendix A.1 is used to get

$$\text{pdf}(R_v) = \left| \frac{\mathrm{d}j(R_v)}{\mathrm{d}R_v} \right| \text{pdf}(j(R_v)), \tag{219}$$

where herein Eq. (205) and Eq. (204) are employed in order to obtain

$$j(R_v) = \frac{(R_v + \frac{1}{2}t_m)^2 6^2 + 3t_m^2}{9R_1^2}, \tag{220}$$

and

$$\left| \frac{\mathrm{d}j(R_v)}{\mathrm{d}R_v} \right| = \frac{2(R_v + \frac{1}{2}t_m)6^2}{9R_1^2}. \tag{221}$$

5.3. Simulations

In this section solutions of the v. Smoluchowski DEQ are given for differently assumed kernels.

Brownian Diffusion Kernel

If the disks coalescence process is assumed to be diffusion limited, the Brownian kernel can be employed

$$K_{m,j}^{\text{disk}} = 4\pi(D_m + D_j)(R_m^h + R_j^h) \qquad 1 \leq m, j \leq j_{\text{cr.}}, \tag{222}$$

see, e.g., [40], where D_j is the diffusion coefficient of an object with hydrodynamic radius R_j^h. The diffusion coefficient [37] is given as

$$D_j = \frac{k_B T}{6\pi\eta R_j^h} \tag{223}$$

where k_B is the Boltzmann constant ($1.38 \cdot 10^{-23}\,\text{J/K}$), T the temperature (25 °C), and η is the viscosity of the solvent, here assumed to be water, i.e., $\eta = 0.891 \cdot 10^{-3}\,\text{Ns/m}^2$ (at 25 °C).

Different models of the hydrodynamic radii are considered:

Disks: The hydrodynamic radius R_j^h of a disk with radius R_j and thickness t_m has been approximately found to be [102]

$$R_j^h = \frac{3R_j}{2}\left(\sqrt{1+\alpha^2} + \frac{1}{\alpha}\ln\left(\alpha + \sqrt{1+\alpha^2}\right) - \alpha\right)^{-1}, \tag{224}$$

where $\alpha = t_m/(2R_j)$.

[48] Here, R_v is used instead of R_j^v to distinguish between the continuous case and the discrete one.

Spheres: A even more rough approximation of the hydrodynamic radius of a disk with radius R_j is given by assuming the hydrodynamic radius of a sphere with radius R_j, i.e., $R_j^h = R_j$. In this case the kernel becomes

$$K_{m,j}^{\text{spheres}} = 4\pi \frac{k_B T}{6\pi\eta} \frac{(R_m + R_j)^2}{R_m R_j} \qquad 1 \leq m, j \leq j_{\text{cr.}}. \tag{225}$$

Identical Spheres: Considering particles of nearly the same size $R_m \approx R_j \; \forall j, m$, the kernel can be approximated as

$$K_{m,j}^{\text{const.}} = K_{\text{const.}} := \frac{8 k_B T}{3\eta} \qquad 1 \leq m, j \leq j_{\text{cr.}}, \tag{226}$$

hence yielding a value of $K_{\text{const.}} = 1.23 \cdot 10^{-17}\,\text{m}^3/\text{s}$ for the assumed physical quantities.

Fig. 35 shows results obtained for different initial mean disk radii ($3, 4.5, 7\,\text{nm}$), and vesiculation radii ($10 - 70\,\text{nm}$, in steps of $10\,\text{nm}$). The Brownian disks diffusion kernel is employed (Eq. (222)-(224)), and the initial disk polydispersities are $p_R(0) = 0.05$. In this figure, the plot of the evolutions of the disk volume fractions $\varphi(t)$ shows that there are negligible differences between the different volume fraction curves for $R^{\text{cr.}} > 10\,\text{nm}$. The vertical dashed lines define the characteristic times τ_φ, where the disk volume fractions are declined to 50 %.

The plot of the evolutions of the mean gyration radii shows that the mean gyration radius at simulation end only depends on the critical radius, but not on the initial mean disk radius. Differences of the mean gyration radii only exist at the beginning of the simulations, where the mean gyration radius can be related to the initial mean disk radius. There are only negligible differences of the mean gyration radii for the different initial mean disk radii at the characteristic times.

The plot of the evolutions of the mean disk radii behaves very similar to the plot of the evolution of the mean gyration radii, but here the mean disk radii still increase a bit at simulation end. However, at this time the disk volume fractions are very low (e.g., the disk volume fraction for the extreme case $\mu_R(0) = 3\,\text{nm}$, and $R^{\text{cr.}} = 70\,\text{nm}$ is only 0.27% of the total volume fraction), and hence the increase of the mean disk radius at simulation end has practically no meaning.

The plot of the evolutions of the polydispersities shows that the transient behavior of the polydispersity depends on the critical radius as well as on the initial mean disk radius: e.g., at time τ_φ the polydispersity becomes larger if the initial mean disk radius becomes smaller. However, the transient behavior of the polydispersity for different initial mean disk radii at fixed critical radius has practically no influence on the scattering intensity at simulation end as Fig. 36 shows.

Fig. 36 shows the effect of the initial mean disk radius and the critical radius on the scattering intensity and on the initial vesicle radius distribution at simulation end: there is mainly an impact of the initial mean disk radius on the vesicle size distribution for the critical radius $R^{\text{cr.}} = 10\,\text{nm}$ (different oscillatory behavior of the vesicle radius distributions), however, the appropriate scattering intensities show that this effect is practically not observable in the scattering intensity. Note here, that the vesicle size distributions oscillate differently strong, but the polydispersities are very similar (ca. $0.115 - 0.135$) and the intensities are nearly identical.

Fig. 37 shows the dependence of the mean vesicle core radius $\mu_{R_v}(R^{\text{cr.}})$ and the vesicle core radius polydispersity $p_{R_v}(R^{\text{cr.}})$ over the critical radius $R^{\text{cr.}}$ for an initial mean disk radius of $4.5\,\text{nm}$ and an initial disk polydispersity of 0.05. Clearly, there is a linear increase of $\mu_{R_v}(R^{\text{cr.}})$

with $R^{\text{cr.}}$. $p_{R_v}(R^{\text{cr.}})$ can be well described by an exponential decay with offset, if the first point is considered as an outlier. From the fit it can be concluded that the lower bound of the minimal achievable polydispersity is around 0.086 nm.

Changing the initial disk polydispersity from 0.5 to 0.025 or 0.1 produces visibly identical results as obtained for 0.5, which are depicted in Fig. 35. Exemplarily, in Fig. 38 the evolution of the polydispersity is plotted for different initial disk polydispersities and critical radii (initial disk radii are 4.5 nm), showing that the influence of the initial disk polydispersity on the transient polydispersity evolution $p_R(t)$ becomes negligible for $t \gtrsim 10^{-6}$ s, if the critical radius is fixed. The main effect of changing the polydispersity is the impact on the oscillatory behavior of the vesicle size distribution for small critical radii, see Fig. 39. However, as Fig. 39 shows, the oscillatory behavior of the different initial vesicle size distributions has visually no influence on the scattering intensity.

Considering different kernel simplification: Since the different kernels (disks kernel, spheres kernel, and constant kernel) encode the disk size differently, the main impact on the different considered quantities is expected for the largest disk size interval, which here, is given by the initial mean disk radius of 3 nm and by the critical radius of 70 nm. Fig. 40 shows these results: all considered quantities for the disks kernel and spheres kernel are nearly identical, only the constant kernel deviates slightly in its transient behavior from the transient behavior of the other two kernels. Also the initial vesicle size distribution (at simulation end) for the constant kernel deviates very slightly from the distributions of the other two kernels, but the appropriate intensities match perfectly each other.

Reaction Limited Kernel

In a colloidal self aggregating system it can be expected that many coalescence processes are not diffusion limited, since the colloidal particles have typically to overcome a certain energy barrier in order to fuse. Hence, not all collisions will lead to a fusion of the colliding particles. In order to model this process, the Brownian diffusion kernel is multiplied with a Boltzmann factor $W(m,j) = \exp(-E_a(m,j)/(k_BT))$ that describes the probability of the particles to coalesce if they collide, where $E_a(m,j)$ is an activation energy function depending on the particles j and m. However, here it is considered that $E_a(m,j) = E_a^{\text{const.}}$ for $1 \leq j,m \leq j_{\text{cr.}}$, and $E_a^{\text{const.}}$ has been found to be in the order of approximately $10 - 30\, k_BT$, see, e.g., [92]. Accordingly, $W(m,j) = W = \exp(-E_a^{\text{const.}})/(k_BT))$, and the v. Smoluchowski DEQ can be written as

$$\begin{aligned}
&\frac{\mathrm{d}n_m(t)}{\mathrm{d}t} = \frac{1}{2}\sum_{j=1}^{m-1} WK_{m-j,j}n_{m-j}(t)n_j(t) - n_m(t)\sum_{j=1}^{j_{\max}} WK_{m,j}n_j(t) \\
\Leftrightarrow \quad &\frac{\mathrm{d}n_m(t)}{\mathrm{d}(Wt)} = \frac{1}{2}\sum_{j=1}^{m-1} K_{m-j,j}n_{m-j}(t)n_j(t) - n_m(t)\sum_{j=1}^{j_{\max}} K_{m,j}n_j(t) \qquad (227) \\
\Leftrightarrow \quad &\frac{\mathrm{d}\tilde{n}_m(\tilde{t})}{\mathrm{d}\tilde{t}} = \frac{1}{2}\sum_{j=1}^{m-1} K_{m-j,j}\tilde{n}_{m-j}(\tilde{t})\tilde{n}_j(\tilde{t}) - \tilde{n}_m(\tilde{t})\sum_{j=1}^{j_{\max}} K_{m,j}\tilde{n}_j(\tilde{t}),
\end{aligned}$$

where $\tilde{t} = Wt$, and $n_m(t) = n_m(\tilde{t}/W) =: \tilde{n}_m(\tilde{t})$, hence showing that if the Brownian diffusion kernel is multiplied with W, the solution of this new reaction limited process is identical to the solution with the Brownian diffusion kernel, only $\tilde{t}$ has to be divided by W in order to get the 'proper' time t.

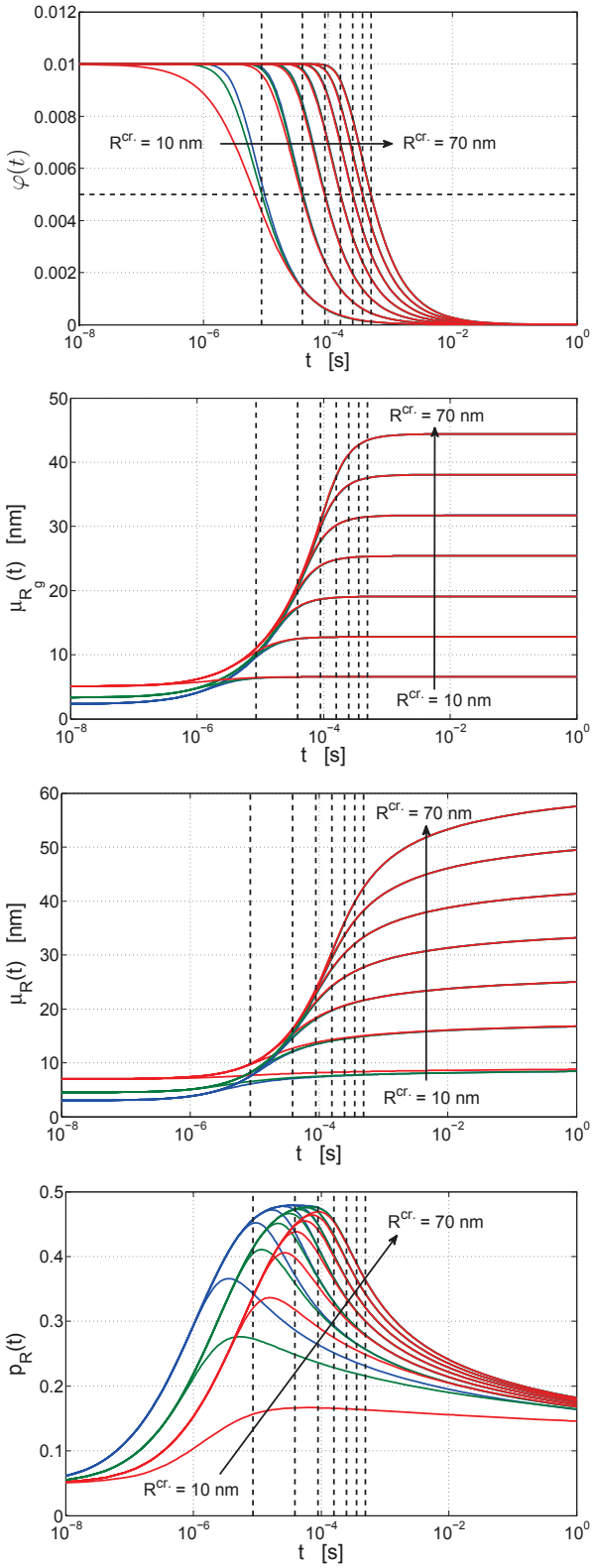


Figure 35: From top to bottom: Evolution of the disk volume fraction $\varphi(t)$, mean gyration radius $\mu_{R_g}(t)$, mean disk radius $\mu_R(t)$, and polydispersity $p_R(t)$ for different initial mean disk radii $(3, 4.5, 7\,\text{nm})$ and different critical radii $R^{\text{cr.}}$ $(10, 20, 30, 40, 50, 60, 70\,\text{nm})$. The Brownian disks diffusion kernel is used, and the initial polydispersity is 0.05. Blue curves: mean start radius 3 nm. Green curves: mean start radius 4.5 nm. Red curves: mean start radius 7 nm. Nearly indistinguishable results are obtained for initial polydispersities 0.025 and 0.1. The vertical lines indicate the half-life time τ_φ of the appropriate disk volume fractions (cf. top diagram).

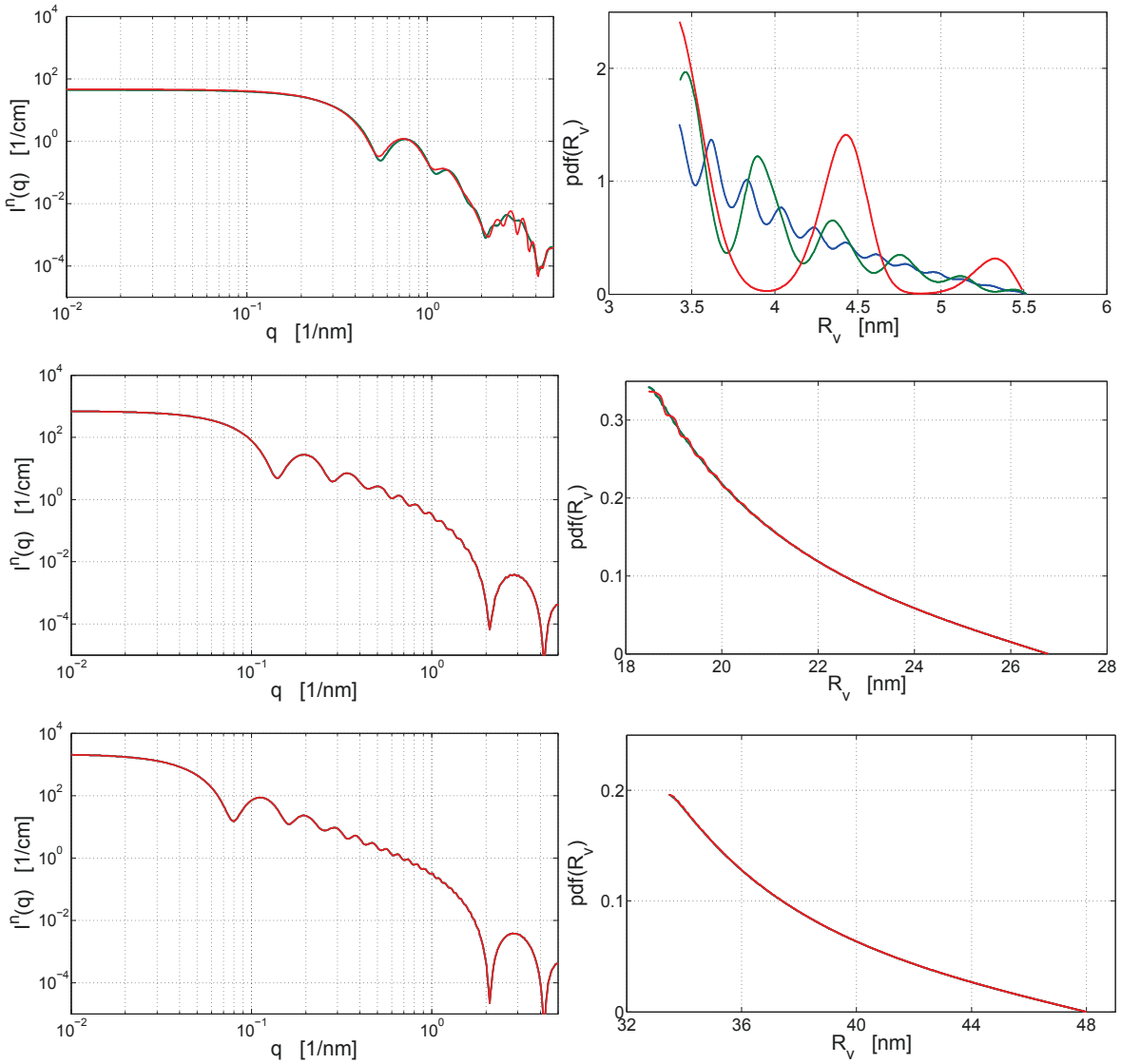


Figure 36: Scattering intensities $I^n(q) = I^n(q, t = 1\,\mathrm{s})$ (left column) and vesicle size distributions $\mathrm{pdf}(R_v)$ (right column) for different initial mean disk radii $3, 4.5, 7\,\mathrm{nm}$ (in this order: blue, green, red) and for critical radii of 10 nm (top row), 40 nm (middle row) and 70 nm (bottom row) at simulation end ($t = 1\,\mathrm{s}$). The Brownian disks diffusion kernel is used, and the initial polydispersities are 0.05. For a critical radius of 10 nm the intensity curves corresponding to initial mean disk radii of 3 nm and 4.5 nm are visually identical, only the intensity curve corresponding to 7 nm is slightly different. The appropriate vesicle size distributions (i.e., for a critical radius of 10 nm) show a similar overall curve behavior, but have a more or less strong oscillatory behavior. These size distributions are right-tailed due to the hard cut-off at the critical/vesiculation radius. Note, for the critical radius 10 nm the vesicle size distributions have polydispersities of 0.135, 0.115, 0.114 and mean radii of 4.06, 3.96, 3.99 nm for the initial mean disk radii of 7, 4.5, 3 nm (in this order). For $R^{\mathrm{cr.}} = 40\,\mathrm{nm}$ the polydispersities are 0.0892 and the mean radii are 20.8 nm. and for $R^{\mathrm{cr.}} = 70\,\mathrm{nm}$ the polydispersities are 0.0864 and the mean radii are 37.6 nm. The intensity curves and vesicle size distributions for a critical radius of 40 nm, and 70 nm are visually identical. Nearly indistinguishable results are obtained for initial polydispersities 0.025 and 0.1, only slight differences exist for the vesicle size distribution corresponding to the critical radius $R^{\mathrm{cr.}}$ =10 nm, cf. Fig. 39.

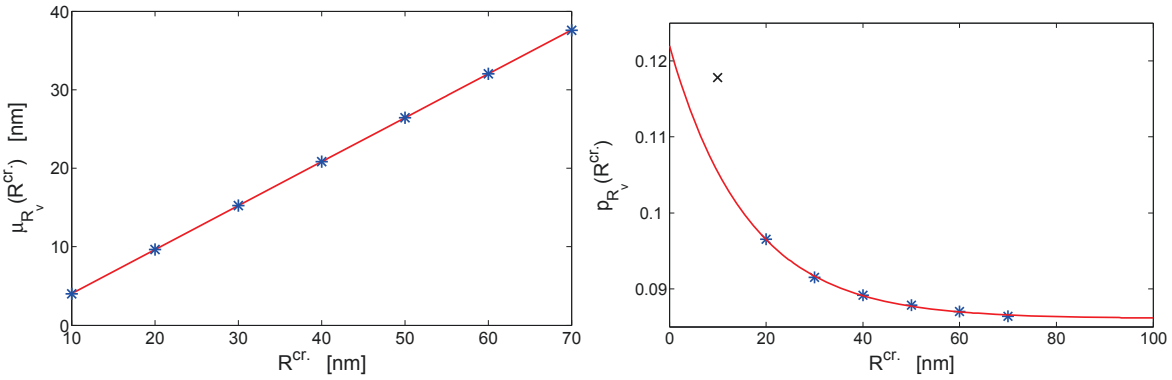


Figure 37: Left plot: determined mean vesicle core radii (blue stars) over critical radii at simulation end, and fit by a line (red curve, the line's offset is -1.569 nm and its slope is 0.6). The initial mean disk radii are 4.5 nm, and the initial polydispersities are 0.05. Right plot: determined vesicle polydispersities (blue stars, and black cross (outlier, not used for the fit)) over critical radii and fit by an decaying exponential (red curve), i.e., $f(R^{\mathrm{cr.}}) = A \cdot \exp(-R^{\mathrm{cr.}}/\tau_{R^{\mathrm{cr.}}}) + b$, where $b = 0.0861\,\mathrm{nm}$, $A = 0.0359\,\mathrm{nm}$, and $\tau_{R^{\mathrm{cr.}}} = 16.13\,\mathrm{nm}$.

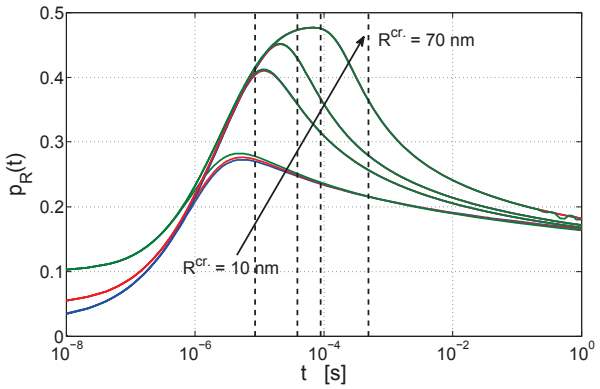


Figure 38: Evolution of the disk polydispersity $p_R(t)$ for different initial disk polydispersities and different critical radii ($R^{\mathrm{cr.}} = 10, 20, 30, 70\,\mathrm{nm}$). The vertical lines indicate the half-life time of the appropriate disk volume fractions.

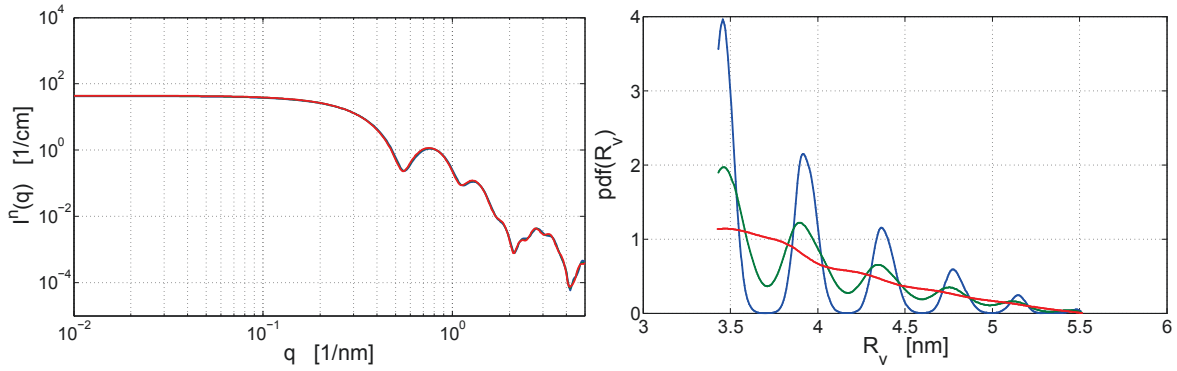


Figure 39: Scattering intensities $I^n(q) = I^n(q, t = 1\,\mathrm{s})$ (left plot) and vesicle size distributions $\mathrm{pdf}(R_v)$ (right plot) for different initial polydispersities $0.025, 0.05, 0.1$ (in this order: blue, green, red) for an initial mean disk radius of 4.5 nm, and a critical radius of 10 nm at simulation end. The Brownian disks diffusion kernel is used. The intensity curves are nearly identical, whereas the size distributions have a similar slope behavior, but they oscillate to a varying extent. Note, the vesicle size distributions have similar polydispersities of 0.12, 0.118, 0.115 and mean radii of 3.94, 4, 4.03 nm for the initial disk polydispersities of 0.025, 0.05, and 0.1 (in this order).

5.3.1. Results

The main results from the Brownian diffusion coalescence simulations are that the initial mean disk radius $(3 - 7\,\mathrm{nm})$ and the initial disk polydispersity $(0.025 - 0.1)$ have practically no influence on the initial vesicle radius distribution and its appropriate scattering intensity (if the critical radius is larger than 10 nm). For the smallest considered critical radius of 10 nm, there are differences (different oscillatory behaviors) within the initial vesicle size distributions (which depend on the initial disk polydispersity and mean disk radius), but this again has no significant impact on the appropriate scattering intensities.

A constant activation energy in a reaction limited kernel can be described by a time scaling of the solution obtained from a Brownian diffusion kernel.

5.4. Experimental Data Fitting

In this section the time evolution of the scattering intensity of a colloidal self-aggregating system — consisting of the three compounds that are given below — is analyzed by means of fitting a reaction limited coalescence model to the intensity data.

TDMAO N,N-tetradecyldimethylamine oxide, $C_{14}H_{29}N(CH_3)_2O$ ($M_w = 257.46\,\mathrm{g{\cdot}mol^{-1}}$, $v_m = 0.4766\,\mathrm{nm}^3$, $\mathrm{SLD}_X = 8.64 \cdot 10^{-4}\,\mathrm{nm}^{-2}$) is a mesoionic surfactant, which is overall neutral but has partial charges (electrons are partially delocalized, drawn away from the nitrogen to the oxygen). In aqueous solutions over the concentration range relevant for this work, it forms cylindrical micelles.

LiPFOS lithium perfluorooctanesulfonate, $C_8F_{17}SO_3^-;^+Li$ ($M_w = 506.06\,\mathrm{g{\cdot}mol^{-1}}$, $v_m = 0.4038\,\mathrm{nm}^3$, $\mathrm{SLD}_X = 17.2 \cdot 10^{-4}\,\mathrm{nm}^{-2}$) is a perfluorinated anionic surfactant. The perfluorinated chain is bulkier, stiffer and more polarizable than alkyl chains. The solubility of perfluorinated chains is also little, with low miscibility in water but also

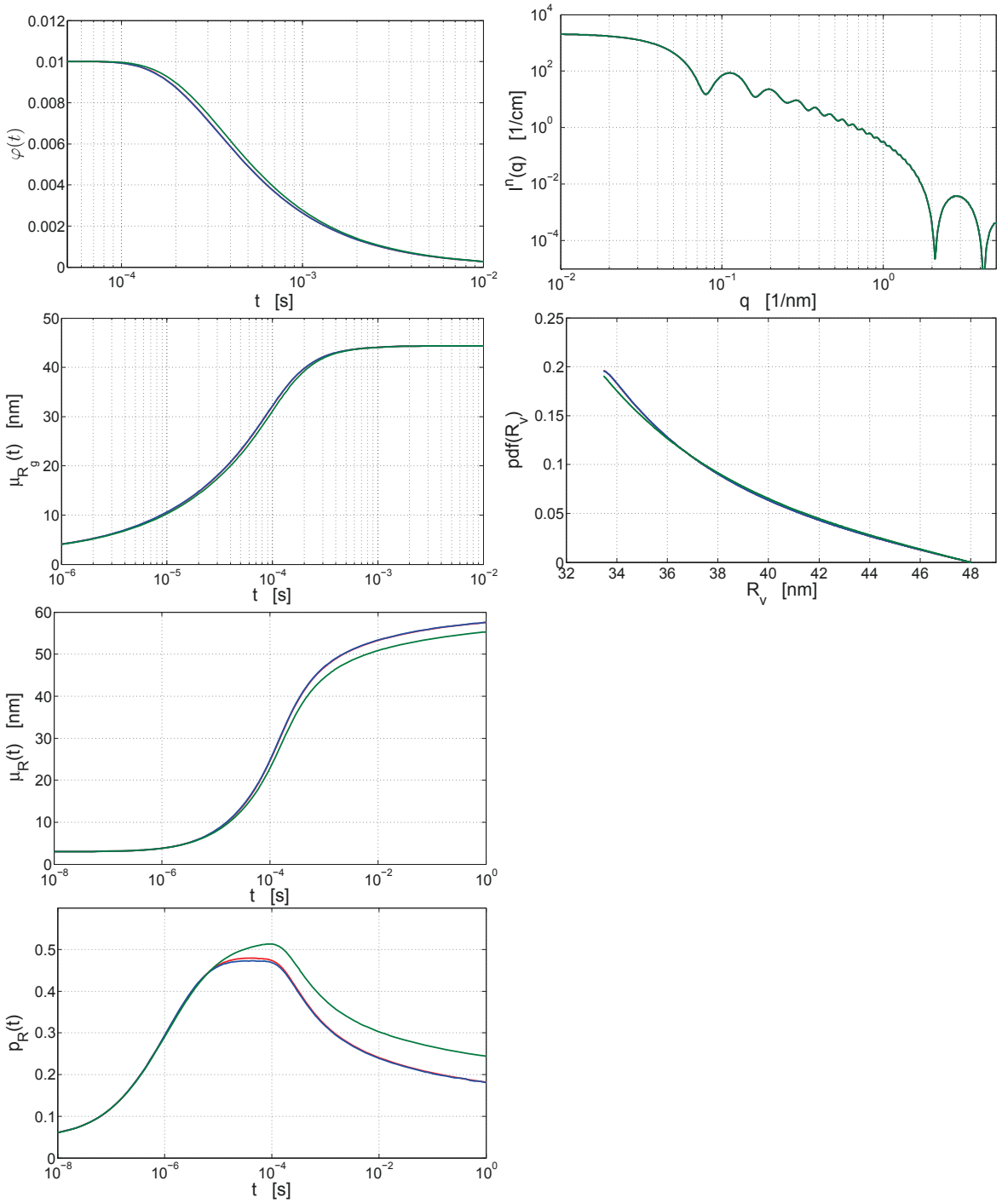


Figure 40: Comparison of different Brownian diffusion kernels. Here, the critical radius is $R^{\text{cr.}} = 70$, the initial mean disk radius is 3 nm, and the initial polydispersity is $p = 0.05$. Red curves: disks kernel, blue curves: spheres kernel, green curves: constant kernel. $\varphi(t)$: volume fraction of disks, $\mu_{R_g}(t)$: mean gyration radius, $\mu_R(t)$: mean disk radius, $p_R(t)$: disk polydispersity, $I^n(q)$: scattering intensity at simulation end, pdf(R_v): vesicle core radius distribution at simulation end. Clearly there are negligible differences for all quantities resulting from employing a disks kernel and a spheres kernel (red curves are often overlaid by the blue curves). Minor transient differences exist between the constant kernel and the disks/spheres kernel. However, the intensity plot shows a perfect match of all different intensity curves. The vesicle size distributions have polydispersities of 0.0864, 0.0864, 0.0868 and mean radii of 3.76, 3.76, 3.77 nm for the disks kernel, spheres kernel, and constant kernel (in this order).

in hydrogenated oils. In aqueous solutions over the concentration range of interest, it forms spheroidal micelles.

Pluronic L35 triblock copolymer with three blocks of two different natures: poly(ethylene oxide) (PEO) and poly(propylene oxide) (PPO), $EO_{11}-PO_{16}-EO_{11}$ ($M_w = 1900\,\mathrm{g{\cdot}mol^{-1}}$, $v_m = 2.98\,\mathrm{nm}^3$, $\mathrm{SLD}_X \approx 9.9 \cdot 10^{-4}\,\mathrm{nm}^{-2}$) is a nonionic amphihilic polymer, where the polypropylene oxide block is rather hydrophobic while the polyethylene oxide block is hydrophilic.

An earlier work [24] suggests that aqueous micellar solutions of these surfactants form, upon mixing, aggregating disk-like micelles which will collapse at a critical size into vesicles. The time lapse of the existence of the rather monodisperse vesicles is controllable by the amount of poloxamers (Pluraonic L35). The poloxamers adsorb onto aggregates and stabilize them by steric interactions, reducing the likeliness of fusion, thus increasing the lifespan of stable solutions of monodisperse vesicles. By adsorbing on the unstable rims of the disks, they also stabilize them, slowing down their growth and increasing the vesiculation radius, thus controlling the size of the monodisperse vesicles.

Here, a TDMAO/(Pluronic L35) solution (concentration of TDMAO: 0.05 mol/L, concentration of Pluronic L35: 0.0005 mol/L) was mixed with a LiPFOS solution (concentration: 0.05 mol/L) in a BioLogic SFM-400 stopped flow apparatus[49] such that the concentrations in the mixed solution were: $c(\mathrm{TDMAO}) = 0.0275\,\mathrm{mol/L}$, $c(\mathrm{LiPFOS}) = 0.000275\,\mathrm{mol/L}$, $c(\text{Pluronic L35}) = 0.0225\,\mathrm{mol/Ls}$, and the kinetics were observed in situ by SAXS.

In detail — following information are friendly provided by Dr. Katharina Bressel —, SAXS experiments were performed at the beamline ID02 at the European Synchrotron Radiation Facility by Dr. Bressel, with a wavelength of 0.1 nm (12 400 eV, relative spread of $2.0 \cdot 10^{-4}$), at different sample-to-detector distances (at 2 m and 4 m, covering a q-regime of $0.02-2.7\,\mathrm{nm}^{-1}$). The 16-bits detector FReLoN Kodak CCD was used, which is $100\times100\,\mathrm{mm}^2$ with a resolution of 2048×2048 pixels (reduced to 512×512 during kinetics to decrease minimum delay between successive frames) and a spatial resolution determined by the point spread function of 80 μm. The beam size at the sample is $100\times300\,\mu\mathrm{m}^2$. The direct beam intensity is captured by a monitor on top of the beam stop, which provides the transmission. The aqueous solutions of pure surfactants are poured into the reservoirs of syringes of the stopped-flow apparatus, which is controlled by a computer; liquid samples are flowing in a quartz capillary of ca. 1.3 mm diameter with ca. 10 μm wall thickness. Data are corrected for transmission, detector aberrations (including electronic noise), scaled with a normalization coefficient determined from the scattering of water ($1.6 \cdot 10^{-2}\,\mathrm{cm}^{-1}$ at 25 °C) and azimuthally averaged. The scattering of solvent (water) was subtracted. Whether the capillary thickness was taken into account during the normalization process is unknown.

5.4.1. Evaluation

Fig. 41 shows the best fit result obtained from adjusting the reaction limited kernel in Eq. (227) such that the appropriate scattering intensity calculated according to Eq. (218) simultaneously fits the measured intensity data of the TDMA/LiPFOS/(Pluronic L35) system (upon mixing it together).

In this fit, the disks diffusion kernel, i.e., Eq. (222)-(224), is used and the coalescence probability has been determined to be $W = 1.1 \cdot 10^{-5}$, yielding $E_a^{\text{const.}} \approx 11.42\,k_BT$, which is well in agreement with [92]. The initial mean disk radius $\mu_R(t = 0)$ as well as the critical

[49] Total mixing volume: 600 μL, flow rate: 6 mL/s, dead volume: 50 μL, dead time: 2.3 ms.

radius $R^{\text{cr.}}$ have also been fitted: $\mu_R(t = 0) = 8\,\text{nm}$, and $R^{\text{cr.}} = 67\,\text{nm}$. The initial disk polydispersity was set to 5 % as this parameter cannot be well adjusted, since (*i*) changing the polydispersity has only a minor impact on the low q intensity regime for $t \lesssim 3\,\text{s}$ (i.e., when there are mainly disks), and (*ii*) there are relatively large model intensity deviations at the low q-regime resulting from a missing structure factor in the employed model. Such a structure factor describing the interaction of disks is not employed here, since the aggregates are anisometric, with certainly an uneven distribution of charges (e.g., an enrichment of charges at the rim), for which there is no analytical structure factor available that takes these facts properly into account. Note, because of the missing structure factor, the fit was achieved by adjusting the model parameters by hand (chi-by-eye approach), hence allowing to include the a priori knowledge 'there is a structure factor, but it cannot be easily modeled'. If otherwise, e.g., a weighted square error objective function would have been taken, an optimal model would be selected, which would also adapt to the low q-regime (for $t \lesssim 3\,\text{s}$, where definitely a structure factor is required. Hence, by adapting the parameters by hand, a model fit was achieved, that intentionally prevented this inconsistency[50].

Deviations at the high q-regime result from the simplification of the bilayer membrane, which here is considered to be homogeneous.

Anyway, the fit captures rather well the kinetics of the TDMA/LiPFOS/(Pluronic L35) system.

5.4.2. Results

The simple reaction limited kernel in Eq. (227) is able to capture the principle kinetic behavior of the TDMA/LiPFOS/(Pluronic L35) system upon mixing. Hence, it can be inferred that the process is based on coalescence, and not on a, e.g., Ostwald ripening process. The fitted parameters: activation energy barrier $E_A \approx 11.42\,k_BT$, initial mean disk radius $\mu_R(0) = 8\,\text{nm}$, and vesiculation (critical) radius $R^{\text{cr.}} = 67\,\text{nm}$ are well in agreement with results from the works of Dr. Bressel [24, 22], which have been obtained from a sequential fitting of models to the scattering curves.

5.5. Conclusion

In this section the process of disk-disk coalescence forming vesicles at a certain critical radius has been analyzed.

The main parameters controlling the coalescence process are the critical radius and the activation energy. If the activation energy is constant, the coalescence process can be seen as a diffusion limited coalescence process that is additionally scaled in time. Other parameters like the initial disk polydispersity, mean disk radius, and simplifications of the diffusion model (spheres kernel or constant kernel instead of a disks kernel) have less influence on the overall coalescence process / initial vesicle size distribution.

The reaction limited kernel has been successfully fitted to an experimental SAXS data set of a TDMAO/LiPFOS/(Pluronic L35) system, proving that the underlying growth process is based on coalescence. In this fit a lot of information has been provided by the many different SAXS intensity curves, and a priori knowledge is used to prevent from a wrong model fitting, hence the obtained parameter values can be seen as quite plausible and reliable. Moreover, the determined parameters are well in agreement with results obtained from earlier works.

[50] In principle the data from the low q-regime for $t \lesssim 3\,\text{s}$ could also be discarded, and then the intensity could be fitted by minimizing a weighted square error objective function. However, the problem of defining the 'useless' data remains, since the structure factor is unknown.

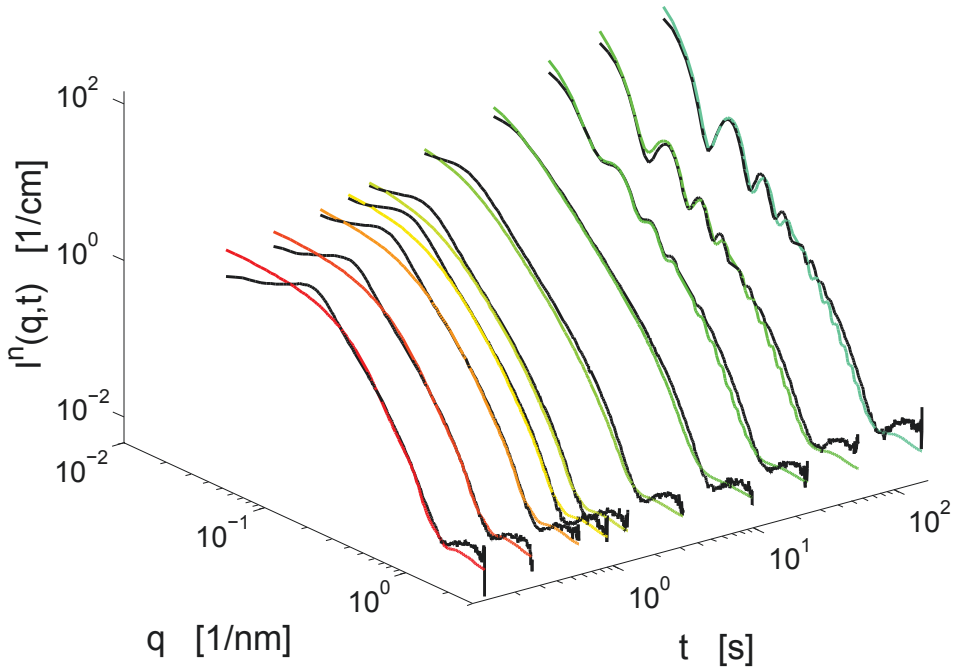


Figure 41: Fitted coaclescence model to the measured SAXS intensities of the selfaggregating system consisting of TDMAO, LiPFOS and Pluronic L35, see text. The black curves are measured intensities. The colored curves are model intensities received from adapting the factor W in the v. Smoluchowski DEQ Eq. (227) such that the appropriate intensities calculated by means of Eq. (218) fit well the measured intensities. The disks diffusion kernel is employed, i.e., Eq. (222)-(224). W has been determined to be $1.1 \cdot 10^{-5}$, i.e., $E_a^{\text{const.}} \approx 11.42\, k_B T$, hence this value is well in agreement with [92]. Furthermore, the initial mean disk radius $\mu_R(t=0)$ and the critical radius $R^{\text{cr.}}$ have been fitted: $\mu_R(t=0) = 8\,\text{nm}$, and $R^{\text{cr.}} = 67\,\text{nm}$. The initial disk polydispersity was set to 5 % — see notes in the text. The bilayer thickness has been determined to be $t_m = 4\,\text{nm}$. Deviations at the high q-regime result from the simplification of the bilayer membrane, which here is considered to be homogeneous. Deviations at the low q-regime result from a missing structure factor — see notes in the text. The relative volume fraction of disks at simulation end ($t_{\text{end}} = 175\,\text{s}$) is 14.5 %.

6. 2-Dimensional Analysis

If anisometric particles are randomly oriented, the scattering is isotropic. However, if anisometric particles are non-randomly orientated, e.g., because of a preferred orientation of the particles resulting from an existing external field (shear-field [117, 39], electric [109] or magnetic field [171]), the scattering can be anisotropic. Hence, the evaluation of such anisotropic pattern is an important issue.

In this section the information contained in a 2-dimensional anisotropic image is analyzed by quantifying its anisotropy. In doing so, different methods can be applied that are implemented in the developed program SASET. Usually, evaluating 2-dimensional images is computationally more demanding than evaluating 1-dimensional images due to the huge amount of data points (number of pixels). However, methods that quantify the anisotropy are commonly rather efficient in order to extract some structural information from a scattering image and presenting this information in some phenomenological or structural parameters.

6.1. Anisotropy Measurement Methods

The different anisotropy measurement methods that are considered (and are implemented in SASET) in the following are the alignment factor, the order parameter, the entropy and the principal component analysis (PCA). Only the order parameter method is capable to additionally determine the orientational distribution function of the particles if certain conditions are fulfilled.

6.1.1. Alignment Factor

Due to [169] the alignment factor is defined as

$$A_f(q) := \frac{\int_0^{2\pi} I(q,\phi)\cos(2\phi)\ \mathrm{d}\phi}{\int_0^{2\pi} I(q,\phi)\ \mathrm{d}\phi}. \tag{228}$$

$I(q,\phi)$ is the intensity measured at (q,ϕ), where q is the magnitude of the scattering vector $\mathbf{q}$ and ϕ an azimuthal angle on the detector plane, and $\phi = 0$ gives the maximum scattering direction. The principle idea is that most azimuthal scattering curves have a partially $\cos(2\phi)$ alignment that can be measured by the above functional. The maximum scattering direction may be determined reliably by a PCA (section 6.1.3) beforehand. For an isotropic image A_f is zero, and is one if the scattering is only along $\phi = 0,\ \pi$. Note, the alignment factor can be generalized by replacing $\cos(2\phi)$ with $\cos(n\phi)$, where $n > 2$ and even. If there is a feature that is (partially) anti-aligned with $\cos(n\phi)$, the appropriate alignment factor can become negative, and -1 is the smallest possible value.

6.1.2. Entropy

Based on information theory (see, e.g., [67, §10]), the quantity

$$A_H = \frac{\ln(N) - H}{\ln(N)} \tag{229}$$

has been defined in order to determine the anisotropy in an intensity ring. H is the discrete entropy defined as

$$H = -\sum_{n=1}^{N} p_n \ln(p_n), \tag{230}$$

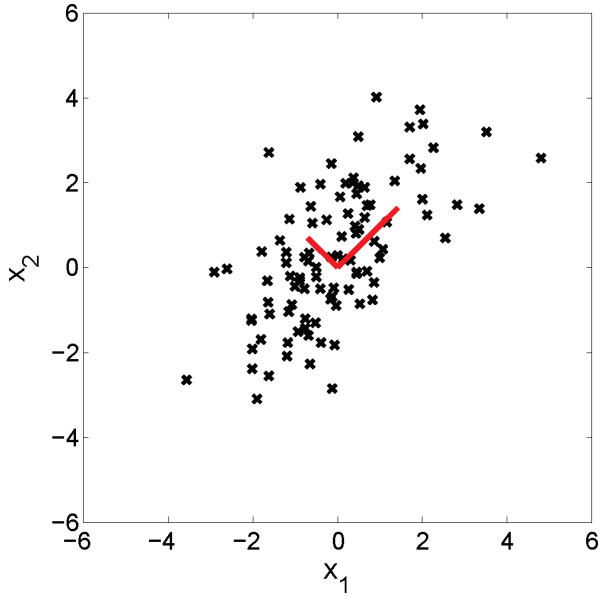


Figure 42: An arbitrary data set (data cloud). The two red lines are aligned along the extremal scattering directions and are orthogonal to each other.

where N is number of intensity values in the ring and

$$p_n = \frac{I_n}{\sum_{\nu=1}^{N} I_\nu} \tag{231}$$

is the n-th normalized intensity such that $\sum_{n=1}^{N} p_n = 1$, where I_n is the n-th intensity. p_n can be seen as relative frequency of photon/neutron occurrence in the n-th pixel in the ring. The discrete entropy is bounded by $0 \leq H \leq \ln(N)$ [67], where $H = \ln(N)$ occurs if $p_n = p_{\text{const.}}\, \forall n$. Therefore $A_H = 0$ if the intensity is isotropic and any deviation from the isotropic image leads to an increase of A_H. $A_H = 1$ is reached for $p_n = 1$ and $p_m = 0 \;\; \forall m \neq n$.

6.1.3. Principal Component Analysis (PCA)

The PCA is a well established method in statistics, signal processing and pattern recognition and is used here to quantify the anisotropy in an image as well as to determine the maximum scattering direction.

Consider a discrete, bivariate data cloud consisting of N data points $\{\mathbf{x}_n\}_{n=1}^{N}$ as depicted in Fig. 42. The projected variance on a complex feature $\mathbf{k}$ is $\left\langle \mathbf{k}^T(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^T\mathbf{k} \right\rangle$, where the $\langle\cdot\rangle$ means the average over the N data points and $\overline{\mathbf{x}} = \langle \mathbf{x}_n \rangle$. Next, a direction $\mathbf{k}$ is searched on which the variance of the projected data becomes extremal subject to keeping the length of $\mathbf{k}$ fixed. This is a constraint optimization problem, which can be solved by means of the Lagrangian method, yielding the Lagrangian function

$$L(\mathbf{k}, \lambda) = \mathbf{k}^T \mathbf{R} \mathbf{k} - \lambda(\mathbf{k}^T\mathbf{k} - 1), \tag{232}$$

where

$$\mathbf{R} := \left\langle (\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^T \right\rangle \tag{233}$$

is the covariance matrix of the data. Taking the partial derivative $\partial L(\mathbf{k}, \lambda)/\partial \mathbf{k}$ and setting

the result equal to zero gives the eigenvalue problem

$$\mathbf{R}\mathbf{k} = \lambda\mathbf{k}, \tag{234}$$

which has two solutions $\{(\lambda_i, \mathbf{k}_i)\}_{i=1}^2$ for the bivariate data set. Assuming[51] $\lambda_1 > \lambda_2$, then the maximum scattering direction is given by $\mathbf{k}_1$ and the appropriate variance is λ_1 as it can be seen by $\mathbf{k}_1^T\mathbf{R}\mathbf{k}_1 = \lambda_1\mathbf{k}_1^T\mathbf{k}_1 = \lambda_1$. Furthermore, it can be shown [15, appendix A] that $\mathbf{k}_2$ is orthogonal to $\mathbf{k}_1$. Therefore, the quantity

$$A = \frac{\sqrt{\lambda_1} - \sqrt{\lambda_2}}{\sqrt{\lambda_1}} \qquad \lambda_1 \geq \lambda_2 \tag{235}$$

has been defined to measure the anisotropy, where $0 \leq A \leq 1$. However, since a scattering image consists of intensity values instead of having a discrete data cloud, the covariance matrix $\mathbf{R}$ is replaced by

$$\mathbf{R}_I = \sum_{n=1}^{N} p_n \mathbf{q}_n \mathbf{q}_n^T. \tag{236}$$

where p_n is the normalized intensity from Eq. (231) and $\mathbf{q}_n$ is the appropriate scattering vector.

It can make sense to project the intensity onto the unit circle, i.e., to normalize with respect to the magnitude of the scattering vector, which is accomplished by substituting $\mathbf{q}_n$ by $\mathbf{q}_n/\|\mathbf{q}_n\|$. This is useful in order to compare different q-regimes with each other.

6.1.4. Order Parameters

Deas [34][52,53] showed that in case the orientational distribution function $D(\alpha)$ of a particle is expanded into Legendre polynomials $P_{2k}(\cdot)$

$$D(\alpha) = \sum_{k=0}^{\infty} D_{2k} P_{2k}(\cos\alpha), \tag{237}$$

where

$$D_{2k} = \frac{4k+1}{2} \int_0^{\pi} D(\alpha) P_{2k}(\cos\alpha) \sin\alpha \, \mathrm{d}\alpha, \tag{238}$$

that the coefficients D_{2k}, called order parameters, can be determined as well from the scattering intensity and the intensity of an aligned model via

$$D_{2k} = \frac{4k+1}{2} \frac{I_{2k}(r^*)}{G_{2k}^0(r^*)}, \tag{239}$$

since the scattering intensity can be expressed as

$$\begin{aligned} I(r^*, \theta^*) &= \sum_{k=0}^{\infty} I_{2k}(r^*) P_{2k}(\cos\theta^*) \\ &= \sum_{k=0}^{\infty} \frac{2}{4k+1} D_{2k} G_{2k}^0(r^*) P_{2k}(\cos\theta^*), \end{aligned} \tag{240}$$

[51] The eigenvalue problem has two identical eigenvalues only if $[\mathbf{R}]_{12} = [\mathbf{R}]_{21} = 0$ and $[\mathbf{R}]_{11} = [\mathbf{R}]_{22}$, which can be seen by calculating its eigenvalues.

[52] Later also in Wilke and Göttlicher [174].

[53] In what follows, the notation from [34] is taken.

and the orthogonality relation

$$\int_0^\pi P_k(\cos\theta^*)P_\nu(\cos\theta^*)\sin\theta^*\ \mathrm{d}\theta^* = \frac{2}{2\nu+1}\delta_{k\nu} \tag{241}$$

can be applied, and $\{G^0_{2k}(r^*)\}_{k=0}^\infty$ are coefficients of an expansion of the intensity of an aligned model (the particle is orientated along the director):

$$I_{\text{aligned}}(r^*,\theta^*) = \sum_{k=0}^\infty G^0_k(r^*)P_k(\cos\theta^*). \tag{242}$$

α is the angle enclosed between the director $\mathbf{f}$ of the particle and a particle axis $\mathbf{n}$ (also called structural unit axis in [108], or sometimes also particle structure normal). The scattering vector is $\mathbf{q} = (r^*,\theta^*,\phi^*)^T$, where the director is taken as the polar axis of the spherical coordinate system (r^*,θ^*,ϕ^*) (r^*: radial distance, θ^*: polar angle, ϕ^*: azimuthal angle[54]). In his work he has shown that Eq. (239) and Eq. (240) are valid under the assumption that (i) the particle uniformly rotates around its molecular axis $\mathbf{n}$ or that there is a rotational symmetry around $\mathbf{n}$, and (ii) $\mathbf{n}$ rotates uniformly around the director $\mathbf{f}$, cf. Fig. 43.

Eq. (242) and Eq. (240) give the intensity of only one particle. If there is a collection of independent scatterers the formulas have to be multiplied by the number of particles.

Next, some normalized coefficients are considered. For this, the orthogonality relation Eq. (241) is applied to Eq. (240), yielding

$$\begin{aligned}\int_0^\pi I(r^*,\theta^*)P_{2\nu}(\cos\theta^*)\sin\theta^*\ \mathrm{d}\theta^* &= \int_0^\pi \sum_{k=0}^\infty \frac{2}{4k+1}\,D_{2k}G^0_{2k}(r^*)P_{2k}(\cos\theta^*)P_{2\nu}(\cos\theta^*)\sin\theta^*\ \mathrm{d}\theta^*\\ &= \sum_{k=0}^\infty \frac{2}{4k+1}\,D_{2k}G^0_{2k}(r^*)\frac{2}{4\nu+1}\delta_{k\nu}\\ &= \left(\frac{2}{4\nu+1}\right)^2 D_{2\nu}G^0_{2\nu}(r^*).\end{aligned} \tag{243}$$

For $\nu = 0$ the result is

$$\int_0^\pi I(r^*,\theta^*)\sin\theta^*\ \mathrm{d}\theta^* = 2^2 D_0 G^0_0(r^*). \tag{244}$$

Dividing the l.h.s. of Eq. (243) by the l.h.s. of Eq. (244) yields

$$\frac{\int_0^\pi I(r^*,\theta^*)P_{2\nu}(\cos\theta^*)\sin\theta^*\ \mathrm{d}\theta^*}{\int_0^\pi I(r^*,\theta^*)\sin\theta^*\ \mathrm{d}\theta^*} =: \langle P_{2\nu}(\cos\theta^*)\rangle_I, \tag{245}$$

and similarly for the r.h.s.

$$\frac{1}{(4\nu+1)^2}\frac{D_{2\nu}G^0_{2\nu}}{D_0G^0_0} = \langle P_{2\nu}(\cos\alpha)\rangle_D\langle P_{2\nu}(\cos\theta^*)\rangle_{I_{\text{aligned}}}, \tag{246}$$

since

$$\frac{1}{4\nu+1}\frac{D_{2\nu}}{D_0} = \frac{1}{4\nu+1}\frac{\frac{4\nu+1}{2}\int_0^\pi D(\alpha)P_{2\nu}(\cos\alpha)\sin\alpha\ \mathrm{d}\alpha}{\frac{1}{2}\int_0^\pi D(\alpha)\sin\alpha\ \mathrm{d}\alpha} =: \langle P_{2\nu}(\cos\alpha)\rangle_D \tag{247}$$

[54]Here, the scattering intensity is actually independent of the azimuthal angle [34].

and

$$\frac{1}{4\nu+1}\frac{G_{2\nu}^0}{G_0^0} = \frac{1}{4\nu+1}\frac{\frac{4\nu+1}{2}\int_0^\pi I_{\text{aligned}}(r^*,\theta^*)P_{2\nu}(\cos\theta^*)\sin\theta^*\ \mathrm{d}\theta^*}{\frac{1}{2}\int_0^\pi I_{\text{aligned}}(r^*,\theta^*)\sin\theta^*\ \mathrm{d}\theta^*} \\ =: \langle P_{2\nu}(\cos\theta^*)\rangle_{I_{\text{aligned}}}. \tag{248}$$

Therefore, it follows

$$\langle P_{2\nu}(\cos\alpha)\rangle_D = \frac{\langle P_{2\nu}(\cos\theta^*)\rangle_I}{\langle P_{2\nu}(\cos\theta^*)\rangle_{I_{\text{aligned}}}}. \tag{249}$$

$\{\langle P_{2\nu}(\cos\alpha)\rangle_D\}$ are normalized order parameters. $\langle P_2(\cos\alpha)\rangle_D$ is often simply called order parameter or Herman's orientation parameter [133]. Sometimes [96, 43, 182] $\langle P_2(\cos\theta^*)\rangle_I$ is used to quantify the anisotropy and in these contexts $\langle P_2(\cos\theta^*)\rangle_I$ is also called Herman's orientation parameter.

The first coefficient D_0 in Eq. (237) is always constant, since

$$D_0 = \frac{1}{2}\int_0^\pi D(\alpha)P_0(\cos\alpha)\sin\alpha\ \mathrm{d}\alpha = \frac{1}{2}, \tag{250}$$

and with Eq. (247) it follows

$$\langle P_{2\nu}(\cos\alpha)\rangle_D = \frac{2}{4\nu+1}D_{2\nu}. \tag{251}$$

Eq. (239), respectively Eq. (249) together with Eq. (251), can be used in order to determine experimental estimations for D_{2k}, namely D_{2k}^{exp}, and therewith to find an approximation for $D(\alpha)$ in Eq. (237). This is accomplished by taking the measured intensity I_{exp} instead of the true intensity I and an assumed model intensity I_{m} is taken for the intensity of the true aligned particle I_{aligned}, i.e.,

$$D(\alpha) \approx D_{\text{exp}}(\alpha) = \sum_{k=0}^{K} D_{2k}^{\text{exp}} P_{2k}(\cos\alpha), \tag{252}$$

where

$$D_{2k}^{\text{exp}} = \frac{4k+1}{2}\frac{\langle P_{2k}(\cos\theta^*)\rangle_{I_{\text{exp}}}}{\langle P_{2k}(\cos\theta^*)\rangle_{I_{\text{m}}}}. \tag{253}$$

K is the maximum number of determined coefficients.

Eq. (252) and Eq. (253) give a direct way to determine the orientational distribution function. However, the solution may be easily influenced by noise in the data and by model inaccuracies such that the solution may contain artifacts as, e.g., oscillations or that the solution becomes negative, cf. Fig. 47. Via the Maximum Entropy Method (MEM) [76], as already suggested in [20], a solution can be found having the aforementioned effects suppressed and is discussed in the next section.

6.1.5. Applying the Maximum Entropy Method (MEM)

The MEM [76] determines the most general solution given some constraints. This solution is characterized by the fact that it contains a maximum of entropy [139] S, "i.e., it is maximally noncomittal with regard to missing information" [76]. Here, the entropy is given by

$$S = -\int_0^\pi \tilde{D}(\alpha)\ln\left(\tilde{D}(\alpha)\right)\sin\alpha\ \mathrm{d}\alpha, \tag{254}$$

where $\tilde{D}(\alpha)$ is a test orientational distribution function. Then, S has to be maximized subject to the following two constraints

(*i*) $\tilde{D}(\alpha)$ is normalized

$$\int_0^\pi \tilde{D}(\alpha)\sin\alpha \, \mathrm{d}\alpha = 1, \tag{255}$$

(*ii*) the experimentally determined coefficients $\{D_{2k}^{\mathrm{exp}}\}_{k=0}^K$ can be recovered, i.e.,

$$\frac{4k+1}{2}\int_0^\pi \tilde{D}(\alpha)P_{2k}(\cos\alpha)\sin\alpha \, \mathrm{d}\alpha = D_{2k}^{\mathrm{exp}}. \tag{256}$$

This optimization problem can be reformulated via the Lagrangian function

$$\begin{aligned} L(\tilde{D}) &= -\int_0^\pi \tilde{D}(\alpha)\ln\big(\tilde{D}(\alpha)\big)\sin\alpha \, \mathrm{d}\alpha + \lambda_0\left(\int_0^\pi \tilde{D}(\alpha)\sin\alpha \, \mathrm{d}\alpha - 1\right) \\ &\quad + \sum_{k=1}^K \lambda_K\left(\frac{4k+1}{2}\int_0^\pi \tilde{D}(\alpha)P_{2k}(\cos\alpha)\sin\alpha \, \mathrm{d}\alpha - D_{2k}^{\mathrm{exp}}\right) \\ &= \int_0^\pi \left[\tilde{D}(\alpha)\ln\big(\tilde{D}(\alpha)\big) + \lambda_0\tilde{D}(\alpha) + \sum_{k=1}^K \lambda_k\frac{4k+1}{2}\tilde{D}(\alpha)P_{2k}(\cos\alpha)\right]\sin\alpha \, \mathrm{d}\alpha \\ &\quad - \lambda_0 - \sum_{k=1}^K \lambda_k D_{2k}^{\mathrm{exp}}, \end{aligned} \tag{257}$$

where $\{\lambda_k\}_{k=0}^K$ is a set of Lagrange parameters. Setting $\tilde{D}(\alpha) = D_{\mathrm{MEM}}(\alpha) + \epsilon\eta(\alpha)$, where $\epsilon \in \mathbb{R}$, and $\eta(\alpha)$ is an arbitrary function, and $D_{\mathrm{MEM}}(\alpha)$ is a desired maximum entropy solution. Then, taking the first functional derivative [84, §9.2] of Eq. (257) and setting it equal to zero yields

$$\int_0^\pi \left[-1 - \ln\big(D_{\mathrm{MEM}}(\alpha)\big) + \lambda_0 + \sum_{k=1}^K \lambda_k\frac{4k+1}{2}P_{2k}(\cos\alpha)\right]\eta(\alpha)\sin\alpha \, \mathrm{d}\alpha = 0. \tag{258}$$

This holds for any $\eta(\alpha)$, therefore it follows

$$-1 - \ln\big(D_{\mathrm{MEM}}(\alpha)\big) + \lambda_0 + \sum_{k=1}^K \lambda_k\frac{4k+1}{2}P_{2k}(\cos\alpha) = 0, \tag{259}$$

from which $D_{\mathrm{MEM}}(\alpha)$ can be determined

$$\begin{aligned} D_{\mathrm{MEM}}(\alpha) &= \exp\left(-1 + \lambda_0 + \sum_{k=1}^K \lambda_k\frac{4k+1}{2}P_{2k}(\cos\alpha)\right) \\ &= \frac{1}{Z}\exp\left(\sum_{k=1}^K \lambda_k\frac{4k+1}{2}P_{2k}(\cos\alpha)\right), \end{aligned} \tag{260}$$

where Z is a normalization constant that can be determined from the first constraint. Hence, the second constraint can be written as

$$D_{2\tilde{k}}^{\mathrm{exp}} \stackrel{!}{=} \frac{4\tilde{k}+1}{2}\int_0^\pi \frac{1}{Z}\exp\left(\sum_{k=1}^K \lambda_k\frac{4k+1}{2}P_{2k}(\cos\alpha)\right)P_{2\tilde{k}}(\cos\alpha)\sin\alpha \, \mathrm{d}\alpha, \tag{261}$$

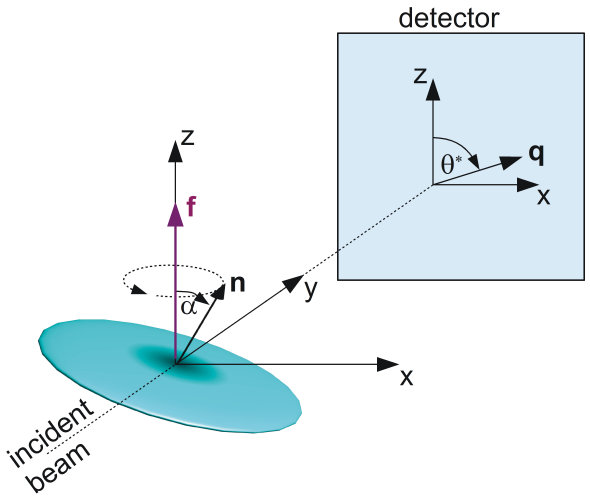


Figure 43: The spheroid rotates uniformly with its short axis, that is along **n**, around the director **f**. The fraction of particles lying in the interval $[\alpha, \alpha + \mathrm{d}\alpha]$ is given by $D(\alpha)\sin\alpha\,\mathrm{d}\alpha$, where $D(\alpha)$ is a Maier-Saupe orientational distribution function. The scattering vector **q** is assumed to lie in the (z, x) plane of the detector.

and the Lagrange parameters $\{\lambda_k\}_{k=1}^K$ can be determined by means of a non-linear optimization routine having the objective function

$$\sum_{\tilde{k}=1}^{K}\left\{D_{2\tilde{k}}^{\exp} - \frac{4\tilde{k}+1}{2}\int_0^\pi \frac{1}{Z}\exp\left(\sum_{k=1}^{K}\lambda_k\frac{4k+1}{2}P_{2k}(\cos\alpha)\right)P_{2\tilde{k}}(\cos\alpha)\sin\alpha\;\mathrm{d}\alpha\right\}^2, \qquad (262)$$

which is to be minimized with respect to $\{\lambda_k\}_{k=1}^K$.

As it is seen from Eq. (260) $D_{\mathrm{MEM}}(\alpha)$ is always positive. Therefore no additional constraint is required, stating that the solution has to be positive. Moreover, if $K = 1$ then $D_{\mathrm{MEM}}(\alpha)$ is the Maier-Saupe distribution.

6.2. Example and Results

The following simulation is considered in order to discuss the described anisotropy measurement methods: A collection of independent oblate shaped ellipsoids of revolution with axes 1, 10, 10 nm rotate uniformly with their short axis **n** around the director **f**, see Fig. 43. The angle enclosed between **f** and **n** is α, and the fraction of oblates lying in the infinitesimal small interval $[\alpha, \alpha + \mathrm{d}\alpha]$ is $D(\alpha)\sin\alpha\,\mathrm{d}\alpha$, where $D(\alpha)$ is an orientational distribution function. Here, $D(\alpha)$ is assumed to be a Maier-Saupe distribution [99], i.e.,

$$D(\alpha) = \frac{1}{\tilde{Z}}\exp(m\cos^2(\alpha)), \qquad (263)$$

where m is a width parameter of the distribution and $\tilde{Z}$ is a normalization constant such that

$$\int_0^\pi D(\alpha)\sin\alpha\;\mathrm{d}\alpha = 1 \qquad (264)$$

holds. In the simulated scattering experiments m is varied between 0.1 and 100. The scattering for $m = 0.1$, 1, 10, 100 is depicted in Fig. 44. In all anisotropy methods the scattering

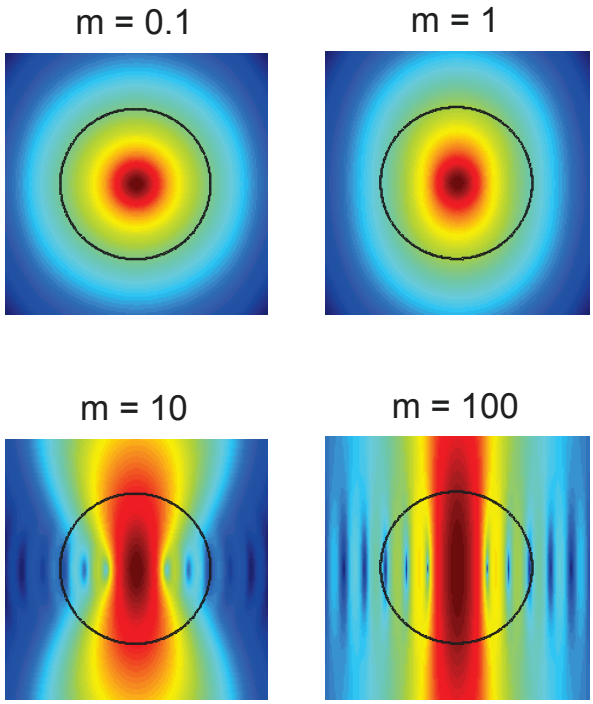


Figure 44: The scattering images for an oblate that has an orientational distribution according to a Maier-Saupe distribution with a parameter $m = 0.1, 1, 10, 100$. The circle corresponds to $\|\mathbf{q}\| = 1.15\,\mathrm{nm}^{-1}$. The intensity is plotted in logarithmic scale and in arbitrary units, from blue (low intensity) to red (high intensity).

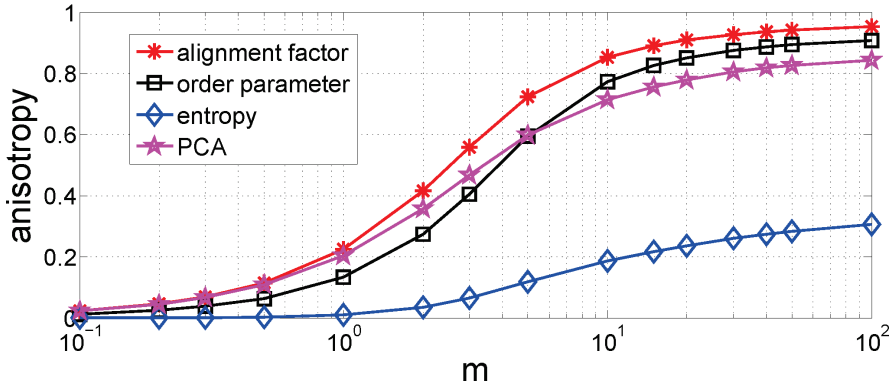


Figure 45: Results of the different anisotropy measurement methods for the simulated scattering experiment. m is the width parameter in the Maier-Saupe orientational distribution function. The order parameter curve only represents the quantity $\langle P_2(\cos\theta^*)\rangle_I$ in Eq. (249).

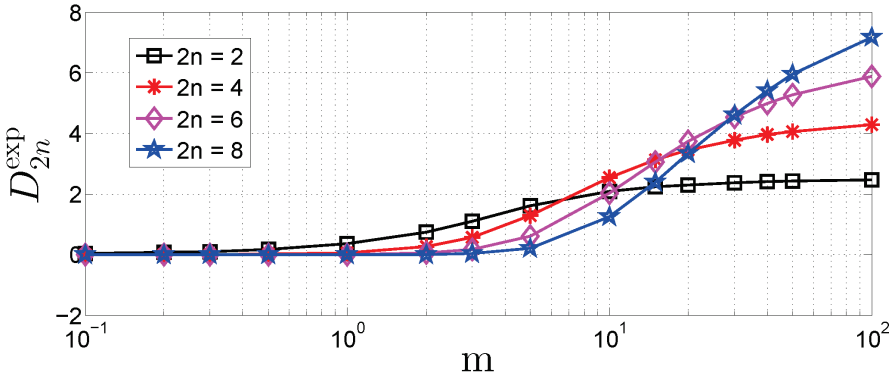


Figure 46: The first four determined experimental order parameters $\{D_{2n}^{\exp}\}_{n=1}^{4}$ for the simulated scattering experiment. m is the width parameter in the assumed Maier-Saupe distribution.

data are taken from an intensity ring at $\|\mathbf{q}\| = 1.15\,\mathrm{nm}^{-1}$. This value was chosen somewhat arbitrarily, but such that the ring contains many intensity values and structure. The results for the alignment factor, order parameter (only $\langle P_2(\cos\theta^*)\rangle_I$ in Eq. (249)), PCA and entropy are depicted in Fig. 45. All curves have a very similar slope behavior. In Fig. 46 the first four experimental orientation parameters $D_{2n}^{\exp}$ (Eq. (253)) are plotted. These parameters are used to reconstruct the orientational distribution functions directly via Eq. (252) as well as to get the orientational distribution functions from the MEM. All solutions are depicted in Fig. 47. As it can be seen from this figure, the MEM gives always solutions that are in good agreement with their underlying Maier-Saupe distributions, while the direct solutions may oscillate and may even be negative.

6.3. Conclusion

Four different anisotropy measurement methods are presented. The alignment factor and the order parameter are already existing methods, while the entropy and the PCA are introduced here. The entropy is well founded in information theory, and can always easily be applied. The PCA can quantify the anisotropy in an image, and additionally it can also find the maximum

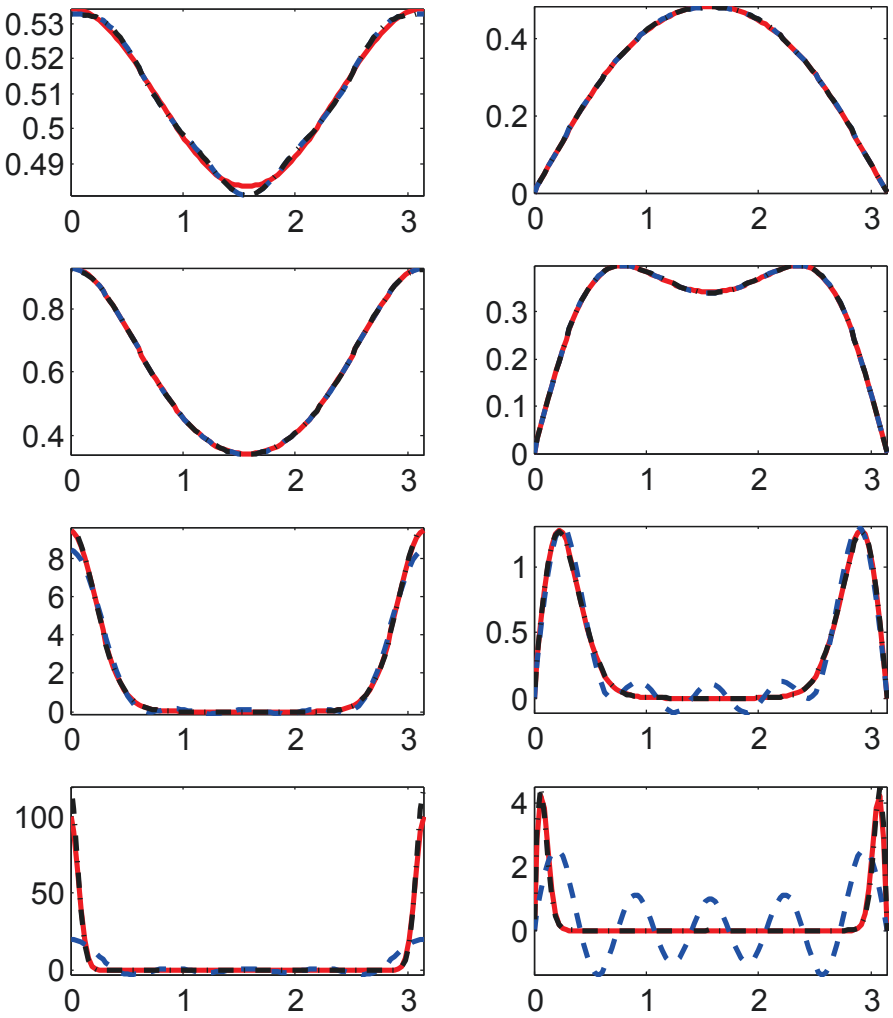


Figure 47: Left column: orientational distribution functions versus α, right column orientational distribution functions times $\sin(\alpha)$ versus α. From top to bottom row: $m = 0.1, 1, 10, 100$. Red and line style '-': true $D(\alpha)$, blue and line style '- -': approximate direct solution, black and line style '-.-.': MEM solution. $D(\alpha)$ is the Maier-Saupe orientational distribution function and m is a parameter controlling the width of the distribution.

scattering direction, which can subsequently be used for determining the alignment factor or the order parameters. However, only the order parameter method is capable of inferring the orientational distribution function of particles (if the shape of the particles is known), which however should be done in consensus with the maximum entropy principle in order to get a robust solution.

7. Indirect Fourier Transform (IFT)

The IFT is used to get structural information about a system from a scattering experiment by means of determining the Pair Distance Distribution Function (PDDF) from which, e.g., size and shape information of scattering objects can be deduced [50, 26]. The IFT is a regularized version of the inverse Fourier transform, which is an ill-posed problem, hence it requires some kind of regularization in order to yield a reasonable solution.

In this section, the mathematical derivation of the method is explained, showing how numerical instabilities arising from such an ill-posed problem are reduced by regularization methods. The next subsection shortly discusses the theory of ill-posed problems in a classical mathematical framework. However, the problem can also be seen as an inference one, and the subsequent two subsections discuss the problem in the Bayesian inference approach as well as in the frequentist one. These two approaches yield the evidence and the generalization error as objective functions — cf. section 2.2. The evidence was already introduced earlier [64], while the present work suggests the generalization error as an alternative; indeed a quite good one. Using these two different objective functions, the inverse Fourier transform is discussed in the context of modern machine learning methods (RVM, SVR, LASSO) in the subsequent subsection yielding new regularization / inference schemes.

For the isotropic scattering case of randomly oriented and non-interacting particles the scattering intensity $\hat{I}(q)$ is related to the PDDF $\hat{p}(r)$ by a first order Fredholm integral operator[55] A:

$$\hat{I}(q) := (A\hat{p})(q) := 4\pi \int_{\Xi} \hat{p}(r) \frac{\sin(qr)}{qr} \, \mathrm{d}r, \tag{265}$$

which is also known as Fourier transform [49]. Solving this equation for $\hat{p}(r)$ can be regarded as an inverse problem, and a short introduction to the theory of inverse problems is given in the next section. As suggested by Glatter [49], it is beneficial to assume that the PDDF is approximated as a linear combination of basis functions $\{\phi_m\}_{m=1}^{M}$

$$\hat{p}(r) \approx p(r) := \sum_{m=1}^{M} w_m \phi_m(r), \tag{266}$$

e.g., b-splines. $\{w_m\}_{m=1}^{M}$ is a set of weights. For an inverse problem it is usually difficult to find a reasonable good solution. Often solutions are corrupted by artifacts, e.g., in form of strong oscillations. This problem can be exemplified by considering a discrete and noisy scattering curve represented by the set $\mathcal{D} := \{(q_n, I_n)\}_{n=1}^{N}$, where I_n is the measured intensity at position q_n, and the data points are assumed to be independent from each other. Then, applying the model $p(r)$ to the integral operator A in Eq. (265) yields for the discrete data set $\mathcal{D}$ the linear system

$$\mathbf{\Psi}\mathbf{w} = \mathbf{I}, \tag{267}$$

where $[\mathbf{I}]_n = I_n$, $[\mathbf{w}]_m = w_m$, and $[\mathbf{\Psi}]_{n,m} = \psi_m(q_n)$ with

$$\psi_m(q_n) := 4\pi \int_{\Xi} \phi_m(r) \frac{\sin(q_n r)}{q_n r} \, \mathrm{d}r. \tag{268}$$

[55] Eq. (265) can be derived from Eq. (5) if the orientationally averaged autocorrelation function of the excess SLD $\Delta\gamma(r) = \langle \Delta\gamma(\mathbf{r}) \rangle_{\Omega} = \left\langle \int \Delta\mathrm{SLD}(\mathbf{u}) \Delta\mathrm{SLD}(\mathbf{u}+\mathbf{r}) \, \mathrm{d}\mathbf{u} \right\rangle_{\Omega}$ is substituted in Eq. (5), and then using $\mathrm{d}\mathbf{r} = r^2 \sin(\theta) \, \mathrm{d}r \, \mathrm{d}\theta \, \mathrm{d}\phi$ and $\left\langle \exp(-i\mathbf{q}^T\mathbf{r}) \right\rangle_{\Omega} = \sin(qr)/(qr)$, see Eq. (18). Hence, the PDDF reads $\hat{p}(r) = \Delta\gamma(r) \cdot r^2$. If the intensity is normalized by a volume, the PDDF has to be divided by the volume, too.

Determining a unique parameter vector $\mathbf{w}$ in Eq. (267) is only possible if $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ and regular. To keep the subsequent discussion more general, consider $M \leq N$, i.e., $\mathbf{\Psi} \in \mathbb{R}^{N\times M}$. Then, Eq. (267) can be multiplied from the left with the transposed matrix of $\mathbf{\Psi}$, yielding

$$\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{w} = \mathbf{\Psi}^T\mathbf{I}, \tag{269}$$

where $\mathbf{\Psi}^T\mathbf{\Psi}$ is a $M \times M$ matrix. Finding a solution for $\mathbf{w}$ is easily possible if $\mathbf{\Psi}^T\mathbf{\Psi}$ is regular. If this is the case, it follows

$$\mathbf{w} = (\mathbf{\Psi}^T\mathbf{\Psi})^{-1}\mathbf{\Psi}^T\mathbf{I}, \tag{270}$$

where $\mathbf{\Psi}^\dagger := (\mathbf{\Psi}^T\mathbf{\Psi})^{-1}\mathbf{\Psi}^T$ is called (Moore-Penrose) pseudoinverse. $(\mathbf{\Psi}^T\mathbf{\Psi})^{-1}$ can be analyzed by performing a Singular Value Decomposition (SVD) [129, §2.6] of the matrix $\mathbf{\Psi}$. The SVD is a basic theorem from linear algebra that states that any matrix, e.g., $\mathbf{\Psi}$, can be decomposed into a product of three matrices

$$\mathbf{\Psi} = \mathbf{U}\mathbf{S}\mathbf{V}^T. \tag{271}$$

Herein, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M] \in \mathbb{R}^{N\times M}$ is an orthogonal matrix build of left singular vectors $\{\mathbf{u}_m\}_{m=1}^M$, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_M] \in \mathbb{R}^{M\times M}$ is an orthogonal matrix build of right singular vectors $\{\mathbf{v}_m\}_{m=1}^M$, and $\mathbf{S} \in \mathbb{R}^{M\times M}$ is a diagonal matrix of singular values $\{s_m\}_{m=1}^M$, i.e., $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_M)$, with $s_1 \geq s_2 \geq \cdots \geq s_m \geq 0$. The singular vectors are normalized to one. With this decomposition the inverse matrix can be written as

$$\begin{aligned}(\mathbf{\Psi}^T\mathbf{\Psi})^{-1} &= (\mathbf{V}\mathbf{S}\mathbf{U}^T\mathbf{U}\mathbf{S}\mathbf{V}^T)^{-1} \\ &= (\mathbf{V}\mathbf{S}\mathbf{S}\mathbf{V}^T)^{-1} \\ &= (\mathbf{V}^T)^{-1}\mathbf{S}^{-1}\mathbf{S}^{-1}\mathbf{V}^{-1} \\ &= \mathbf{V}\mathbf{S}^{-1}\mathbf{S}^{-1}\mathbf{V}^T,\end{aligned} \tag{272}$$

where $[\mathbf{S}^{-1}]_{mm} = [\mathrm{diag}(s_1^{-1}, \ldots, s_M^{-1})]_{mm} = s_m^{-1}$. Hence, the matrix $\mathbf{\Psi}^T\mathbf{\Psi}$ is regular if none of the singular values s_m is zero, otherwise the matrix is singular. In practice s_m can become very small, especially if the number of parameters M is large, which can then lead to huge errors in the solution. This fact can be exemplified via Eq. (280) (see the next section), which shows that the upper bound of the relative error in the solution is given by the product of the condition number times the relative error in the data. The condition number of the pseudoinverse is given in Eq. (287), hence [63, §4]

$$C(\mathbf{\Psi}) = \|\mathbf{\Psi}\|\,\|\mathbf{\Psi}^\dagger\| = s_1 \cdot s_M^{-1}, \tag{273}$$

where the Euclidean norm (2-norm) is taken. In practice, $C(\mathbf{\Psi})$ can be extremely large, e.g., 10^h with often $10 \lessapprox h \lessapprox 100$, which then yields a severely, numerically unstable solution. In a nested model class[56], $C(\mathbf{\Psi})$ increases with the number of freely adjustable parameters, i.e., if the model becomes more complex, and decreases if there are less parameters, i.e., if the model complexity decreases. Hence, a model is searched that is as less complex as possible under the constraint that it can explain the data — in accordance with the principle of Occam's razor, cf. section 2.2. This idea can be implemented via regularization methods and is already discussed in section 2.2 in the context of ML estimators, Eq. (46). Note that minimizing $\|\mathbf{I} - \mathbf{\Psi}\mathbf{w}\|^2$ with respect to $\mathbf{w}$ is equivalent to Eq. (269) [97, §6.9, theorem 1]. Hence, e.g., the regularized optimization problem [158, 159]

$$\arg\min_{\mathbf{w}} \left\{\|\mathbf{I} - \mathbf{\Psi}\mathbf{w}\|^2 + \gamma\|\mathbf{w}\|^2\right\} \tag{274}$$

[56] E.g., the polynomial class is a nested one: the polynomial of order m is contained in the polynomial of order $m+1$.

can be solved. However, it would additionally require to optimally determine the regularization parameter $\gamma > 0$, i.e., the model complexity, which is the topic of section 7.2. The optimal parameter vector in Eq. (274) is found by setting the first derivative of it, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}\left[(\mathbf{I}-\boldsymbol{\Psi}\mathbf{w})^T(\mathbf{I}-\boldsymbol{\Psi}\mathbf{w})+\gamma\mathbf{w}^T\mathbf{w}\right] = 2\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{w} - 2\boldsymbol{\Psi}^T\mathbf{I} + 2\gamma\mathbf{w}, \tag{275}$$

equal to zero, yielding

$$\mathbf{w}^\gamma = (\boldsymbol{\Psi}^T\boldsymbol{\Psi}+\gamma\mathbf{1})^{-1}\boldsymbol{\Psi}^T\mathbf{I}, \tag{276}$$

where $\mathbf{1}$ is the identity matrix of dimension $M \times M$. By virtue of the SVD, Eq. (271), this formula can be rewritten as

$$\begin{aligned}
\mathbf{w}^\gamma &= (\mathbf{V}\mathbf{S}\mathbf{S}\mathbf{V}^T+\gamma\mathbf{1})^{-1}\mathbf{V}\mathbf{S}\mathbf{U}^T\mathbf{I} \\
&= (\mathbf{V}^T)^{-1}(\mathbf{S}\mathbf{S}+\gamma\mathbf{1})^{-1}\mathbf{V}^{-1}\mathbf{V}\mathbf{S}\mathbf{U}^T\mathbf{I} \\
&= \mathbf{V}(\mathbf{S}\mathbf{S}+\gamma\mathbf{1})^{-1}\mathbf{S}\mathbf{U}^T\mathbf{I} \\
&= \sum_{m=1}^{M}\mathbf{v}_m\frac{s_m}{s_m^2+\gamma}\mathbf{u}_m^T\mathbf{I},
\end{aligned} \tag{277}$$

showing that the filter coefficient $s_m/(s_m^2+\gamma)$ is inherently 'stabilized' by means of γ if s_m is close to zero or zero (i.e., it prevents this term, as well as $\mathbf{w}^\gamma$, to become unreasonably large) and γ sufficiently large.

The optimal parameter vector can often easily be calculated, while the determination of the optimal model complexity is usually non-trivial. In this work now, this issue is considered in the context of statistical inference, see section 2.2, and the following aspects are discussed in the next subsections:

1. How to build an objective function in order to perform model selection?
2. Is there a method to get a numerically more stable solution (i.e., by using other regularization approaches)?

The main approaches to solve these two questions are coming from ideas of statistical inference, and not from a classical mathematical approach. Hence, the Bayesian inference is applied or the frequentist one. Indeed, from a Bayesian viewpoint there is no inverse problem only an inference one, cf. footnote 57 on page 124. In a frequentist approach, the goal is to find a model with good generalization properties, i.e., the optimal model should optimally predict unseen data. A classical mathematical approach would only look for a model that can explain the current data, and which is at the same time a more or less smooth function. Nevertheless, a short introduction to the mathematical theory of inverse problems is given in the next section, in order define some technical terms often used in this context, as well as to show how ill-posed problems (in the sense of Hadamard) become well-posed (in the sense of Nashed) if discretized.

7.1. Mathematical Introduction to Inverse Problems

As discussed in chapter 2.1 calculating the intensity for a given SLD is a direct problem, while the reverse procedure is often called inverse problem.[57] Mathematically, the direct problem as well as the inverse problem are formally given by means of an operator

$$\mathcal{K} : X \to Y \tag{278}$$

that acts between a set of causes X and a set of effects Y [59, 72]. Then, the direct problem is the evaluation of $\mathcal{K}$ for a given $f \in D(\mathcal{K}) \subset X$, i.e.,

$$\mathcal{K}(f) = g, \qquad f \in D(\mathcal{K}) \subset X, \quad g \in Y. \tag{279}$$

$D(\mathcal{K})$ is the domain of the operator $\mathcal{K}$. In the inverse problem an element $g \in Y$ is given and the task is to find a f such that Eq. (279) holds.

An important notion in the mathematical problem description is the technical term well-posedness. In the sense of Hadamard [60] a problem is well-posed if three conditions are fulfilled:

1. There is a solution for all admissible data. (Existence)
2. The solution is unique. (Uniqueness)
3. The solution depends continuously on the data. (Continuity)

In this section only operators from the set of linear and continuous operators, denoted by $\mathcal{L}(X, Y)$, are considered, which act between real Hilbert spaces of functions X and Y.

However, even if a problem is well-posed in the sense of Hadamard, it may be that small errors in the data lead to large errors in the solution, hence continuity is a necessary but not a sufficient condition for robustness [74, §1.3], [128]: Let $\mathcal{K} \in \mathcal{L}(X, Y)$ and assume that the inverse operator $\mathcal{K}^{-1}$ exists and is continuous, then a solution is called well-conditioned [74, §1.5] if small deviations δg in the datum g only lead to small deviations δf of the true solution f. Following procedure is applied in order to elucidate the instability issue: Since $\mathcal{K}$ and $\mathcal{K}^{-1}$ are linear and continuous operators we have for $\mathcal{K}f = g$ and $\delta f = \mathcal{K}^{-1}\,\delta g$ the inequalities $\|\mathcal{K}\|\|f\|_X \geq \|g\|_Y$ and $\|\delta f\|_X \leq \|\mathcal{K}^{-1}\|\,\|\delta g\|_Y$, which can be combined to yield

$$\frac{\|\delta f\|_X}{\|f\|_X} \leq C(\mathcal{K}) \frac{\|\delta g\|_Y}{\|g\|_Y}, \tag{280}$$

where (see, e.g., [74, §1.5])

$$C(\mathcal{K}) = \|\mathcal{K}\|\,\|\mathcal{K}^{-1}\| \geq 1 \tag{281}$$

and is called condition number. Hence the relative error in the solution, i.e., $\|\delta f\|_X / \|f\|_X$, is bounded by the product of the relative error in the data times the condition number. It follows that the inverse problem is well-conditioned if the condition number is close to one, and if it is considerably larger than one it is called ill-conditioned [74, §1.5].

[57] From a classical mathematical point of view, the problem of finding the causes given the effects can be thought as an inverse problem. However, from a Bayesian point of view [141, §6.4.2][77], the problem is not an inverse one, but an inference problem, in which a priori probabilities about hypotheses (causes) have to be adjusted according to some observed data (noisy effects), in order to infer the a posteriori probabilities of the hypotheses. Nevertheless, the fundamentals in this chapter shed some light on the intrinsic problems that arise when trying to solve the problem with 'classical' mathematics. Therefore, the notion inverse problem is kept.

In general, the datum $g \in Y$ does not have to lie in the range of the operator, $R(\mathcal{K})$, because $R(\mathcal{K}) \subset Y$, i.e., $\mathcal{K}$ can be non-surjective. In such a case an approximate solution might exist by means of the pseudosolution[58] f^+ given by [128]

$$\|\mathcal{K}f^+ - g\|_Y \leq \|\mathcal{K}f - g\|_Y \qquad \forall f \in D(\mathcal{K}) \tag{282}$$

or equivalently by

$$\mathcal{K}^*\mathcal{K}f^+ = \mathcal{K}^*g. \tag{283}$$

$\mathcal{K}^*$ is the adjoint operator of $\mathcal{K}$. If there is a solution it is not unique unless the nullspace of $\mathcal{K}$, $N(\mathcal{K})$, is trivial, i.e., $\mathcal{K}$ is injective. Then, $\mathcal{K}^*\mathcal{K}$ is regular such that the inverse $(\mathcal{K}^*\mathcal{K})^{-1}$ exists [97, §6.9]. Getting a unique solution for a non-trivial $N(\mathcal{K})$ can be easily accomplished [128]: since $X = N(\mathcal{K})^\perp \oplus N(\mathcal{K})$ a pseudosolution f^+ can uniquely be decomposed as $f^+ = f^\| + f^\dagger$, where $f^\dagger \in N(\mathcal{K})^\perp$ and $f^\| \in N(\mathcal{K})$, and therefore by only taking the contribution $f^\dagger$ a unique pseudosolution is selected. $f^\dagger$ can be selected by choosing the pseudosolution with minimum norm, i.e.,

$$f^\dagger = \arg\inf_{f^+}\{\|f^+\|_X : \mathcal{K}^*\mathcal{K}f^+ = \mathcal{K}^*g\}, \tag{284}$$

since

$$\|f^+\|_X^2 = \|f^\| + f^\dagger\|_X^2 = \|f^\|\|_X^2 + \|f^\dagger\|_X^2 \geq \|f^\dagger\|_X^2. \tag{285}$$

$f^\dagger$ is called generalized pseudosolution. A solution of Eq. (282),(284) is only guaranteed to exist if $R(\mathcal{K})$ is closed (cf. [97, §3, theorem 1,2]), because then $R(\mathcal{K}) \oplus R(\mathcal{K})^\perp = Y$, otherwise $R(\mathcal{K}) \oplus R(\mathcal{K})^\perp$ is a proper subset of Y. Define $D(\mathcal{K}^\dagger) := R(\mathcal{K}) \oplus R(\mathcal{K})^\perp \subset Y$, then the operator $\mathcal{K}^\dagger : D(\mathcal{K}^\dagger) \to X$ defined by the mapping

$$\mathcal{K}^\dagger g = f^\dagger \tag{286}$$

is called generalized inverse operator, and the appropriate condition number reads (see, e.g., [74, §1.5])

$$C(\mathcal{K}) = \|\mathcal{K}\|\,\|\mathcal{K}^\dagger\| \geq 1. \tag{287}$$

It can be shown that $\mathcal{K}^\dagger$ is linear [128, Eq. (15)-(17)], and $\mathcal{K}^\dagger$ is continuous if and only if $R(\mathcal{K})$ is closed [128, theorem 3]. Thus, the well-posedness conditions of Hadamard are all together fulfilled by $\mathcal{K}^\dagger$ if and only if $R(\mathcal{K})$ is closed. Therefore, in the sense of Nashed, a problem is well-posed if and only if $R(\mathcal{K})$ is closed [59, definition 3.10].

As a concluding result from the discussion above, it can be summarized that if the range of a linear and continuous operator is closed, a unique solution exists, which depends continuously on the data (hence the well-posedness conditions are altogether fulfilled), and is given by the generalized pseudosolution, Eq. (284). This is an important result if applied to practical problems involving a finite number of measured data points: In this case the range of the operator $\mathcal{K}$ is a finite dimensional space, and since in a normed linear space any finite-dimension subspace is closed [97, §2.12, theorem 1,2] it follows that the inverse problem is well-posed in the sense of Nashed. However, the problem is usually ill-conditioned, hence the solution is non-robust, i.e., it is numerically unstable [74, §1.5]. In order to achieve a robust solution regularization can be applied as already explained in the previous section, cf. Eq. (274).

[58] Also called normal solution.

7.2. Objective Function and Model Selection

Consider a discrete and noisy scattering data set $\mathcal{D}$ as in section 7, but now, it is assumed that the errors (standard deviations) σ_n in the measured intensities I_n are available, thus $\mathcal{D} := \{(q_n, I_n, \sigma_n)\}_{n=1}^N$, where it is assumed that the data points are statistically independent from each other. Then, the optimization problem from Eq. (274) becomes

$$\mathbf{w}^\gamma = \arg\min_{\mathbf{w}} \left\{ J(\mathbf{w}) := \frac{1}{2} \left\| \mathbf{\Sigma}^{-1/2} (\mathbf{I} - \mathbf{\Psi}\mathbf{w}) \right\|^2 + \frac{\gamma}{2} \|\mathbf{L}\mathbf{w}\|^2 \right\}, \tag{288}$$

where

$$\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2) \tag{289}$$

is the error covariance matrix, and $\mathbf{L}$ is an additionally introduced regularization matrix, which extracts certain features from $\mathbf{w}$ that are to be penalized. The factors $1/2$ in Eq. (288) are introduced for consistency with section 2.2. Eq. (288) is known as Tikhonov regularization (see, e.g., [158, 159]) or ridge regression [71]. It is a weighted linear least squares problem[59], which has a unique solution if the intersection of the nullspaces $N(\mathbf{\Sigma}^{-1/2}\mathbf{\Psi})$ and $N(\sqrt{\gamma}\mathbf{L})$ is trivial, i.e.,

$$N(\mathbf{\Sigma}^{-1/2}\mathbf{\Psi}) \cap N(\sqrt{\gamma}\mathbf{L}) = \{\varnothing\}, \tag{291}$$

otherwise any non-trivial $\tilde{\mathbf{w}} \in N(\mathbf{\Sigma}^{-1/2}\mathbf{\Psi}) \cap N(\sqrt{\gamma}\mathbf{L})$ could be added to the solution while keeping $J(\mathbf{w}^\gamma + \tilde{\mathbf{w}})$ in Eq. (288) unchanged. Here, $\mathbf{L}$ is taken as discrete second order derivative operator[60,61] with zero boundary conditions [63, §8.1], i.e.,

$$\mathbf{L} = \mathbf{L}_{D^2} := \begin{pmatrix} -2 & 1 & & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & & \ddots & & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & \end{pmatrix} \in \mathbb{R}^{M \times M}, \tag{292}$$

and

$$\|\mathbf{L}_{D^2}\mathbf{w}\|^2 = \sum_{m=1}^{M} (w_{m-1} - 2w_m + w_{m+1})^2, \quad \text{where } w_0 = w_{M+1} = 0. \tag{293}$$

$\mathbf{L}_{D^2}$ has full rank, hence for $\gamma > 0$ the nullspace $N(\sqrt{\gamma}\mathbf{L}_{D^2})$ is trivial and therefore the solution is unique.

Tikhonov [158] has considered Eq. (288) as a mathematical approach to solve linear ill-posed problems (as given in discrete form in Eq. (267). Later, Glatter [49] independently invented the Tikhonov formula for stabilizing the inverse Fourier transform. In this classical mathematical approach the principle idea is to find a solution that explains the data and at the same time is smooth. In the Tikhonov formula, Eq. (288), the regularization parameter γ

[59] In order to see that Eq. (288) is a weighted linear least squares problem, the formula can be arranged as

$$\mathbf{w}^\gamma = \arg\min_{\mathbf{w}} \left\| \begin{pmatrix} \mathbf{\Sigma}^{-1/2}\mathbf{\Psi} \\ \sqrt{\gamma}\mathbf{L} \end{pmatrix} \mathbf{w} - \begin{pmatrix} \mathbf{\Sigma}^{-1/2}\mathbf{I} \\ \mathbf{0} \end{pmatrix} \right\|^2. \tag{290}$$

[60] Depending on the boundary conditions the second derivative matrix can be defined differently, see, e.g., [63, §8.1 and §8.2]. Here, the boundary condition is that the solution shall be zero outside of the considered interval, hence $w_0 = w_{M+1} = 0$.

[61] Another regularization matrix is given in the work of Glatter [49, 26], which is reproduced in the appendix F.1.

controls the model complexity, and needs to be adjusted optimally, i.e., by means of performing model selection. There are many different approaches for performing model selection, e.g., in the context of the IFT, by means of a 'stability plot' / 'point of inflection'[62] [49] or perceptual criteria [149], but they are introduced ad-hoc, and have no deep fundamental basis. The main problem with such methods is that usually the solution is not unique and therefore the user has a certain freedom to select an 'optimal' solution, thus it is not an objective method (instead it is a subjective one, and hence different users might select different solutions). Moreover, with such methods inter-model comparison is not possible, and additionally, a few methods might even not yield a solution in some cases.

Alternatively the Tikhonov formula, Eq. (288), can be interpreted in the Bayesian inference framework as well as within a frequentist one, cf. section 2.2, which then allow to build objective functions for model selection based on these approaches. These objective functions even allow to perform inter-model comparison — see section (7.2.3).

In the frequentist approach, the empirical error should not be minimized directly, since overfitting would occur if the model complexity is too high, which is a natural property of a maximum likelihood (ML) estimator — see section 2.2. Hence, the empirical error term of a relatively complex model is conveniently augmented by a regularization term (yielding the Tikhonov formula Eq. (288)) in order to control the model complexity. Then, the model complexity is fine-tuned by adjusting the regularization parameter γ. In this frequentist approach γ should be adjusted such that it yields a model with optimal generalization property, i.e., the selected model should predict (in average) unseen i.i.d. data sets as well as possible, hence the generalization error is the desired objective function. This approach is followed in the next subsection.

In the Bayesian approach, already used earlier in the work of Hansen [64], a priori hypotheses (model parameters $\mathbf{w}$, hyperparameter γ) are adjusted according to an observed data set $\mathcal{D}$, yielding an a posteriori distribution of the hypotheses given the observed data set — cf. section 2.2. Then, in the hierarchical inference approach, Eq. (288) gives the most probable set of parameters (MAP solution) for a given but unknown hyperparameter γ that controls the uncertainty of the parameters; and the marginal likelihood $\mathrm{pdf}(\gamma|\mathcal{D})$, also called evidence, yields the appropriate objective function for the model complexity, and which has to be maximized with respect to γ to select the MAP solution of γ. This approach is followed in subsection after next.

Before the objective functions for model selection are given in the context of the two different inference approaches, the Tikhonov optimization problem in Eq. (288) is extended for the case of a constant background intensity $I_{\mathrm{bkg.}}$. Since $I_{\mathrm{bkg.}}$ yields an additional intensity contribution that may have in principle any value, it should not be penalized by the regularization term. Thus, the optimization problem reads

$$\arg\min_{\mathbf{w},I_{\mathrm{bkg.}}} \left\{ \tilde{J}(\mathbf{w}, I_{\mathrm{bkg.}}\mathbf{e}_1) := \frac{1}{2}\left\|\boldsymbol{\Sigma}^{-1/2}\big(\mathbf{I} - (\boldsymbol{\Psi}\mathbf{w} + I_{\mathrm{bkg.}}\mathbf{e}_1)\big)\right\|^2 + \frac{\gamma}{2}\|\mathbf{L}\mathbf{w}\|^2 \right\}, \tag{294}$$

where $\mathbf{e}_1 = [1, 1, \ldots, 1]^T \in \mathbb{R}^N$. The derivative of $\tilde{J}(\mathbf{w}, I_{\mathrm{bkg.}}\mathbf{e}_1)$ with respect to $I_{\mathrm{bkg.}}\mathbf{e}_1$ gives

$$\begin{aligned} \frac{\partial \tilde{J}(\mathbf{w}, I_{\mathrm{bkg.}}\mathbf{e}_1)}{\partial (I_{\mathrm{bkg.}}\mathbf{e}_1)} &= \frac{\partial}{\partial (I_{\mathrm{bkg.}}\mathbf{e}_1)} \left((\mathbf{I} - \boldsymbol{\Psi}\mathbf{w} - I_{\mathrm{bkg.}}\mathbf{e}_1)^T \boldsymbol{\Sigma}^{-1} (\mathbf{I} - \boldsymbol{\Psi}\mathbf{w} - I_{\mathrm{bkg.}}\mathbf{e}_1) \right) \\ &= 2\boldsymbol{\Sigma}^{-1}(I_{\mathrm{bkg.}}\mathbf{e}_1) - 2\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \boldsymbol{\Psi}\mathbf{w}), \end{aligned} \tag{295}$$

[62] The 'stability plot' / 'point of inflection' method is as follows: The empirical error and the regularization term (without the regularization parameter) are plotted versus the regularization parameter. Then, an optimal regularization parameter can be obtained from the position of a point that lies within a plateau of the regularization term, where additionally the empirical error is small.

and setting the result equal to zero yields:

$$I_{\text{bkg.}}\mathbf{e}_1 = (\mathbf{I} - \mathbf{\Psi}\mathbf{w}) \tag{296}$$

$$\Rightarrow \qquad I_{\text{bkg.}}N = \mathbf{e}_1^T(\mathbf{I} - \mathbf{\Psi}\mathbf{w})$$

$$\Leftrightarrow \qquad I_{\text{bkg.}} = \frac{1}{N}\sum_{n=1}^{N} I_n - \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} \Psi_m(q_n)w_m, \tag{297}$$

where Eq. (296) is multiplied from the left with $\mathbf{e}_1^T$ in order to get Eq. (297). Substituting $I_{\text{bkg.}}$ back into Eq. (294) gives the remaining optimization problem

$$\arg\min_{\mathbf{w}} \left\{ \tilde{J}(\mathbf{w}) := \frac{1}{2}\left\|\mathbf{\Sigma}^{-1/2}(\mathbf{I}_c - \mathbf{\Psi}_c\mathbf{w})\right\|^2 + \frac{\gamma}{2}\|\mathbf{L}\mathbf{w}\|^2 \right\}, \tag{298}$$

where

$$\mathbf{I}_c := \mathbf{I} - \overline{I}\mathbf{e}_1 \tag{299}$$

with

$$\begin{aligned} \overline{I} :&= \frac{1}{N}\sum_{n=1}^{N} I_n \\ &= \frac{1}{N}\mathbf{e}_1^T\mathbf{I}, \end{aligned} \tag{300}$$

and

$$\begin{aligned} \mathbf{\Psi}_c :&= \begin{pmatrix} \Psi_1(q_1) - \overline{\Psi}_1 & \cdots & \Psi_M(q_1) - \overline{\Psi}_M \\ \vdots & \ddots & \vdots \\ \Psi_1(q_N) - \overline{\Psi}_1 & \cdots & \Psi_M(q_N) - \overline{\Psi}_M \end{pmatrix} \\ &= \left(\mathbf{1} - \frac{1}{N}\mathbf{e}_1\mathbf{e}_1^T\right)\mathbf{\Psi} \end{aligned} \tag{301}$$

with

$$\overline{\Psi}_m := \frac{1}{N}\sum_{n=1}^{N} \Psi_m(q_n). \tag{302}$$

$\mathbf{1} \in \mathbb{R}^{N\times N}$ is the identity matrix. Eq. (298) shows that the optimization problem in Eq. (288) is not different from the one in Eq. (288), it only requires a centering of $\mathbf{I}$ and $\mathbf{\Psi}$ via Eq. (299),(301) in advance. After the optimization problem in Eq. (298) is solved, the background intensity can be calculated for the optimal parameter vector $\mathbf{w}^*$ via Eq. (297).

7.2.1. Objective Function within a Frequentist Approach

As explained in section 2.2, in a frequentist approach the Kullback-Leibler divergence can be minimized, or equivalently the generalization error, in order to perform model selection. Hence here, the generalization error is used as the objective function for performing model selection.

The generalization error with respect to the conditional distribution $\widehat{\text{pdf}}(\mathbf{I}|\mathbf{q})$, being a heteroscedastic Gaussian likelihood function consisting of N independently distributed data points, reads

$$\mathcal{E}'_{\widehat{\text{pdf}}(\mathbf{I}|\mathbf{q})}(p) = \sum_{n=1}^{N} \int \widehat{\text{pdf}}(I_n|q_n) \left(\frac{\left(I_n - (Ap)(q_n)\right)^2}{2\sigma^2(q_n)} \right) \mathrm{d}I_n, \tag{303}$$

where terms independent of the model function $p(r)$ are neglected as they are not relevant for a minimization of the generalization error with respect to $p(r)$. Eq. (303) has a minimum if $(Ap)(q_n) = \int I_n \widehat{\mathrm{pdf}}(I_n|q_n)\ \mathrm{d}I_n\ \forall n$, cf. Eq. (52).

In this work now, the generalization error is estimated by means of the cross-validation method, which is already described in algorithm 1 of section 2.2. The factor 1/2 in Eq. (303) is replaced by $1/N$ in order to make the analysis independent of the data set size — it is just a matter of convenience —, and thus the optimal estimate of the generalization error should be close to one. However, the question arises in how many different data sets the original data set has to be divided in the cross-validation method. Consider the two extreme cases: (*i*) If the data set is divided into two disjoint sets, then there are only two training sets and two validation sets, yielding training sets that might be too small or too different from the original complete data set. Hence, statistics might be poor. (*ii*) If the data set is divided into N disjoints sets (N is the number of data points, and the method is then called leave-one-out cross-validation), then there are N training sets, but which are all very similar, hence statistics might be poor again[63]. In practice it is found that often 5- or 10-folded cross-validation is a good choice. Last but not least, it is noted that the cross-validation method can be easily applied to the two considered cases: scattering data containing a background intensity and scattering data without a background.

The optimal parameter vector in Eq. (288) (and similarly in Eq. (298)) is found by setting the first derivative of $J(\mathbf{w})$, i.e.,

$$\nabla J(\mathbf{w}) = \left(\mathbf{\Psi}^T\mathbf{\Sigma}^{-1}\mathbf{\Psi}\mathbf{w} - \mathbf{\Psi}^T\mathbf{\Sigma}^{-1}\mathbf{I}\right) + \gamma\mathbf{L}^T\mathbf{L}\mathbf{w}, \tag{304}$$

equal to zero, yielding

$$\mathbf{w}^\gamma = (\mathbf{\Psi}^T\mathbf{\Sigma}^{-1}\mathbf{\Psi} + \gamma\mathbf{L}^T\mathbf{L})^{-1}\mathbf{\Psi}^T\mathbf{\Sigma}^{-1}\mathbf{I}. \tag{305}$$

As outlined in section 2.2 parameter uncertainties can be determined from the distribution of parameters obtained from i.i.d. data sets (which might have been obtained from repeating the experiment or by applying the bootstrap method). However, here a more convenient procedure is considered, which builds on the idea of analyzing the curvature of the likelihood at its maximum likelihood value, and which is described in section 3.2. Unfortunately, this approach would require calculating the inverse of the Hessian matrix of the negative logarithm of the likelihood function, which might give unreasonably large uncertainties, because the likelihood function describes an inverse set-up, and thus the Hessian matrix can be singular or close to singular. Moreover, here a regularized solution $\mathbf{w}^\gamma$ is used, which is in general not the maximum likelihood solution, hence the likelihood is not approximated around a maximum, as required. Both problems can be circumvented if the regularization term is also taken into account. Accordingly, the Hessian matrix is calculated of $J(\mathbf{w})$ at the minimum position $\mathbf{w}^\gamma$

$$\nabla\nabla J(\mathbf{w})\big|_{\mathbf{w}=\mathbf{w}^\gamma} = \nabla\nabla J(\mathbf{w}) = \mathbf{\Psi}^T\mathbf{\Sigma}^{-1}\mathbf{\Psi} + \gamma\mathbf{L}^T\mathbf{L}. \tag{306}$$

Hence this Hessian matrix is used as a substitute and the parameter uncertainties read $\Delta w_m = [\Delta\mathbf{w}]_m = \sqrt{[(\nabla\nabla J(\mathbf{w}))^{-1}]_{mm}}$, $1 \le m \le M$. This approach is introduced ad-hoc in the frequentist approach, therefore Δw_m are uncertainties which are not easy to interpret. However, if the perspective is changed to the Bayesian approach (cf. next section), then Δw_m

[63] The issue of finding the optimal data partition size is related to the bias-variance tradeoff of frequentist statistics, see, e.g., [111, §6.4.4]: Let $\mathcal{D}$ a data set and $\theta(\mathcal{D})$ an estimator of a quantity $\hat{\theta}$. Then, the mean squared error $\left\langle(\theta(\mathcal{D}) - \hat{\theta})^2\right\rangle_{\mathcal{D}}$ is equal to $\left\langle(\theta(\mathcal{D}) - \overline{\theta})^2\right\rangle_{\mathcal{D}} + (\overline{\theta} - \hat{\theta})^2$, where $\overline{\theta} = \langle\theta(\mathcal{D})\rangle_{\mathcal{D}}$. Hence, in practice one should use an estimator $\theta(\mathcal{D})$ that has a small variance $\left\langle(\theta(\mathcal{D}) - \overline{\theta})^2\right\rangle_{\mathcal{D}}$ as well as a small bias squared $(\overline{\theta} - \hat{\theta})^2$ in order get an overall small mean squared error.

can be interpreted as the standard deviation of the m-the parameter within the a posteriori distribution, which gives the uncertainty Δw_m a clear meaning. Therefore, this method is applied, here.

7.2.2. Objective Function within a Bayesian Approach

In order to define the objective function for the Tikhonov formula Eq. (288) within a Bayesian approach, the formula is derived for it: Consider as in the previous, frequentist approach that the noise model is a heteroscedastic Gaussian one with mean $\mathbf{\Psi w}$ and covariance $\mathbf{\Sigma}$. Then, the likelihood function reads

$$\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\mathbf{w}) = \mathcal{N}(\mathbf{I}|\mathbf{\Psi w},\mathbf{\Sigma}) \propto \exp\left\{-\frac{1}{2}\left\|\mathbf{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{\Psi w})\right\|^2\right\}. \tag{307}$$

If also the a priori distribution is a Gaussian distribution with zero mean and covariance $\left(\gamma \mathbf{L}^T\mathbf{L}\right)^{-1}$, i.e.,

$$\mathrm{pdf}(\mathbf{w}|\gamma) = \mathcal{N}\left(\mathbf{w}|\mathbf{0},(\gamma\mathbf{L}^T\mathbf{L})^{-1}\right) \propto \exp\left\{-\frac{1}{2}\gamma\|\mathbf{Lw}\|^2\right\}, \tag{308}$$

then by means of the formulas from appendix E.1 the a posteriori distribution of the parameters is a Gaussian one and reads as follows

$$\begin{aligned}\mathrm{pdf}(\mathbf{w}|\mathbf{q},\boldsymbol{\sigma},\mathbf{I},\gamma) &= \frac{\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\mathbf{w})\mathrm{pdf}(\mathbf{w}|\gamma)}{\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma)} \\ &= \mathcal{N}\left(\mathbf{w}|\mathbf{B}\mathbf{\Psi}^T\mathbf{\Sigma}^{-1}\mathbf{I},\mathbf{B}\right) \\ &\propto \exp\left\{-\frac{1}{2}\left\|\mathbf{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{\Psi w})\right\|^2 - \frac{1}{2}\gamma\|\mathbf{Lw}\|^2\right\},\end{aligned} \tag{309}$$

where

$$\mathbf{B} = \left(\mathbf{\Psi}^T\mathbf{\Sigma}^{-1}\mathbf{\Psi} + \gamma\mathbf{L}^T\mathbf{L}\right)^{-1}. \tag{310}$$

From the a posteriori distribution it can be seen that finding the maximum a posteriori (MAP) solution $\mathbf{w}^\gamma = \arg\max_{\mathbf{w}} \mathrm{pdf}(\mathbf{w}|\mathbf{q},\boldsymbol{\sigma},\mathbf{I},\gamma)$ is equivalent to the minimization of $J(\mathbf{w})$ in Eq. (288), since the exponential function is strictly monotonically increasing. In this setting γ can be interpreted as a hyperparameter, cf. section 2.2, and similarly the a posteriori distribution of γ on a higher hierarchical level of model specification can be calculated as

$$\mathrm{pdf}(\gamma|\mathbf{q},\boldsymbol{\sigma},\mathbf{I}) = \frac{\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma)\mathrm{pdf}(\gamma)}{\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma})}. \tag{311}$$

$\mathrm{pdf}(\gamma)$ is the a priori distribution of γ and $\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma}) = \int \mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma)\mathrm{pdf}(\gamma)\,\mathrm{d}\gamma$. If $\mathrm{pdf}(\gamma)$ is assumed to be constant[64] it follows $\mathrm{pdf}(\gamma|\mathbf{q},\boldsymbol{\sigma},\mathbf{I}) \propto \mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma)$, i.e., the a posteriori distribution in the second level is proportional to the marginal likelihood (evidence) of the first level in Eq. (309). Then, an optimal regularization parameter can be chosen as $\gamma^* = \arg\max_\gamma \mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma)$, i.e., one that maximizes the evidence in Eq. (309). γ^* gives the most probable model. The evidence can be expressed via the formulas in appendix E.1, as

$$\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma) = \mathcal{N}(\mathbf{I}|\mathbf{0},\mathbf{W}), \tag{312}$$

[64] In this case the prior $\mathrm{pdf}(\gamma)$ is not normalized, and hence is called improper (see, e.g., [16, §2.4.2]), but it can be used here anyway, since the a posteriori distribution can be correctly normalized.

where

$$\mathbf{W} = \mathbf{\Sigma} + \mathbf{\Psi}(\gamma \mathbf{L}^T \mathbf{L})^{-1} \mathbf{\Psi}^T. \tag{313}$$

The evidence is the desired objective function in the Bayesian approach for determining the optimal model (complexity). However, in practice it is more convenient to maximize the logarithm of the evidence (because of numerical stability), i.e., to maximize

$$\begin{aligned} Q_B(\gamma) &= \ln \mathrm{pdf}(\mathbf{I}|\mathbf{q}, \boldsymbol{\sigma}, \gamma) \\ &= -\frac{1}{2} \left(N \ln(2\pi) + \ln(\det(\mathbf{W})) + \mathbf{I}^T \mathbf{W}^{-1} \mathbf{I} \right), \end{aligned} \tag{314}$$

and we also call this expression evidence.

Given a certain γ, appropriate standard deviations of the model parameters can be obtained from the covariance matrix $\mathbf{B}$ in Eq. (310) of the a posteriori distribution $\mathrm{pdf}(\mathbf{w}|\mathbf{q}, \boldsymbol{\sigma}, \mathbf{I}, \gamma)$ via $[\Delta \mathbf{w}]_m = \sqrt{[\mathbf{B}]_{mm}}$, $1 \leq m \leq M$.

In case the intensity data contain an unknown background intensity $I_{\mathrm{bkg.}}$, the likelihood function reads

$$\begin{aligned} \widetilde{\mathrm{pdf}}(\mathbf{I}|\mathbf{q}, \boldsymbol{\sigma}, \mathbf{w}) &= \mathcal{N}(\mathbf{I}|\mathbf{\Psi}\mathbf{w} + I_{\mathrm{bkg.}}(\mathbf{w})\mathbf{e}_1, \mathbf{\Sigma}) \\ &\propto \exp\left\{ -\frac{1}{2} \left\| \mathbf{\Sigma}^{-1/2} (\mathbf{I} - (\mathbf{\Psi}\mathbf{w} + I_{\mathrm{bkg.}}(\mathbf{w})\mathbf{e}_1)) \right\|^2 \right\}. \end{aligned} \tag{315}$$

In this formula, it is considered that $I_{\mathrm{bkg.}}$ is also distributed, since unknown, and in general it can be considered that $I_{\mathrm{bkg.}}$ is actually a function of $\mathbf{w}$ — expressing the idea that $I_{\mathrm{bkg.}}$ shall be optimally selected for each parameter vector $\mathbf{w}$. $I_{\mathrm{bkg.}}(\mathbf{w})$ is determined as follows: Firstly, consider the a priori distribution $\mathrm{pdf}(\mathbf{w}|\gamma)$ from Eq. (308), then the a posteriori distribution reads

$$\begin{aligned} \widetilde{\mathrm{pdf}}(\mathbf{w}|\mathbf{q}, \boldsymbol{\sigma}, \mathbf{I}, \gamma) &= \frac{\widetilde{\mathrm{pdf}}(\mathbf{I}|\mathbf{q}, \boldsymbol{\sigma}, \mathbf{w}) \mathrm{pdf}(\mathbf{w}|\gamma)}{\widetilde{\mathrm{pdf}}(\mathbf{I}|\mathbf{q}, \boldsymbol{\sigma}, \gamma)} \\ &\propto \exp\left\{ -\frac{1}{2} \left\| \mathbf{\Sigma}^{-1/2} (\mathbf{I} - (\mathbf{\Psi}\mathbf{w} + I_{\mathrm{bkg.}}(\mathbf{w})\mathbf{e}_1)) \right\|^2 - \frac{1}{2} \gamma \|\mathbf{L}\mathbf{w}\|^2 \right\}. \end{aligned} \tag{316}$$

Next, for a certain $\mathbf{w}$ the background $I_{\mathrm{bkg.}}(\mathbf{w})$ has to be determined optimally, which can be achieved by taking the negative logarithm of Eq. (316) yielding the cost function $\tilde{J}(\mathbf{w}, I_{\mathrm{bkg.}}\mathbf{e}_1)$, where $\tilde{J}$ is defined in Eq. (294). Optimizing for $I_{\mathrm{bkg.}}(\mathbf{w})$, as similarly done in Eq. (295)-(297), yields the background intensity in dependence of $\mathbf{w}$

$$\begin{aligned} I_{\mathrm{bkg.}}(\mathbf{w}) &= \frac{1}{N} \mathbf{e}_1^T (\mathbf{I} - \mathbf{\Psi}\mathbf{w}) \\ &= \overline{I} - \frac{1}{N} \mathbf{e}_1^T \mathbf{\Psi}\mathbf{w}, \end{aligned} \tag{317}$$

and $\overline{I}$ is defined in Eq. (300). Back-substitution of this result into the likelihood function Eq. (315) yields

$$\begin{aligned} \widetilde{\mathrm{pdf}}(\mathbf{I}|\mathbf{q}, \boldsymbol{\sigma}, \mathbf{w}) &= \mathcal{N}\big(\mathbf{I}|\mathbf{\Psi}\mathbf{w} + \overline{I}\mathbf{e}_1 - \frac{1}{N}\mathbf{e}_1\mathbf{e}_1^T\mathbf{\Psi}\mathbf{w}, \mathbf{\Sigma}\big) \\ &= \mathcal{N}\Big(\mathbf{I}\big|\big(\mathbf{1} - \frac{1}{N}\mathbf{e}_1\mathbf{e}_1^T\big)\mathbf{\Psi}\mathbf{w} + \overline{I}\mathbf{e}_1, \mathbf{\Sigma}\Big) \\ &= \mathcal{N}(\mathbf{I}|\mathbf{\Psi}_c\mathbf{w} + \overline{I}\mathbf{e}_1, \mathbf{\Sigma}) \\ &\propto \exp\left\{ -\frac{1}{2} \Big\| \mathbf{\Sigma}^{-1/2} \Big(\mathbf{I} - \big(\mathbf{\Psi}_c\mathbf{w} + \overline{I}\mathbf{e}_1 \big) \Big) \Big\|^2 \right\} \\ &\propto \exp\left\{ -\frac{1}{2} \Big\| \mathbf{\Sigma}^{-1/2} \big(\mathbf{I}_c - \mathbf{\Psi}_c\mathbf{w} \big\|^2 \right\}, \end{aligned} \tag{318}$$

where $\mathbf{\Psi}_c$ is defined in Eq. (301), and $\mathbf{I}_c$ is defined in Eq. (299). Combining this likelihood function with the a priori distribution in Eq. (308) and applying the formulas in appendix E.1, gives the a posteriori distribution

$$\begin{aligned}\widetilde{\mathrm{pdf}}(\mathbf{w}|\mathbf{q},\boldsymbol{\sigma},\mathbf{I},\gamma) &= \frac{\widetilde{\mathrm{pdf}}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\mathbf{w})\mathrm{pdf}(\mathbf{w}|\gamma)}{\widetilde{\mathrm{pdf}}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma)} \\ &= \mathcal{N}(\mathbf{w}|\tilde{\mathbf{B}}\mathbf{\Psi}_c^T\mathbf{\Sigma}^{-1}(\mathbf{I}-\overline{I}\mathbf{e}_1),\tilde{\mathbf{B}}) \\ &= \mathcal{N}(\mathbf{w}|\tilde{\mathbf{B}}\mathbf{\Psi}_c^T\mathbf{\Sigma}^{-1}\mathbf{I}_c,\tilde{\mathbf{B}}) \\ &\propto \exp\left\{-\frac{1}{2}\left\|\mathbf{\Sigma}^{-1/2}\big(\mathbf{I}_c-\mathbf{\Psi}_c\mathbf{w}\big)\right\|^2-\frac{1}{2}\gamma\|\mathbf{L}\mathbf{w}\|^2\right\},\end{aligned} \tag{319}$$

where

$$\tilde{\mathbf{B}} = \left(\gamma\mathbf{L}^T\mathbf{L}+\mathbf{\Psi}_c^T\mathbf{\Sigma}^{-1}\mathbf{\Psi}_c\right)^{-1}. \tag{320}$$

Applying the formulas in appendix E.1 again, the evidence reads

$$\widetilde{\mathrm{pdf}}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma) = \mathcal{N}\big(\mathbf{I}|\overline{I}\mathbf{e}_1,\tilde{\mathbf{W}}\big), \tag{321}$$

where

$$\tilde{\mathbf{W}} = \mathbf{\Sigma}+\mathbf{\Psi}_c(\gamma\mathbf{L}^T\mathbf{L})^{-1}\mathbf{\Psi}_c^T. \tag{322}$$

Similarly as above, the objective function is taken as

$$\begin{aligned}\tilde{Q}_B(\gamma) &= \ln\widetilde{\mathrm{pdf}}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\gamma) \\ &= -\frac{1}{2}\left(N\ln(2\pi)+\ln(\det(\tilde{\mathbf{W}}))+\mathbf{I}_c^T\tilde{\mathbf{W}}^{-1}\mathbf{I}_c\right).\end{aligned} \tag{323}$$

Thus, in case the data contain a background, the evidence is calculated similarly as in the case where there is no background contained in the data, only the intensity vector $\mathbf{I}$ and the matrix $\mathbf{\Psi}$ have to be substituted by their centered versions, i.e., $\mathbf{I}_c$ and $\mathbf{\Psi}_c$.

Standard deviations of parameters are obtained as $[\Delta\mathbf{w}]_m = \sqrt{[\tilde{\mathbf{B}}]_{mm}}$, $1 \le m \le M$.

7.2.3. Evaluations

Scattering data are simulated in order to analyze the performance of the two different objective function approaches. For this purpose the scattering of encapsulated silica nanoparticles is considered (core-shell objects). The intensity is calculated for the case of neutron scattering, and three different matrix SLDs are assumed: in one case the matrix is contrast matched to the silica particles, in another case the matrix is contrast matched to the shell, and in the last case the zero average contrast (ZAC) condition is assumed, i.e., the matrix is adjusted such that it matches in average the average SLD of the encapsulated particles.
In detail, it is considered that the silica nanoparticles are lognormally distributed with mean radius 8.6 nm and polydispersity 0.14, the shell thickness has a Gaussian distribution with mean 4.4 nm and polydispersity 0.1, the overall volume fraction of the encapsulated particles is 0.015, and the SLDs are considered to be:

SLD of silica nanoparticles: $3.45 \cdot 10^{-4}\,\mathrm{nm}^{-2}$

SLD of the shell: $3.01 \cdot 10^{-5}\,\mathrm{nm}^{-2}$ (which is the SLD of DOPC, cf. section 4.4.3)

SLDs of the matrix (three cases are considered):

(i) **Matching core SLD:** SLD of the matrix is $3.59 \cdot 10^{-4}\,\mathrm{nm}^{-2}$, such that it approximately matches the SLD of the silica nanoparticles.

(*ii*) **Matching shell SLD:** SLD of the matrix is $3.07 \cdot 10^{-5}\,\mathrm{nm}^{-2}$, such that it approximately matches the SLD of the shell of the encapsulated nanoparticles.

(*iii*) **ZAC condition:** SLD of the matrix is $1.38 \cdot 10^{-4}\,\mathrm{nm}^{-2}$, such that it approximately matches the average SLD of the encapsulated particles.

Such a system is interesting, since it appears in practice (cf. section 4.4.3) and at the same time allows to model different PDDFs via contrast variation, i.e, by changing the D_2O/H_2O ratio of the matrix. The model scattering intensity is calculated as

$$I(q) = I_{\text{core-shell}}(q) + I_{\text{bkg.}}, \tag{324}$$

where $I_{\text{core-shell}}(q)$ is the scattering intensity of the core-shell objects (which is an integration over the core-shell amplitude form factor, Eq. (182), of the polydisperse core-shell objects; times the number distribution of the core-shell objects), and $I_{\text{bkg.}}$ is assumed to be a q-independent background. $I_{\text{bkg.}}$ is considered to result purely from incoherent scattering and is essentially proportional to the amount of hydrogen in the sample. The theoretical incoherent scattering values for the different components (silica, light and heavy water, phospholipid) are used (see, e.g., [138, 5]), i.e., the incoherent level changes with the solvent isotopic composition.

In order to create a discrete and noisy data set the model scattering intensity is calculated for $N = 130$ (typical detector resolution in a SANS experiment) equidistantly distributed q-values in $0.01 - 2\,\mathrm{nm}^{-1}$, yielding the set $\{I(q_n)\}_{n=1}^N$. Next, these intensity values $I(q_n)$ are multiplied with a constant c, such that $c \cdot \sum_{n=1}^N I(q_n)$ is in the order of 10^6 — hence simulating a SANS experiment with ca. 10^6 detected neutrons. Subsequently, a simulated experimentally observed data set $\{I^c(q_n)\}_{n=1}^N$ is generated by creating realizations of two independent Poisson processes $I^c(q_n) \sim [\mathrm{Poiss}(c \cdot I_{\text{core-shell}}(q_n)) + \mathrm{Poiss}(c \cdot I_{\text{bkg.}})]$. The finally used intensity data set is $\{I_n\}_{n=1}^N$, which is obtained by the rescaling $I_n = I^c(q_n)/c$, and appropriate error bars[65] are calculated as $\sigma_n = \sqrt{I^c(q_n)}/c$. Hence, the complete data set is $\{(q_n, I_n, \sigma_n)\}_{n=1}^N$.

Performing intra-model comparison

The case of performing intra-model comparison is considered. Since the mean diameter of the scattering objects is $2 \cdot (8.6 + 4.4)\,\mathrm{nm} = 26\,\mathrm{nm}$, 40, respectively 100, b-spline basis functions of order three, i.e., $\phi_m(r) = B_3((r - r_m)/h)$, are placed in r-space equidistantly in the interval [0,40] nm, where h is the spacing. The definition of the third order b-spline is (see, e.g., [1])

$$B_3(x) := \frac{1}{6} \begin{cases} (2 - |x|)^3 - 4(1 - |x|)^3, & |x| \le 1, \\ (2 - |x|)^3, & 1 \le |x| \le 2, \\ 0, & |x| \ge 2. \end{cases} \tag{325}$$

In order to check if the b-splines' set-ups are capable to capture the features of the goal PDDFs, noiseless data sets are used in the IFT and their appropriate PDDF solutions are determined and compared with the PDDFs obtained from the direct inversion of $I_{\text{core-shell}}(q)$ via the inverse Fourier transform formula (see, e.g., [51] or [21, §13])

$$\hat{p}(r) = \frac{1}{2\pi^2} \int I_{\text{core-shell}}(q) q r \sin(qr)\, \mathrm{d}q. \tag{326}$$

As Fig. 48 shows, the different b-splines' set-ups are perfect.

[65] The error bar calculation is not consistent with the generation of how the intensity data are generated, but it is the method that is used in practice, and therefore also applied here. The proper generation of error bars would be to sum up the square roots of the coherent scattering intensity and of the incoherent one and to divide the result by c.

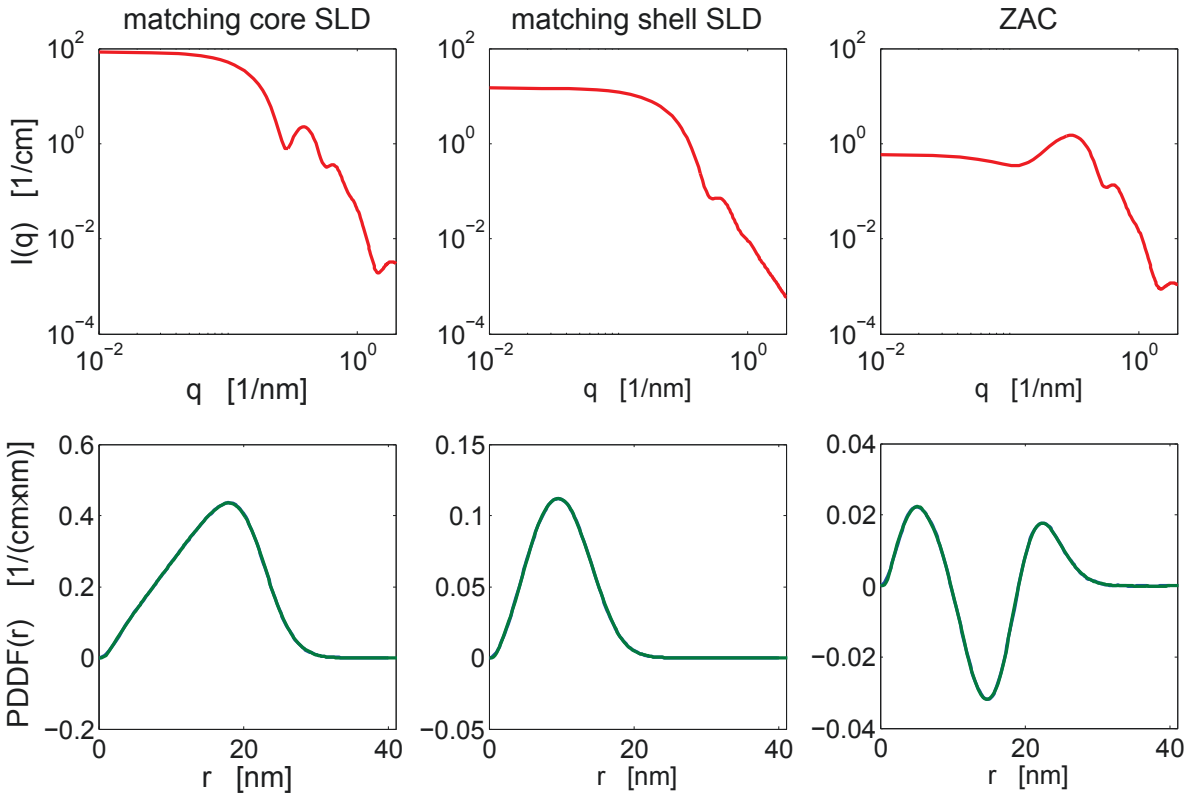


Figure 48: B-spline set-up test. First row: scattering intensities of polydisperse core-shell objects of the three different contrast conditions. Second row: the appropriate PDDFs, where the red curves are the direct PDDF solutions (i.e., the true PDDFs) obtained via Eq. (326), and the blue and green curves are the PDDF solutions obtained by the IFT method having 40, respectively 100, b-splines equidistantly placed in [0,40] nm. However, the true PDDFs are perfectly covered by the two different b-splines' set-ups. Accordingly, the set-ups of the b-spline basis functions (40 and 100 b-splines) are appropriate to capture the features of the true PDDFs.

Fig. 49 and 50 show the results obtained for the set-up with 100 splines and by applying the 5-folded cross-validation method (frequentist approach) and the evidence method (Bayesian approach) to the considered data sets. In Fig. 49 the intensity background is subtracted automatically by the centering method, see section 7.2, and in Fig. 50 the intensity background is removed from the optimization by subtracting it before the optimization method is applied — simulating a user who would have removed it (optimally) 'by hand' before the IFT is applied. Since the data are simulated, the true (noiseless) intensity values are known, and they have been used to determine an error curve, which is called true error curve in the following. A point of this curve is determined[66] by the weighted mean squared error between the true (noiseless) intensity values $\hat{\mathbf{I}}$ and the modeled intensities $\mathbf{\Psi w}$: $\frac{1}{N}\|\mathbf{\Sigma}^{-1/2}(\hat{\mathbf{I}} - \mathbf{\Psi w})\|^2$.

Following aspects are observed in Fig. 49 and 50:

- The (most right) minima positions within the cross-validation curves are in good agreement with the minima positions of the true error curves as well as with the evidence maxima positions.

  Note: (*i*) Changes of the cross-validation curves in the γ-interval of ca. $[10^{-2}, 10^{6}]$ become visible if there is a zoom in. (*ii*) There can be multiple minima in a cross-validation error curve. However, the principle of Occam's razor is applied here in order to get a unique solution: A minimum is selected, which has the highest γ-value, meaning that the appropriate model is the simplest one that is consistent with the data. The other local minima are considered to be spurious.

- If the γ-position of the maximum of the evidence is taken or the γ-position of the minimum of the true error in order to determine the appropriate PDDF, the PDDF would look very similar or identical to the PDDF that is obtained for 5-folded cross-validation.

- It is not possible to derive unique stable points from the stability plots. However, the minima of the cross-validation curves (in case of multiple minima in a curve: the minimum associated to the largest γ), respectively the maxima of the evidence, fall into a γ-interval, where one would suppose a 'stable point'.

- The PDDFs show a slight tendency to oscillate at large r-values, where the true PDDFs are equal to zero. In this regime, the determined uncertainty tubes[67] of the PDDFs are in the order of the amplitudes of the oscillations of the PDDFs.

As these first results show, the cross-validation method seems to be a useful tool to determine objectively an optimal PDDF.

In order to get a statistically more meaningful analysis of the two different inference approaches, 32 i.i.d. intensity data sets are simulated for each set-up (different contrasts, and background is removed a priori or is determined automatically) and the PDDF solutions (having 40 and 100 b-splines) are plotted in a single diagram, see Fig. 51-53. Additionally, in these figures PDDFs obtained for 2-/10-folded and leave-one-out cross-validation are included. Following aspects are observed in these figures — details within the plots are visible within the electronic version of this document:

- In average all PDDFs obtained from the different objective functions (evidence, 2-/5-/10-folded and leave-one-out cross-validation, true error curve) are in good agreement with the true PDDFs.

[66] The true error is essentially the discrete formula of the second term in the last line of Eq. (50) in section 2.2, but without the pre-factor $N/2$ as assumed here for convenience. It has as lower bound the value zero.

[67] An uncertainty tube is obtained by substituting w_m by $w_m \pm \Delta w_m$ in Eq. (266).

- The determined PDDF uncertainty regimes shown in Fig. 49, 50 are in good agreement with the spreads of the PDDFs shown in Fig. 51-53.
- The spreads of the PDDFs obtained from the true error curves are in average the smallest one.
- The spreads of the PDDFs obtained from 5- and 10-folded cross-validation are only slightly higher or comparable as the spreads of the PDDFs obtained from the true error curves.
- Especially for all considered shell matched cases, but also for a few other cases, the spreads of the PDDFs obtained from the evidences are slightly higher as they are for 5- and 10-folded cross-validations and for the true error cases.
- There is a significant outlier PDDF for 2-folded cross-validation of the ZAC case, and the spreads of the PDDFs are also slightly higher as they are for 5- and 10-folded cross-validations and for the true error cases.
- Leave-one-out cross-validation gives spreads of the PDDFs that are comparable with 5- and 10-folded cross-validation — there is only a slightly stronger oscillation of a PDDF for the ZAC case.
- Solutions for 40 and 100 b-splines are nearly identical.
- Additionally, the average misfit error[68] $\mathcal{E}$ is plotted into the diagrams[69]. The observed behavior of the PDDFs is in agreement with the determined misfit errors.

In summary: 5- and 10-folded cross-validations are good methods to determine reliably the PDDF in the considered intra-model comparison approach. Moreover, it has been found that the cross-validation method performs as well as the evidence method.

[68] $\mathcal{E} = \frac{1}{N_{\text{curves}}} \sum_{n=1}^{N_{\text{curves}}} \sqrt{\frac{1}{r_{\max}} \int_{r=0}^{r_{\max}} \left(p_n(r) - p_{\text{true}}(r)\right)^2 \, \mathrm{d}r}$

[69] $\mathcal{E}$ is shown without units within the following figures to keep the diagrams uncluttered.

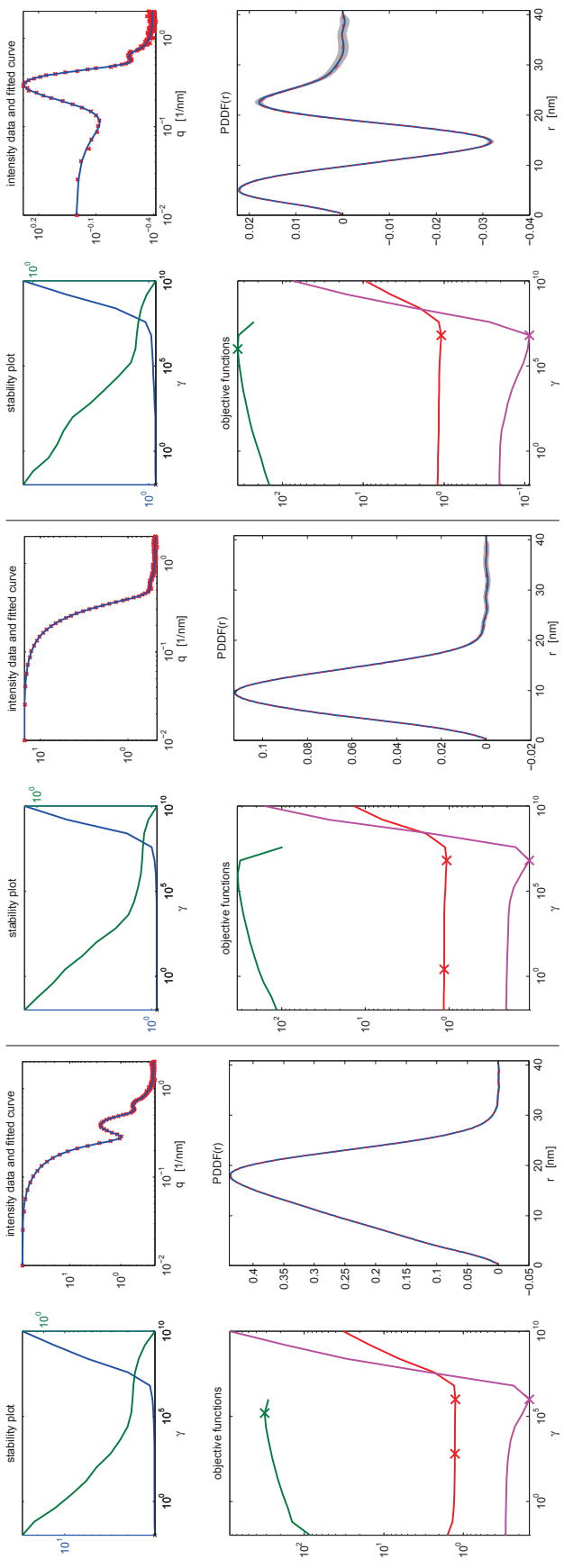


Figure 49: Intra-model comparison for the three considered solvent SLDs: approximately matching core SLD (left block), matching shell SLD (middle block), ZAC condition (right block). The intensity background (that exists in the simulated data) is determined automatically as described in section 7.2. 100 b-splines are used. In each block four diagrams are presented: (*i*) left top: a stability diagram that shows the empirical error (blue curve) and the norm of the (partial) regularization term $\|\mathbf{Lw}\|^2$ (green curve) vs. the regularization parameter γ; (*ii*) left bottom: an objective functions' diagram that shows the evidence (green curve), the (5-folded) cross-validation error curve (red curve), and the true error curve (magenta) — crosses in the curves show minima, respectively maxima; (*iii*) right top: an intensity plot (intensities are in units of cm^{-1}) showing the simulated data set (red crosses) and the model intensity for an optimal regularization parameter γ^*, which is taken as the position of the most right minimum of the cross-validation curve; (*iv*) right bottom: the PDDF determined for γ^* (blue curve, and the gray area shows the uncertainty regime), and the true PDDF (red, dashed curve) (PDDFs are in units of $\mathrm{cm}^{-1}\cdot\mathrm{nm}^{-1}$). The found background values for γ^* are: core matched: $I_{\mathrm{bkg.}} = 0.27\,\mathrm{cm}^{-1}$ (true value $0.23\,\mathrm{cm}^{-1}$), shell matched: $I_{\mathrm{bkg.}} = 0.5\,\mathrm{cm}^{-1}$ (true value $0.49\,\mathrm{cm}^{-1}$), ZAC: $I_{\mathrm{bkg.}} = 0.42\,\mathrm{cm}^{-1}$ (true value $0.41\,\mathrm{cm}^{-1}$).

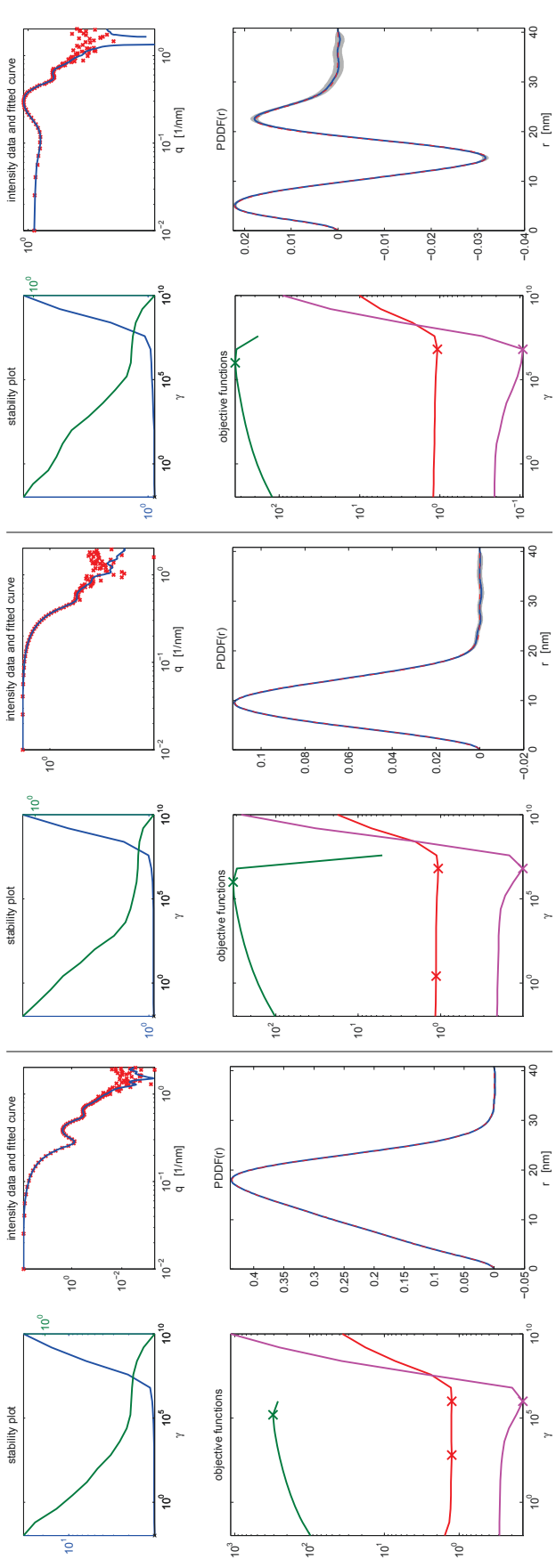


Figure 50: Intra-model comparison for the three considered solvent SLDs: matching core SLD (left block), matching shell SLD (middle block), ZAC condition (right block). The same simulated data sets are used as in the evaluation depicted in Fig. 49, but here the intensity background (that exists in the simulated data) is removed from the optimization by subtracting it a priori — simulating a user who has removed it optimally 'by hand'. 100 b-splines are used. In each block are the following four diagrams: (*i*) left top: a stability diagram that shows the empirical error (blue curve) and the norm of the regularization term (green curve) vs. the regularization parameter γ; (*ii*) left bottom: an objective functions' diagram that shows the evidence (green curve), the (5-folded) cross-validation error curve (red curve), and the true error curve (magenta) — crosses in the curves show minima, respectively maxima; (*iii*) right top: an intensity plot (intensities are in units of cm^{-1}) showing the simulated data set (red crosses) and the model intensity for an optimal regularization parameter γ^*, which is taken as the position of the most right minimum of the cross-validation curve; (*iv*) right bottom: the PDDF determined for γ^* (blue curve, and the gray area shows the uncertainty regime), and the true PDDF (red, dashed curve) (PDDFs are in units of $\mathrm{cm}^{-1}\cdot\mathrm{nm}^{-1}$).

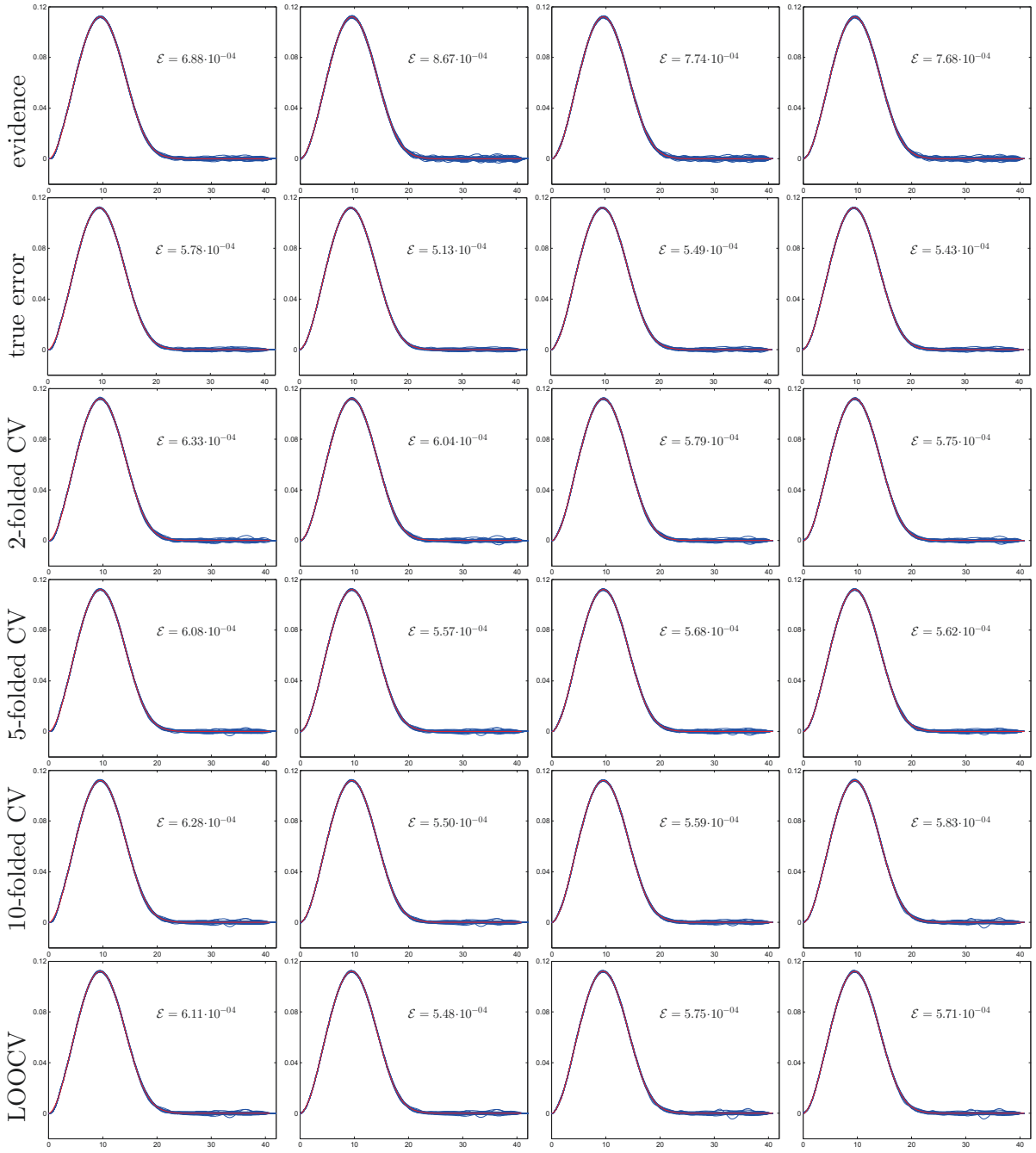


Figure 51: Obtained PDDF solutions ($p(r)/(\mathrm{cm}^{-1} \cdot \mathrm{nm}^{-1})$ vs. r/nm) from the different objective functions and for different numbers of b-splines from intensity data of the 'matching shell SLD' case. For each method 32 i.i.d. scattering data sets are simulated and accordingly the 32 appropriate determined PDDFs are plotted in one diagram (blue curves). The red curves show the true PDDFs. 1st column: 40 b-splines and background is automatically subtracted. 2nd column: 40 b-splines and background is subtracted a priori. 3rd column: 100 b-splines and background is automatically subtracted. 4th column: 100 b-splines and background is subtracted a priori. 1st row: PDDFs obtained via evidence method. 2nd row: PDDFs obtained via true error method. Rows 3-6: PDDFs obtained via Cross-Validation (CV) method (2-folded CV, 5-folded CV, 10-folded CV, Leave-One-Out CV (LOOCV)). $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.

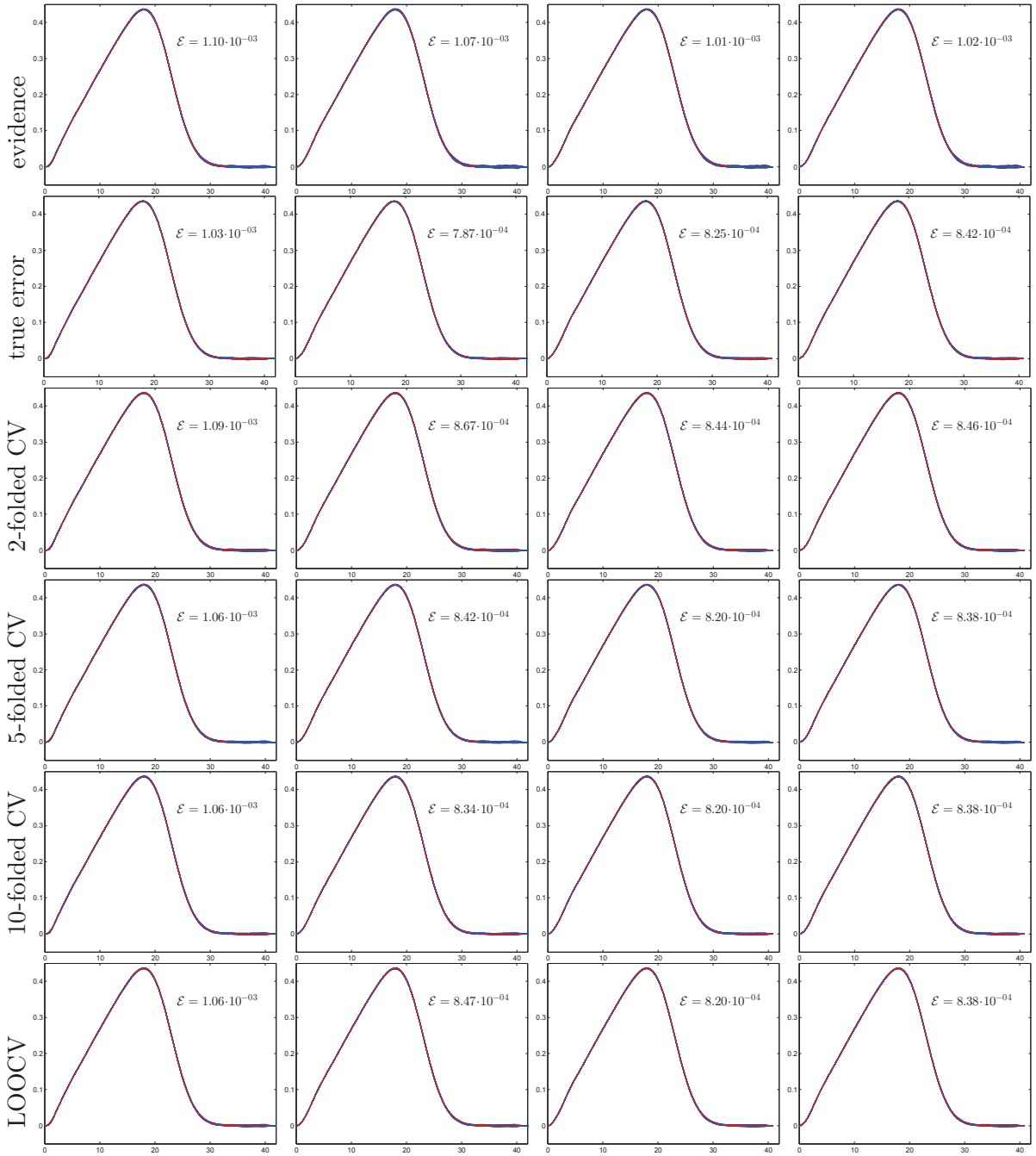


Figure 52: Obtained PDDF solutions ($p(r)/(\mathrm{cm}^{-1} \cdot \mathrm{nm}^{-1})$ vs. r/nm) from the different objective functions and for different numbers of b-splines from intensity data of the 'matching core SLD' case. For each method 32 i.i.d. scattering data sets are simulated and accordingly the 32 appropriate determined PDDFs are plotted in one diagram (blue curves). The red curves show the true PDDFs. 1st column: 40 b-splines and background is automatically subtracted. 2nd column: 40 b-splines and background is subtracted a priori. 3rd column: 100 b-splines and background is automatically subtracted. 4th column: 100 b-splines and background is subtracted a priori. 1st row: PDDFs obtained via evidence method. 2nd row: PDDFs obtained via true error method. Rows 3-6: PDDFs obtained via Cross-Validation (CV) method (2-folded CV, 5-folded CV, 10-folded CV, Leave-One-Out CV (LOOCV)). $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.

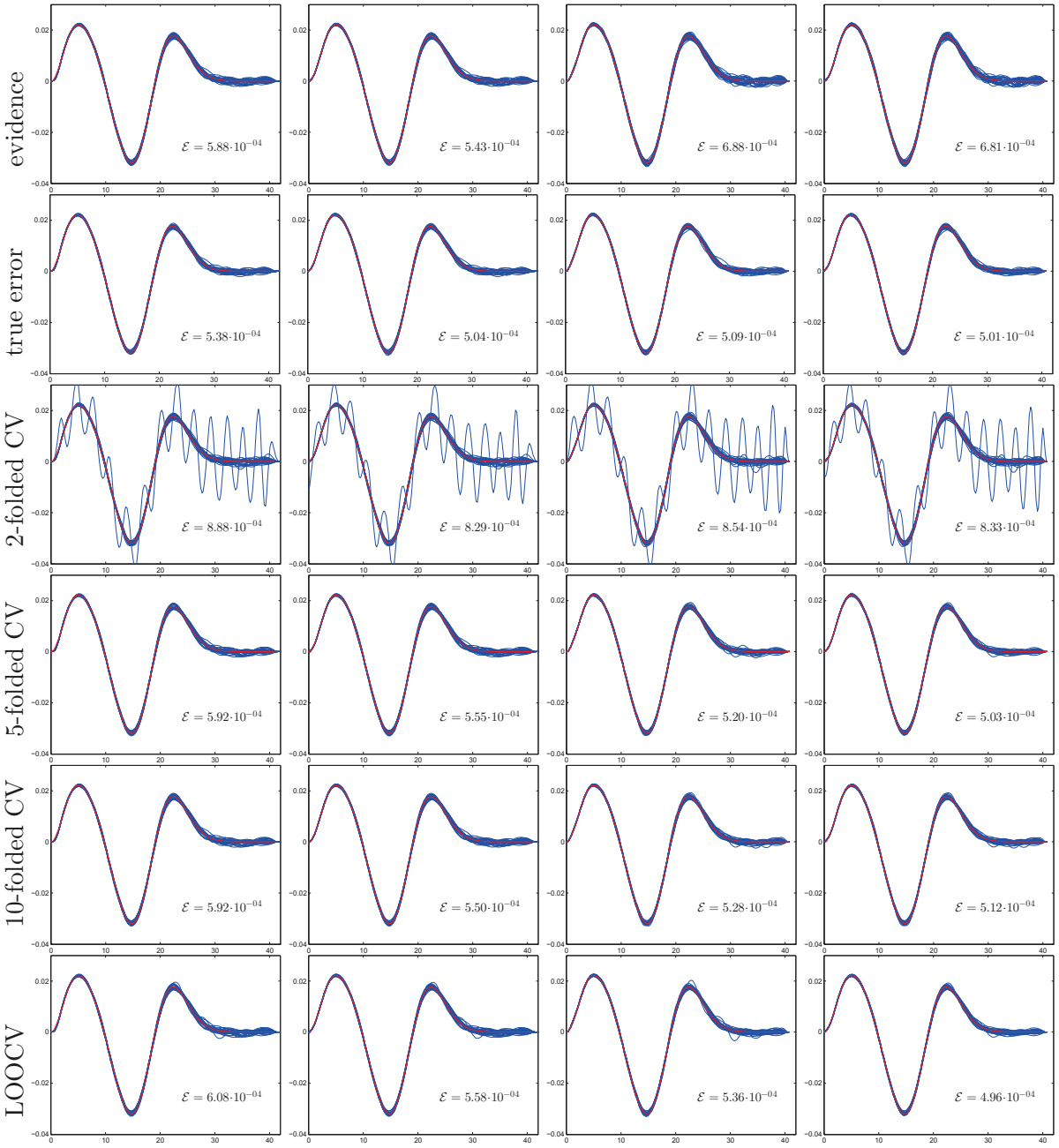


Figure 53: Obtained PDDF solutions ($p(r)/(\mathrm{cm}^{-1} \cdot \mathrm{nm}^{-1})$ vs. r/nm) from the different objective functions and for different numbers of b-splines from intensity data of the ZAC condition case. For each method 32 i.i.d. scattering data sets are simulated and accordingly the 32 appropriate determined PDDFs are plotted in one diagram (blue curves). The red curves show the true PDDFs. 1st column: 40 b-splines and background is automatically subtracted. 2nd column: 40 b-splines and background is subtracted a priori. 3rd column: 100 b-splines and background is automatically subtracted. 4th column: 100 b-splines and background is subtracted a priori. 1st row: PDDFs obtained via evidence method. 2nd row: PDDFs obtained via true error method. Rows 3-6: PDDFs obtained via Cross-Validation (CV) method (2-folded CV, 5-folded CV, 10-folded CV, Leave-One-Out CV (LOOCV)). $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.

Performing inter-model comparison

In order to study the possibility of using the objective functions for performing inter-model comparison the following procedure is considered: In the IFT method the upper integration bound in Eq. (265), respectively in Eq. (268), is set to $r_{\max}$, and its value is changed from 10 nm to 50 nm in steps of 1 nm and the IFT is performed for each $r_{\max}$. For each $r_{\max}$ set-up b-splines of order three are placed equidistantly in steps of 1 nm in $[0, r_{\max}]$. Since the goal is to determine the true PDDF, it is desirable that $r_{\max}$ is close to the actual largest particle size[70] $\hat{r}_{\max}$. As a consequence of this, it is expected that the oscillatory behavior of a determined PDDF is suppressed (mainly in the regime lying outside of the physically relevant domain of the true PDDF) as long as $r_{\max}$ is not much below $\hat{r}_{\max}$ (otherwise strong oscillatory behavior can be expected).

Exemplarily, Fig. 54 shows the results of the different objective functions (5-folded cross-validation, true error, evidence) plotted versus $r_{\max}$ and γ for a ZAC data set, where the background is subtracted before the IFT procedure is applied. In order to have easier access to the important information contained in the surface plots of the objective functions, the figure contains 'projection' plots of the objective functions' surfaces that have been achieved as follows: the minimum-function is applied to the Cross-Validation (CV) and true error surfaces with respect to γ, respectively the maximum-function is applied to the evidence surface with respect to γ (i.e., $\min_\gamma\{\mathrm{CV}(r_{\max}, \gamma)\}$, $\min_\gamma\{(\text{true error})(r_{\max}, \gamma)\}$, $\max_\gamma\{\text{evidence}(r_{\max}, \gamma)\}$). These projection plots show also (local) extrema as crosses within the curves. Finally, for the global extrema appropriate PDDFs can be determined. However, in order to get a statistically more meaningful analysis 32 i.i.d. intensity data sets are simulated and the PDDF solutions are plotted in a single diagram[71], see Fig. 55 — moreover appropriate PDDFs obtained from 10-folded cross-validation are included. The same procedure has been performed for the same data, but the intensity background has been subtracted before the IFT method has been applied. Additionally, the same procedure has been performed for the shell-matched case and the core-matched case and the obtained PDDFs are shown in Fig. 56,57.

Following aspects are observed in Fig. 55-57 (PDDFs are obtained from the global extrema of the objective functions; the figures also contain means and standard deviations of the determined $r_{\max}$ values):

- For all objective functions the PDDFs obtained for the background removal method are similar to the PDDFs obtained for the case where the background is removed optimally 'by hand' before the IFT is applied.

- The PDDFs obtained from the global minima of the true error objective functions are all close to the true PDDFs. These PDDFs have the smallest spreads — compared to the PDDFs obtained from the extrema of the other objective functions.

- The PDDFs obtained from the global maxima of the evidence objective functions are close to the true PDDFs too, but have slightly higher spreads as the PDDFs obtained from the global minima of the true error objective functions. Moreover the PDDFs show a tendency to underestimate the maximal dimension $\hat{r}_{\max}$: the mean values of $r_{\max}$ of the PDDFs obtained from the global minima of the true error objective functions (which are considered as good estimates of the true $\hat{r}_{\max}$ values) are slightly higher than the

[70] Because of the assumed size distributions, particles can be arbitrary large, but the probability of such sizes is negligibly small. The actual largest size can be visually inferred from an inspection of the true PDDFs.

[71] In appendix F.1 the same procedure is applied on the ZAC data set again, but $\mathbf{L}$ is taken as a discrete first order derivative matrix (the regularization matrix that is used in the work of Glatter [49, 26]). As this analysis shows the $\mathbf{L}_{D^2}$ operator used in this work gives better solutions.

mean values of $r_{\max}$ of the PDDFs obtained from the global maxima of the evidences, see tables within Fig. 55-57.

- The PDDFs obtained from the global minima of the 5- and 10-folded cross-validation objective functions are in average also close to the true PDDFs, but some outlier PDDFs exist — $\hat{r}_{\max}$ is overestimated —, which show oscillations for $\hat{r}_{\max} \lesssim r < r_{\max}$. Therefore, the standard deviations of $r_{\max}$ of the PDDFs obtained from these two objective functions are much higher than the ones obtained from the true error objective functions. Moreover, the 5-folded cross-validation method has one strong outlier in Fig. 56, which the 10-folded cross-validation method does not have, but this might be a hazard.

- The misfit errors are in agreement with the observations.

In summary: The true error objective functions yield PDDFs that are closest to the true PDDFs. The evidence objective functions yield PDDFs that are close to the true PDDFs but show higher spreads than the PDDFs obtained from the true error objective functions. The 5- and 10-folded cross-validation objective functions yield in average good PDDF results, but have a tendency to produce outliers, which oscillate more or less strongly. The origin of these outliers is considered to be a result of spurious extrema in the objective functions, cf. Fig. 54 as well as Fig. 49,50.

In order to get rid of spurious extrema the following 2-step procedure is applied, which is a repeated application of the principle of Occam's razor:

(i) The projection plot is modified: Instead of taking the global minimum/maximum with respect to γ of an objective function $f(r_{\max}, \gamma)$, where $r_{\max}$ is fixed, the local minimum/maximum is taken that has the largest γ (in agreement with the principle of Occam's razor) and whose minimum/maximum is non-spurious.

In detail — for the true error and cross-validation objective functions: Let $U(r_{\max})$ the set of all local minima positions of the objective function $f(r_{\max}, \gamma)$ (true error or cross-validation), where $r_{\max}$ is fixed. Then, $\tilde{U}(r_{\max}) = \{\gamma' \in U(r_{\max}) |\, f(r_{\max}, \gamma') < \beta \min_\gamma f(r_{\max}, \gamma)\}$ is a 'cleaned up version' of $U(r_{\max})$: it does not have the possibly contained spurious minima positions of $U(r_{\max})$, which would give too simple models that have relatively high objective function values; here $\beta = 2$ has been found to be appropriate. Finally, the new minimum projection function reads

$$\text{new-min}_\gamma\big(f(\gamma, r_{\max})\big) = \max\big(\tilde{U}(r_{\max})\big). \tag{327}$$

Analogously, a new maximum projection function can be defined for the evidence function.

(ii) Determine the local extrema in the new projection curves with respect to $r_{\max}$, and then choose an local extremum, which has the smallest $r_{\max}$ (hence is in agreement with the principle of Occam's razor, since this selection then has the lowest number of splines, and therefore it is the easiest model that can explain the data) and which is non-spurious (e.g., in the min-projection(CV) diagram of Fig. 54 the first left local minimum is not acceptable — it is much higher than the other local minima).

In detail — for the true error and cross-validation objective functions: Let V the set of all local minima positions of $\tilde{f}(r_{\max}) := \text{new-min}_\gamma\big(f(\gamma, r_{\max})\big)$. Then $\tilde{V} = \{r'_{\max} \in V |\, \tilde{f}(r'_{\max}) < \beta' \min_{r_{\max}} \tilde{f}(r_{\max})\}$ is a 'cleaned up version' of V: it does not have the

possibly contained spurious minima positions of V, which would give too simple models that have relatively high objective function values; here $\beta = 2$ has been found to be appropriate.

Then the optimal $r_{\max}$ value is taken as $r_{\max}^{\text{opt.}} = \min_{r_{\max}}(\tilde{V})$. The appropriate optimal γ of $r_{\max}^{\text{opt.}}$ is $\gamma_{r_{\max}^{\text{opt.}}} = \text{new-min}_{\gamma}\big(f(r_{\max}^{\text{opt.}}, \gamma)\big)$.

The same procedure can be applied for the evidence objective function, if the function is multiplied with -1.

The results of this model selection scheme are also depicted in Fig. 55-57. Following aspects are observed:

- The PDDFs obtained for the evidence objective functions are unchanged compared to the PDDFs obtained from the original (global) extrema selection method.

- The PDDFs obtained for the true error objective functions are nearly unchanged compared to the PDDFs obtained from the original (global) extrema selection method — slight differences are only evident from the misfit errors and the mean values and the standard deviations presented in the tables.

- The PDDFs obtained for the 5- and 10-folded cross-validation objective functions are altogether quite good and there is no outlier curve any more. Moreover, in average the PDDFs are only slightly worse than the PDDFs obtained for the true error objections functions, but they are slightly better than the PDDFs obtained for the evidence objective functions. As a result, the standard deviations of the $r_{\max}$ values are smaller than the ones obtained from the original extrema selection method. Their mean values of $r_{\max}$ are closer to the mean values of $r_{\max}$ obtained for the true error objective function than the mean values of $r_{\max}$ obtained for the evidence objective functions.

- The standard deviations of $r_{\max}$ are in average slightly smaller for 10-folded cross-validation than for 5-folded cross-validation. Moreover, the mean values of $r_{\max}$ obtained for 10-folded cross-validation are in average slightly closer to the mean values of $r_{\max}$ obtained for the true error curve than the mean values of $r_{\max}$ obtained for 5-folded cross-validation. Hence 10-folded cross-validation seems to be to perform slightly better in average than 5-folded cross-validation.

- The misfit errors are in agreement with the observations.

In summary: the considered extrema selection method gives very good PDDF results for 5- and 10-folded cross-validation, and which are very close to the PDDFs obtained for the true error objective function. The considered extrema selection method has no influence on the PDDFs obtained for the evidence objective function and it only has a slight one on the PDDFs obtained for the true error objective function. There is no influence on the evidence method, since in the Bayesian approach the principle of Occam's razor is automatically incorporated — see section 2.2 — and therefore it works automatically well. On the contrary, the cross-validation method only gives an estimate of the generalization error, which can produce spurious minima. Therefore, additional information is required to select a 'good' minimum, e.g., as done here, one that is in consensus with the principle of Occam's razor. In doing so, the 5-/10-folded cross-validation method seems to be a very good objective function for performing inter-model comparison.

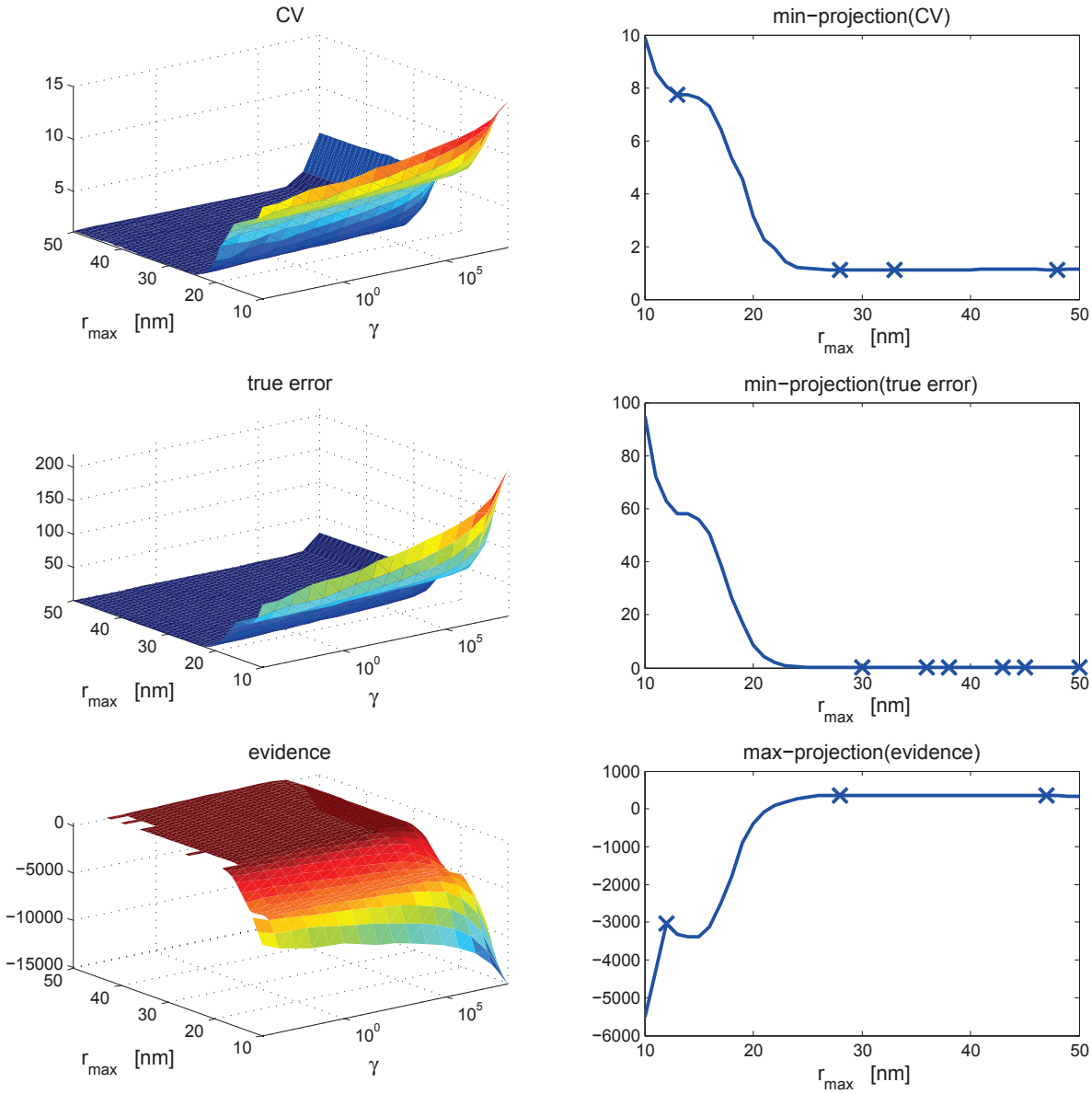
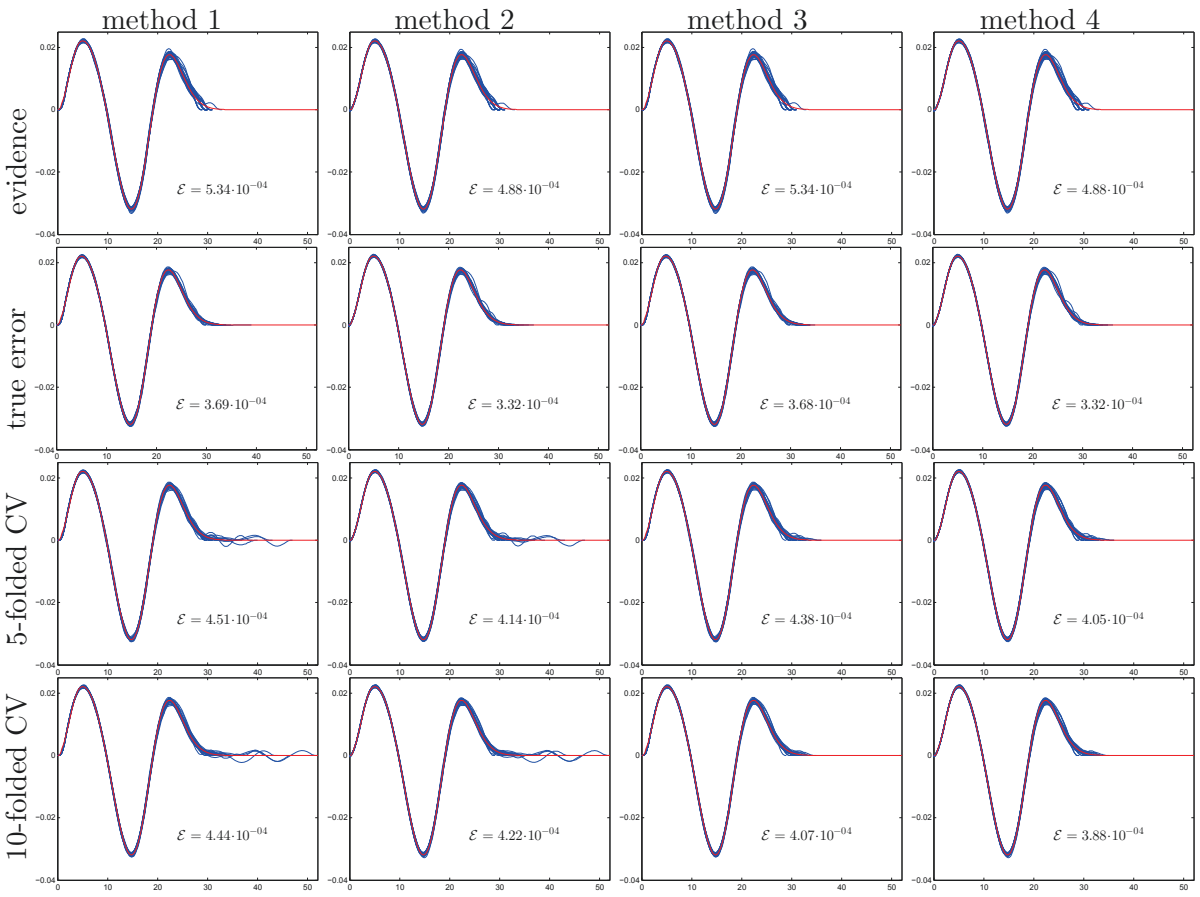
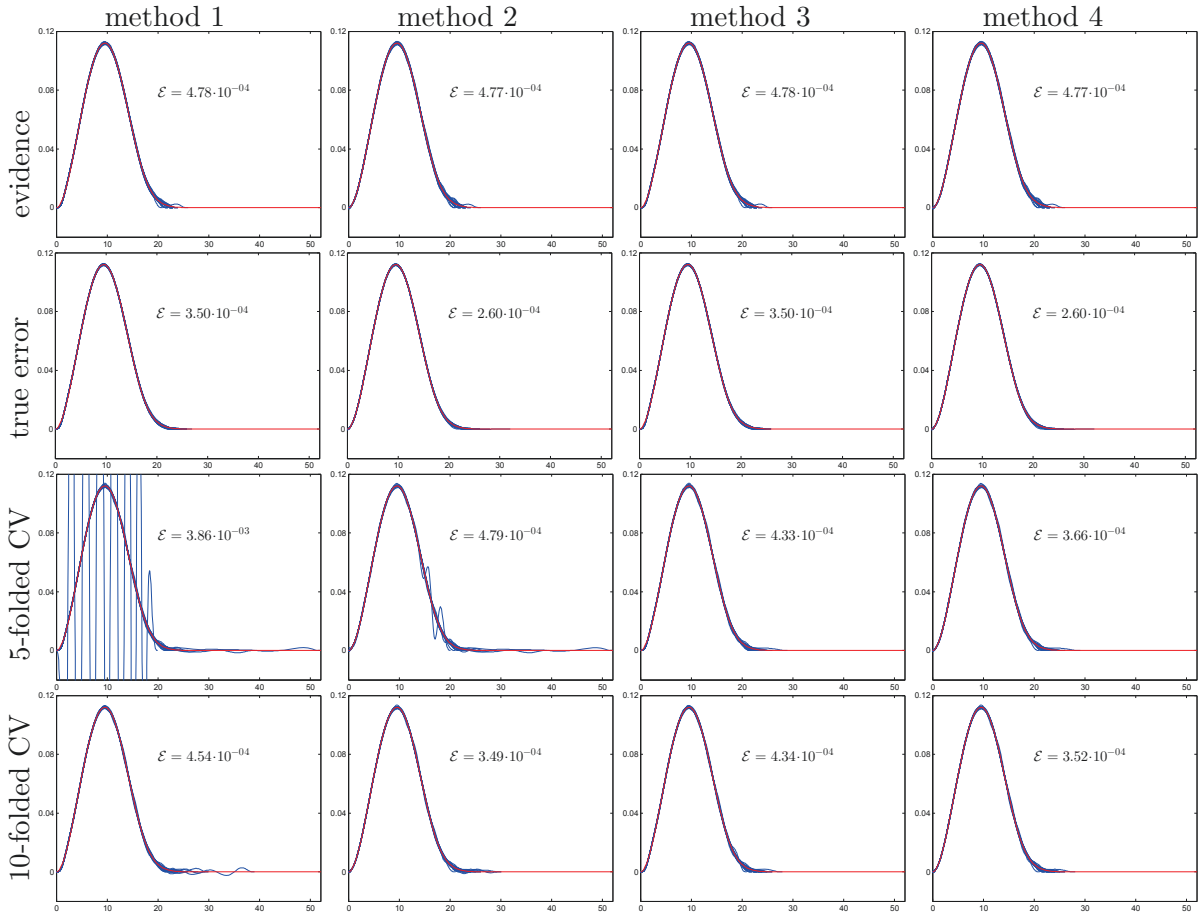


Figure 54: Plots of objective functions for a ZAC data set (background is subtracted before the IFT procedure is applied). In the left column are the surfaces of the objective functions (5-folded Cross-Validation (CV), true error, evidence) shown over $r_{\max}$ and γ. The right column shows appropriate 'projection' plots — from top to bottom: $\min_\gamma \big(\mathrm{CV}(r_{\max},\gamma)\big)$, $\min_\gamma \big((\text{true error})(r_{\max},\gamma)\big)$, $\max_\gamma \big(\text{evidence}(r_{\max},\gamma)\big)$. Crosses indicate extrema. There are some spurious extrema: the most left minimum in the min-projection(CV) curve, and the most left maximum in the max-projection(evidence) curve — the PDDFs corresponding to these extrema oscillate strongly.



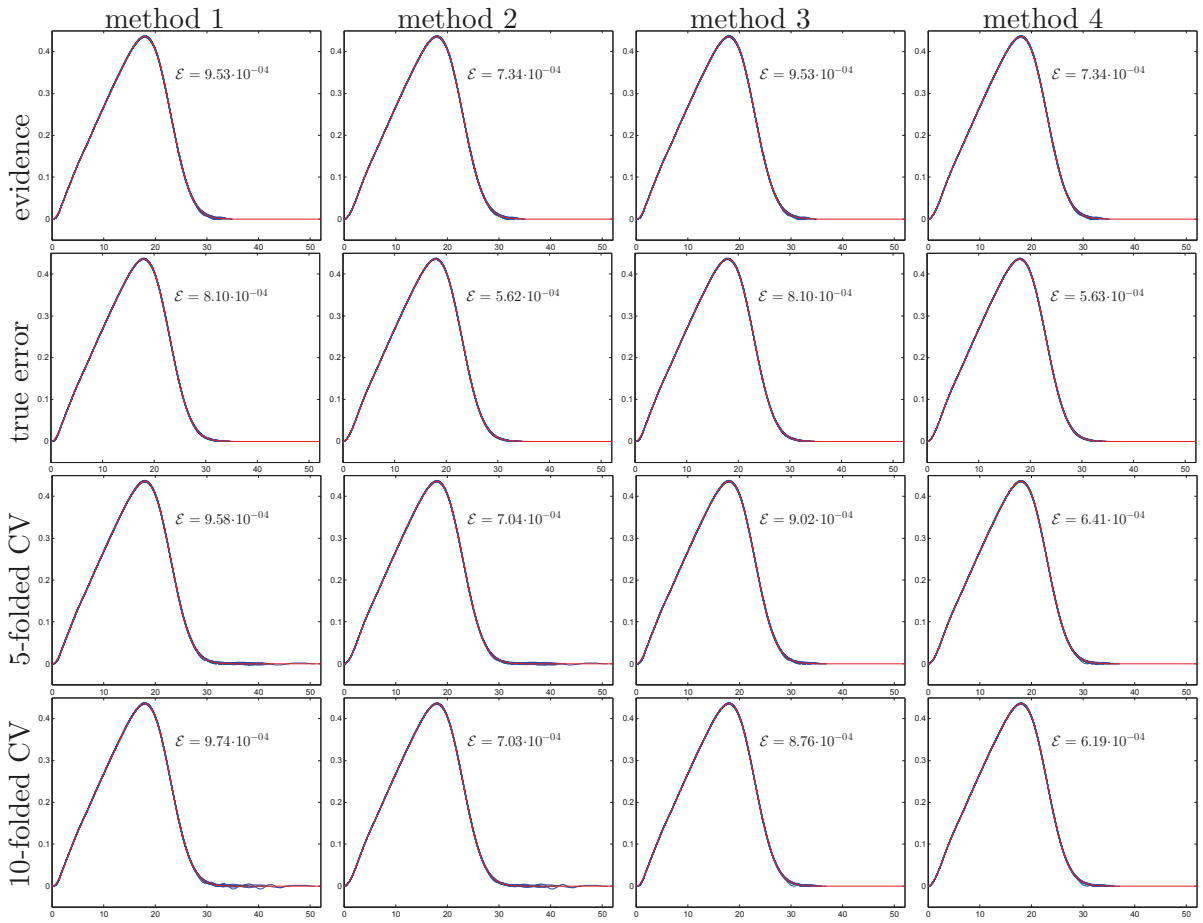
	method 1	method 2	method 3	method 4
evidence	(28.2, 0.79) nm	(28.2, 0.78) nm	(28.2, 0.79) nm	(28.2, 0.78) nm
true error	(30.8, 1.57) nm	(30.8, 1.40) nm	(30.6, 1.10) nm	(30.8, 1.24) nm
5-folded CV	(31.2, 4.07) nm	(31.1, 3.96) nm	(29.6, 1.70) nm	(29.6, 1.70) nm
10-folded CV	(32.0, 5.20) nm	(31.9, 5.23) nm	(29.7, 1.33) nm	(29.6, 1.39) nm

Figure 55: Obtained PDDF solutions ($p(r)/(\mathrm{cm}^{-1}\cdot\mathrm{nm}^{-1})$ vs. r/nm) for the inter-model comparison approach using different objective functions. For each method 32 i.i.d. scattering data sets are simulated for the ZAC condition case, and accordingly the 32 appropriately determined PDDFs are plotted in one diagram (blue curves). The red curves show the true PDDFs. 1st column (method 1): PDDFs are calculated for the global extrema of the objective functions and background is subtracted automatically. 2nd column (method 2): PDDFs are calculated for the global extrema of the objective functions and background is subtracted a priori. 3rd column (method 3): PDDFs are calculated for extrema in accordance with the principle of Occam's razor and background is subtracted automatically. 4rd column (method 4): PDDFs are calculated for extrema in accordance with the principle of Occam's razor and background is subtracted a priori. 1st row: PDDFs obtained via evidence method. 2nd row: PDDFs obtained via true error method. 3rd row: PDDFs obtained via 5-folded Cross-Validation (CV) method. 4th row: PDDFs obtained via 10-folded CV method. The table at the bottom shows the mean and standard deviation of the determined $r_{\max}$ values (notation: (mean, std. dev.)). $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.



	method 1	method 2	method 3	method 4
evidence	(21.0, 0.92) nm	(20.8, 0.98) nm	(21.0, 0.92) nm	(20.8, 0.98) nm
true error	(22.4, 1.04) nm	(22.8, 1.83) nm	(22.3, 0.92) nm	(22.6, 1.83) nm
5-folded CV	(23.1, 5.81) nm	(23.0, 5.88) nm	(21.8, 1.46) nm	(21.6, 1.54) nm
10-folded CV	(22.8, 3.23) nm	(22.3, 2.05) nm	(21.8, 1.46) nm	(21.8, 1.48) nm

Figure 56: Obtained PDDF solutions ($p(r)/(\mathrm{cm}^{-1}\cdot\mathrm{nm}^{-1})$ vs. r/nm) for the inter-model comparison approach using different objective functions. For each method 32 i.i.d. scattering data sets are simulated for the 'matching shell SLD' case, and accordingly the 32 appropriately determined PDDFs are plotted in one diagram (blue curves). The red curves show the true PDDFs. 1st column (method 1): PDDFs are calculated for the global extrema of the objective functions and background is subtracted automatically. 2nd column (method 2): PDDFs are calculated for the global extrema of the objective functions and background is subtracted a priori. 3rd column (method 3): PDDFs are calculated for extrema in accordance with the principle of Occam's razor and background is subtracted automatically. 4rd column (method 4): PDDFs are calculated for extrema in accordance with the principle of Occam's razor and background is subtracted a priori. 1st row: PDDFs obtained via evidence method. 2nd row: PDDFs obtained via true error method. 3rd row: PDDFs obtained via 5-folded Cross-Validation (CV) method. 4th row: PDDFs obtained via 10-folded CV method. The table at the bottom shows the mean and standard deviation of the determined $r_{\max}$ values (notation: (mean, std. dev.)). $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.



	method 1	method 2	method 3	method 4
evidence	(30.9, 0.84) nm	(30.9, 0.84) nm	(30.9, 0.84) nm	(30.9, 0.84) nm
true error	(32.4, 0.79) nm	(32.2, 0.76) nm	(32.3, 0.67) nm	(32.2, 0.67) nm
5-folded CV	(33.9, 4.15) nm	(34.0, 4.12) nm	(31.7, 1.31) nm	(31.7, 1.41) nm
10-folded CV	(33.6, 4.12) nm	(33.7, 4.12) nm	(31.7, 1.15) nm	(31.8, 1.20) nm

Figure 57: Obtained PDDF solutions ($p(r)/(\mathrm{cm}^{-1}\cdot\mathrm{nm}^{-1})$ vs. r/nm) for the inter-model comparison approach using different objective functions. For each method 32 i.i.d. scattering data sets are simulated for the 'matching core SLD' case, and accordingly the 32 appropriately determined PDDFs are plotted in one diagram (blue curves). The red curves show the true PDDFs. 1st column (method 1): PDDFs are calculated for the global extrema of the objective functions and background is subtracted automatically. 2nd column (method 2): PDDFs are calculated for the global extrema of the objective functions and background is subtracted a priori. 3rd column (method 3): PDDFs are calculated for extrema in accordance with the principle of Occam's razor and background is subtracted automatically. 4rd column (method 4): PDDFs are calculated for extrema in accordance with the principle of Occam's razor and background is subtracted a priori. 1st row: PDDFs obtained via evidence method. 2nd row: PDDFs obtained via true error method. 3rd row: PDDFs obtained via 5-folded Cross-Validation (CV) method. 4th row: PDDFs obtained via 10-folded CV method. The table at the bottom shows the mean and standard deviation of the determined $r_{\max}$ values (notation: (mean, std. dev.)). $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.

7.2.3.1. Results

It has been found that model selection for the IFT based on 5- or 10-folded cross-validation, as an estimator of the generalization error within a frequentist approach, is a very good alternative to the already existing evidence objective function within a Bayesian approach. For the inter-model comparison case the cross-validation method even shows a slightly better performance. However, if the cross-validation scheme is used, it is essential to additionally apply the principle of Occam's razor in order to prevent from selecting wrong models.

7.3. Stable Solution

As discussed in sections 7 and 7.1 a major problem with the IFT is that the solution may easily contain artifacts (e.g., oscillations) as a result of the bad numerical conditioning, which means that the solution is numerically unstable. In this section, a few typical machine learning methods, namely the Relevance Vector Machine (RVM) [160], Support Vector Regression (SVR) [164], and Least Absolute Shrinkage and Selection Operator (LASSO) [155] are applied to the IFT method with the goal to get a more stable solution.

The RVM is usually seen as a typical Bayesian approach, and SVR as a frequentist one. However, the RVM, SVR and LASSO approaches differ only in their loss / likelihood function, respectively in their regularization term / prior, hence their formulas can be interpreted within a Bayesian framework as well as in a frequentist one. Nevertheless in practice calculating the evidence is only easily feasible for the RVM, while applying cross-validation is inexpensive for SVR and LASSO, whereas cross-validation would be quite demanding for the RVM (the RVM has as many Lagrangian parameters as elements in the parameter vector, see next section).

The RVM typically yields solutions which only contain basis function contributions that are supported by the data and are otherwise suppressed. Hence, often solutions are sparse, meaning that many basis function parameters $w_m = [\mathbf{w}]_m$ are set to zero. Formally sparseness means $|\{w_m | w_m \neq 0\}| \ll M$. A sparse approach can be interesting as a method that may suppress all artifacts, which are not supported by the data. Solutions found by SVR can also be sparse, but are usually less sparse as obtained via the RVM [160]. LASSO is an approach, which typically also gives sparse solutions.

7.3.1. Relevance Vector Machine (RVM)

The RVM is a sparse Bayesian learning[72] method. In this method, the a priori distribution from the Bayesian approach in section 7.2.2 is substituted for a more general one. It works as follows: Consider the a priori distribution given in Eq. (308) with $\mathbf{L} = \mathbf{1}$, where $\mathbf{1}$ is the $M \times M$ identity matrix, i.e.,

$$\mathrm{pdf}(\mathbf{w}|\gamma) = \mathcal{N}\big(\mathbf{w}|\mathbf{0}, (\gamma\mathbf{1})^{-1}\big) \propto \exp\left\{-\frac{1}{2}\mathbf{w}^T(\gamma\mathbf{1})\mathbf{w}\right\}. \tag{328}$$

This shows that each parameter $w_m = [\mathbf{w}]_m$ has the identical standard deviation $\gamma^{-1/2}$. A more general approach would be to assign an individual standard deviation $\gamma_m^{-1/2} = [\boldsymbol{\gamma}]_m^{-1/2}$ to each parameter w_m. This can be accomplished by defining $\boldsymbol{\Gamma} := \mathrm{diag}(\boldsymbol{\gamma})$, and then using the a priori distribution

$$\mathrm{pdf}(\mathbf{w}|\boldsymbol{\Gamma}) = \mathcal{N}\big(\mathbf{w}|\mathbf{0}, \boldsymbol{\Gamma}^{-1}\big) \propto \exp\left\{-\frac{1}{2}\mathbf{w}^T\boldsymbol{\Gamma}\mathbf{w}\right\}. \tag{329}$$

The RVM uses this prior together with the likelihood function in Eq. (307) in order to maximize the evidence (and then to determine the a posteriori distribution of the parameters) and hence to perform inference [160, 161, 162]. However, the solution would miss a certain degree of smoothness, because of the missing regularization matrix $\mathbf{L}$. Accordingly, the original regularization matrix $\mathbf{L}$ is introduced into the a priori distribution again:

$$\mathrm{pdf}(\mathbf{w}|\boldsymbol{\Gamma}) = \mathcal{N}\big(\mathbf{w}|\mathbf{0}, (\mathbf{L}^T\boldsymbol{\Gamma}\mathbf{L})^{-1}\big) \propto \exp\left\{-\frac{1}{2}\mathbf{w}^T\mathbf{L}^T\boldsymbol{\Gamma}\mathbf{L}\mathbf{w}\right\}. \tag{330}$$

[72] The idea of sparse Bayesian learning has been tried to apply on the inverse Laplace transform in dynamic light scattering [119].

In order to apply the evidence maximization algorithm given in [161], which is described in the text below, the covariance matrix of the a priori distribution needs to be diagonal. In order to accomplish this, the following substitution is performed: Let $\mathbf{L}$ be regular, which is the case for the second order discrete differential operator with boundary conditions zero as given in Eq. (292), and define

$$\tilde{\mathbf{w}} := \mathbf{L}\mathbf{w}, \tag{331}$$

hence

$$\mathbf{w} = \mathbf{L}^{-1}\tilde{\mathbf{w}}, \tag{332}$$

which is substituted into Eq. (330), yielding

$$\mathrm{pdf}(\tilde{\mathbf{w}}|\boldsymbol{\Gamma}) = \mathcal{N}\big(\tilde{\mathbf{w}}|\mathbf{0},\boldsymbol{\Gamma}^{-1}\big) \propto \exp\left\{-\frac{1}{2}\tilde{\mathbf{w}}^T\boldsymbol{\Gamma}\tilde{\mathbf{w}}\right\}. \tag{333}$$

Accordingly, with Eq. (332) the likelihood function in Eq. (307) becomes

$$\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\tilde{\mathbf{w}}) = \mathcal{N}(\mathbf{I}|\boldsymbol{\Psi}_\mathbf{L}\tilde{\mathbf{w}},\boldsymbol{\Sigma}) \propto \exp\left\{-\frac{1}{2}\big\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{I}-\boldsymbol{\Psi}_\mathbf{L}\tilde{\mathbf{w}})\big\|^2\right\}, \tag{334}$$

where

$$\boldsymbol{\Psi}_\mathbf{L} := \boldsymbol{\Psi}\mathbf{L}^{-1}. \tag{335}$$

Using the formulas from appendix E.1, the appropriate a posteriori distribution reads

$$\begin{aligned}\mathrm{pdf}(\tilde{\mathbf{w}}|\mathbf{q},\boldsymbol{\sigma},\mathbf{I},\boldsymbol{\Gamma}) &= \frac{\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\tilde{\mathbf{w}})\mathrm{pdf}(\tilde{\mathbf{w}}|\boldsymbol{\Gamma})}{\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\boldsymbol{\Gamma})}\\ &= \mathcal{N}\big(\tilde{\mathbf{w}}|\boldsymbol{\mu},\mathbf{B}_\mathbf{L}\big)\\ &\propto \exp\left\{-\frac{1}{2}\big\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{I}-\boldsymbol{\Psi}_\mathbf{L}\tilde{\mathbf{w}})\big\|^2 - \frac{1}{2}\tilde{\mathbf{w}}^T\boldsymbol{\Gamma}\tilde{\mathbf{w}}\right\},\end{aligned} \tag{336}$$

where the mean of the posterior is

$$\boldsymbol{\mu} = \mathbf{B}_\mathbf{L}\boldsymbol{\Psi}_\mathbf{L}^T\boldsymbol{\Sigma}^{-1}\mathbf{I}, \tag{337}$$

and the posterior covariance matrix reads

$$\mathbf{B}_\mathbf{L} = (\boldsymbol{\Psi}_\mathbf{L}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}_\mathbf{L} + \boldsymbol{\Gamma})^{-1}. \tag{338}$$

Moreover, using the formulas from appendix E.1, the evidence reads

$$\mathrm{pdf}(\mathbf{I}|\mathbf{q},\boldsymbol{\sigma},\boldsymbol{\Gamma}) = \mathcal{N}(\mathbf{I}|\mathbf{0},\mathbf{W}_\mathbf{L}), \tag{339}$$

where the covariance matrix is

$$\mathbf{W}_\mathbf{L} = \boldsymbol{\Sigma} + \boldsymbol{\Psi}_\mathbf{L}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Psi}_\mathbf{L}^T. \tag{340}$$

There are different strategies to locally maximize[73] the evidence with respect to $\boldsymbol{\Gamma}$, respectively $\boldsymbol{\gamma}$, cf. [161]. E.g., by the first order optimality condition (i.e., setting the derivative of the evidence with respect to $\boldsymbol{\gamma}$ equal to zero and solving for $\boldsymbol{\gamma}$ [161]) two different iterative update strategies for $\boldsymbol{\gamma}$ can be derived. One gives the iterative re-estimation rule

$$\gamma_m^{\mathrm{new}} = \frac{1}{\mu_m^2 + [\mathbf{B}_\mathbf{L}]_{mm}}, \tag{341}$$

[73] The evidence can have many local maxima [28, §2].

where $\mu_m = [\boldsymbol{\mu}]_m$, and which is identical to an expectation maximization update step, hence guarantying that the evidence is locally maximized [161], and the other re-estimation rule is

$$\gamma_m^{\text{new}} = \frac{1 - \gamma_m [\mathbf{B_L}]_{mm}}{\mu_m^2}, \tag{342}$$

which has a faster convergence in practice, but there is no guaranty that the evidence is locally maximized. Indeed in this work tests have shown that the second rule produces solutions that are not acceptable, and therefore the first update rule is used here.

In order to get an initially good value of $\boldsymbol{\Gamma}$, it is set to $\boldsymbol{\Gamma}_{\text{init}} = \text{diag}(\gamma^*, \ldots, \gamma^*)$, where γ^* is the value that maximizes the evidence in Eq. (312). Next, the iterative procedure is performed as long as there are significant relative changes in the elements of $\boldsymbol{\Gamma}$. After the convergence of the algorithm, there is the optimal $\boldsymbol{\Gamma}^*$, which is used to calculate the optimal and most probable parameter vector $\mathbf{w}^*$, via Eq. (332), (337) and (338).

In case the intensity data contain an unknown constant background intensity, the intensity vector $\mathbf{I}$ and the basis function matrix $\boldsymbol{\Psi}_\mathbf{L}$ must first be centered as described in section 7.2.2. Subsequently, the RVM can be applied on the centered quantities.

7.3.1.1. Evaluations

The PDDF results that have been obtained for the inter-model comparison approach (see last section) show that the PDDFs obtained for the ZAC case have the highest oscillations at large r values. Therefore the RVM approach is applied to the ZAC case in order to suppress these oscillations. Fig. 58 shows the results obtained for the RVM in comparison with the original (Tikhonov) method (Eq. (288), respectively Eq. (294) and using the evidence, 10-folded cross-validation and true error objective functions) for two different $r_{\max}$ values (35 and 50 nm) and for the two different methods of treating the background intensity (background subtracted automatically and background subtracted before the RVM is applied). Following aspects are observed:

1. The RVM is superior, compared to the original method and using the different objective functions, in damping oscillatory behavior of the PDDFs in the regimes outside of the domains of the true PDDFs (i.e., for $r > \hat{r}_{\max}$).

2. In the regime $r < \hat{r}_{\max}$, the PDDFs obtained via the RVM deviate slightly stronger from the true PDDFs than the PDDFs obtained from the original method. The PDDFs obtained from the RVM show a tendency to be edged, cf. Fig. 59.

3. The average misfit error $\mathcal{E}$ is in agreement with the two previous observations.

7.3.1.2. Results

The RVM is able to suppress undesired artifacts in the regime outside of the physically relevant domain of the true PDDF. However, the method has a strong tendency of producing sparse solutions, and as a result PDDF solutions can be slightly more edged than the ones obtained from the original (Tikhonov) method; nevertheless, principles features of the true PDDFs are retained. The edges may be explained by the reasoning that if there is a parameter that is supported by the data, an adjacent parameter may be damped, and hence yielding a more edged curve.

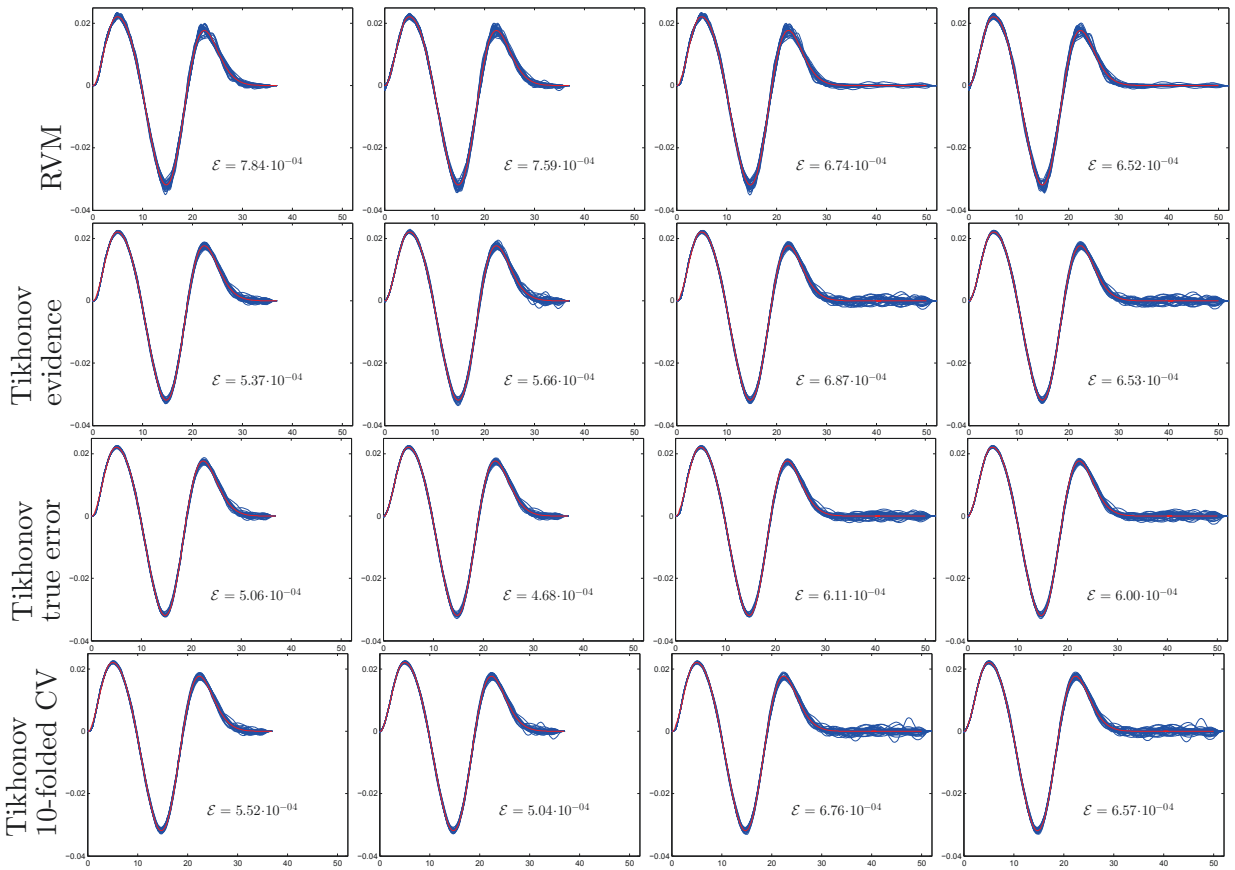


Figure 58: Obtained PDDFs ($p(r)/(\mathrm{cm}^{-1} \cdot \mathrm{nm}^{-1})$ vs. r/nm) (blue curves) for the RVM approach (first row). Moreover, for comparison reasons, PDDFs (blue curves) obtained from the original (Tikhonov) method (i.e., minimizing Eq. (288), respectively Eq. (294) using different objective functions (2nd row: evidence, 3rd row: true error, 4th row: 10-folded cross-validation) are plotted. 1st column: $r_{\max} = 35\,\mathrm{nm}$ and background is determined automatically. 2nd column: $r_{\max} = 35\,\mathrm{nm}$ and background is determined a priori. 3rd column: $r_{\max} = 50\,\mathrm{nm}$ and background is subtracted automatically. 4th column: $r_{\max} = 50\,\mathrm{nm}$ and background is subtracted a priori. Red curves: true PDDFs. $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.

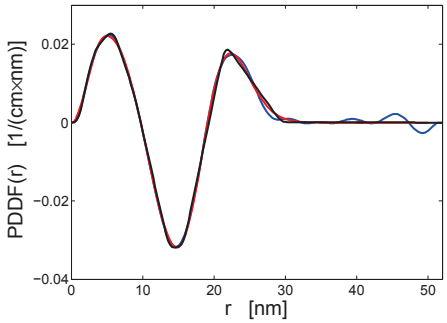


Figure 59: Obtained PDDFs ($p(r)$ vs. r) for the RVM (black curve) and for the original (Tikhonov) approach using the evidence as objective function (blue curve). Red curve: true PDDF. Clearly, the strong oscillatory behavior of the PDDF at large r is suppressed via the RVM, whereas it is present in the PDDF obtained from original approach. However, the PDDF solution obtained from the RVM is more edged, especially in the regime of the last peak.

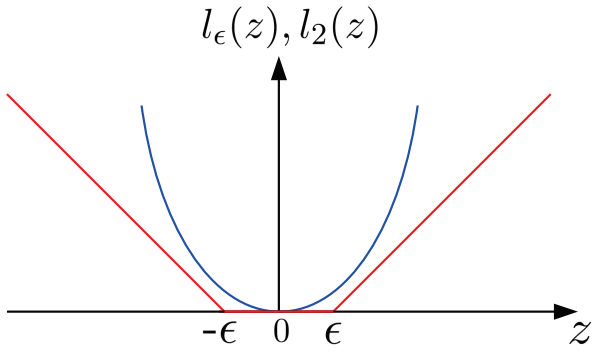


Figure 60: A linear ϵ-insensitive loss function $l_\epsilon(z)$ (red), and an ordinary quadratic loss function $l_2(z)$ (blue).

7.3.2. Support Vector Regression (SVR)

SVR [11, 164] is a method that has been developed in the field of statistical learning theory[74] in order to perform inference. SVR uses an ϵ-insensitive loss function (see below) in a regularized approach for regression, which can yield sparse solutions. However, solutions are usually not as sparse as they are for the RVM [160]. The following description is based on [16, §7.1.4].

In the SVR method a modified version of the regularization problem given in Eq. (288) is minimized: Instead of the squared loss function used in Eq. (288) an ϵ-insensitive loss function [164, §11.1] is used, and here a linear one

$$l_\epsilon\left(\frac{(Ap)(q_n)+I_{\text{bkg.}}-I_n}{\sigma_n}\right)=\begin{cases}0 & \text{if } \left|\frac{(Ap)(q_n)+I_{\text{bkg.}}-I_n}{\sigma_n}\right|<\epsilon\\ \left|\frac{(Ap)(q_n)+I_{\text{bkg.}}-I_n}{\sigma_n}\right|-\epsilon & \text{otherwise,}\end{cases} \tag{343}$$

which is illustrated in Fig. 60. A is the integral operator from Eq. (265), p the basis function model given in Eq. (266), and $I_{\text{bkg.}}$ a possible constant intensity background (see section 7.2). Hence, the weighted misfit $[(Ap)(q_n)+I_{\text{bkg.}}-I_n]/\sigma_n$ yields an error that increases linearly beyond the ϵ-insensitive region. If model fitting is performed with an ϵ-insensitive loss function it is expected that overfitting is suppressed (depending on ϵ), since the ϵ-insensitive loss function prevents the model from following the noise in the data. As a consequence of not following the noise, artifacts should be suppressed in the PDDF solution. In general, solutions are less sensitive to noise in the intensity data and therefore are more robust.

The new optimization problem reads

$$\arg\min_{\mathbf{w},I_{\text{bkg.}}}\left\{\sum_{n=1}^{N} l_\epsilon\left(\frac{(Ap)(q_n)+I_{\text{bkg.}}-I_n}{\sigma_n}\right)+\frac{1}{2}\gamma\|\mathbf{L}\mathbf{w}\|^2\right\}. \tag{344}$$

In a Bayesian approach the linear ϵ-insensitive loss function corresponds to a likelihood function of the from $\frac{1}{2(1+\epsilon)}\exp(-l_\epsilon(z))$ [143], and Eq. (344) can be seen as the MAP solution for the model parameters. However determining γ via maximizing the appropriate evidence is practically not easily tractable. Therefore here, a frequentist viewpoint is taken and the generalization error is estimated by the cross-validation method in order to determine the optimal model complexity, i.e., γ^*. Anyway, Eq. (344) is non-differentiable whenever the loss function has a kink. Therefore, this optimization problem is reformulated via the introduction

[74] Statistical learning theory (see, e.g., [164, 19]) gives a theoretical frequentist framework of performing inference.

of slack variables $\{(\xi_n, \hat{\xi}_n)\}_{n=1}^N$, that are zero if $-\epsilon \leq [(Ap)(q_n) + I_{\text{bkg.}} - I_n]/\sigma_n \leq \epsilon$, otherwise are set to ξ_n and $\hat{\xi}_n$ such that $-\epsilon - \xi_n \leq [(Ap)(q_n) + I_{\text{bkg.}} - I_n]/\sigma_n \leq \epsilon + \hat{\xi}_n$ holds. Then, the corresponding optimization problem follows

$$\min_{\mathbf{w}, I_{\text{bkg.}}, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}}} \left\{ \frac{1}{\gamma} \sum_{n=1}^{N} (\xi_n + \hat{\xi}_n) + \frac{1}{2} \|\mathbf{L}\mathbf{w}\|^2 \right\} \tag{345}$$

$$\text{subject to} \begin{cases} \xi_n, \hat{\xi}_n \geq 0 & \forall n = 1, \ldots, N \\ \frac{I_n}{\sigma_n} \leq \frac{(Ap)(q_n) + I_{\text{bkg.}}}{\sigma_n} + \epsilon + \xi_n & \forall n = 1, \ldots, N \\ \frac{I_n}{\sigma_n} \geq \frac{(Ap)(q_n) + I_{\text{bkg.}}}{\sigma_n} - \epsilon - \hat{\xi}_n & \forall n = 1, \ldots, N, \end{cases} \tag{346}$$

where, because of later convenience, the target function in Eq. (345) is divided by γ^{-1}, letting the minimization problem unchanged. This is a convex optimization problem, hence the solution is unique [115] and can be found by quadratic programming[75] [70]. The problem requires minimization over $M + 2N + 1$ variables and $4N$ side conditions, and it is called the primal problem [115]. A dual problem can be derived from the Lagrangian function of this problem, which is usually solved in practice, since its constraints are somewhat simpler [137, §6], and it only needs minimization over $2N$ variables and $2N + 1$ side conditions. For this purpose, the corresponding Lagrangian optimization problem is considered[76], which reads

$$\begin{aligned} \min_{\mathbf{w}, I_{\text{bkg.}}, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}}} \max_{\mathbf{a}, \hat{\mathbf{a}}, \boldsymbol{\mu}, \hat{\boldsymbol{\mu}}} \Big\{ & \mathcal{L}(\mathbf{w}, I_{\text{bkg.}}, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}}, \boldsymbol{\mu}, \hat{\boldsymbol{\mu}}, \mathbf{a}, \hat{\mathbf{a}}) := \\ & \frac{1}{\gamma} \sum_{n=1}^{N} (\xi_n + \hat{\xi}_n) + \frac{1}{2} \mathbf{w}^T \mathbf{L}^T \mathbf{L} \mathbf{w} - \sum_{n=1}^{N} (\mu_n \xi_n + \hat{\mu}_n \hat{\xi}_n) \\ & - \sum_{n=1}^{N} a_n \left(\epsilon + \xi_n + \left(\frac{\mathbf{w}^T \boldsymbol{\psi}(q_n)}{\sigma_n} + \frac{I_{\text{bkg.}}}{\sigma_n} \right) - \frac{I_n}{\sigma_n} \right) \\ & - \sum_{n=1}^{N} \hat{a}_n \left(\epsilon + \hat{\xi}_n - \left(\frac{\mathbf{w}^T \boldsymbol{\psi}(q_n)}{\sigma_n} + \frac{I_{\text{bkg.}}}{\sigma_n} \right) + \frac{I_n}{\sigma_n} \right) \Big\} \end{aligned} \tag{347}$$

$$\text{subject to} \begin{cases} a_n, \hat{a}_n \geq 0 & \forall n = 1, \ldots, N \\ \mu_n, \hat{\mu}_n \geq 0 & \forall n = 1, \ldots, N, \end{cases} \tag{348}$$

where $\mathcal{L}(\cdot)$ is the Lagrangian function and $\{(a_n, \hat{a}_n, \mu_n, \hat{\mu}_n)\}_{n=1}^N$ are the Lagrangian parameters, also called dual variables. Moreover, $(Ap)(q_n)$ is substituted by $\mathbf{w}^T \boldsymbol{\psi}(q_n)$, where $\boldsymbol{\psi}(q_n) \in \mathbb{R}^M$, and $[\boldsymbol{\psi}(q_n)]_m := \psi_m(q_n)$, where $\psi_m(q_n)$ is defined in Eq. (268). Optimization with respect to the primal variables ($I_{\text{bkg.}}, \{w_m\}_{m=1}^M$, and $\{(\xi_n, \hat{\xi}_n)\}_{n=1}^N$) is accomplished

[75] In this work the MATLAB function `quadprog()` (using an interior point method) is used for solving quadratic programs.

[76] See, e.g., [70] for a good tutorial on convex optimization

by setting the derivative with respect to these variables to zero, yielding

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \mathbf{w} = (\mathbf{L}^T\mathbf{L})^{-1} \sum_{n=1}^{N} \frac{a_n - \hat{a}_n}{\sigma_n} \boldsymbol{\psi}(q_n), \tag{349}$$

$$\frac{\partial \mathcal{L}}{\partial I_{\text{bkg.}}} = 0 \quad \Rightarrow \quad \sum_{n=1}^{N} \frac{a_n - \hat{a}_n}{\sigma_n} = 0 \qquad \text{(if there is a bias)}, \tag{350}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_n} = 0 \quad \Rightarrow \quad a_n + \mu_n = \gamma^{-1} \qquad \forall n = 1, \ldots, N, \tag{351}$$

$$\frac{\partial \mathcal{L}}{\partial \hat{\xi}_n} = 0 \quad \Rightarrow \quad \hat{a}_n + \hat{\mu}_n = \gamma^{-1} \qquad \forall n = 1, \ldots, N, \tag{352}$$

where $\mathbf{L}$ is assumed to be regular. Substituting the obtained expressions back into the Lagrangian function $\mathcal{L}(\cdot)$, except the result from the second line, yields the dual problem

$$\begin{aligned} \max_{\mathbf{a},\hat{\mathbf{a}}} \Bigg\{ \tilde{\mathcal{L}}(\mathbf{a}, \hat{\mathbf{a}}) := &-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (a_i - \hat{a}_i)(a_j - \hat{a}_j) \left(\frac{\boldsymbol{\psi}(q_i)}{\sigma_i} \right)^T (\mathbf{L}^T\mathbf{L})^{-1} \left(\frac{\boldsymbol{\psi}(q_j)}{\sigma_j} \right) \\ &- \epsilon \sum_{n=1}^{N} (a_n + \hat{a}_n) + \sum_{n=1}^{N} \frac{I_n}{\sigma_n} (a_n - \hat{a}_n) \Bigg\} \end{aligned} \tag{353}$$

$$\text{subject to} \begin{cases} 0 \leq a_n, \hat{a}_n \leq \gamma^{-1} & n = 1, \ldots, N \\ \sum_{n=1}^{N} \frac{a_n - \hat{a}_n}{\sigma_n} = 0 & \text{(if there is a bias)}, \end{cases} \tag{354}$$

where the box constraints $0 \leq a_n, \hat{a}_n \leq \gamma^{-1}$ follow from the constraints in Eq. (348) and the conditions in Eq. (351) and Eq. (352). Having determined the optimal dual parameters $\{a_n^*, \hat{a}_n^*\}_{n=1}^N$ via quadratic programming the optimal weights $\{w_m^*\}_{m=1}^M$ can be determined from Eq. (349).

Determination of the background intensity: The optimal bias $I_{\text{bkg.}}^*$ is determined by exploiting the complementary condition of the Karush-Kuhn-Tucker (KKT) conditions [16, Appendix E],[70] at the optimal solution, which states that the product of a primal constraint function at an optimal point with its corresponding optimal Lagrangian multiplier (dual variable) is zero. Therefore,

$$a_n^*(\epsilon + \xi_n^* + (\mathbf{w}^*)^T \boldsymbol{\psi}(q_n)/\sigma_n + I_{\text{bkg.}}^*/\sigma_n - I_n/\sigma_n) = 0, \tag{355}$$

$$\hat{a}_n^*(\epsilon + \hat{\xi}_n^* - (\mathbf{w}^*)^T \boldsymbol{\psi}(q_n)/\sigma_n - I_{\text{bkg.}}^*/\sigma_n + I_n/\sigma_n) = 0, \tag{356}$$

$$\mu_n^* \xi_n^* = 0, \tag{357}$$

$$\hat{\mu}_n^* \hat{\xi}_n^* = 0. \tag{358}$$

Using Eq. (351),(352), the last two equations can be written as

$$(\gamma^{-1} - a_n^*)\xi_n^* = 0, \tag{359}$$

$$(\gamma^{-1} - \hat{a}_n^*)\hat{\xi}_n^* = 0. \tag{360}$$

For $0 < a_n^* < \gamma^{-1}$, i.e., for a_n^* within the box constraints, it follows from Eq. (355) that $\epsilon + (\mathbf{w}^*)^T \boldsymbol{\psi}(q_n)/\sigma_n + I_{\text{bkg.}}^*/\sigma_n - I_n/\sigma_n = 0$, where ξ_n^* is zero, because $a_n^* \neq \gamma^{-1}$ together with the condition in Eq. (359)). Thus

$$I_{\text{bkg.}}^* = I_n - \epsilon\sigma_n - (\mathbf{w}^*)^T \boldsymbol{\psi}(q_n). \tag{361}$$

An analogous computation is possible for any $\hat{a}_n^*$ that is within the box constraints. In order to have a numerically more accurate calculation of $I_{\text{bkg.}}^*$ an average from all a_n^*, $\hat{a}_n^*$ that lie within the box constraints is taken.

Support Vectors: If data point (I_n, q_n, σ_n) is within the ϵ-tube the slack variables ξ_n^* and $\hat{\xi}_n^*$ are zero and $|(\mathbf{w}^*)^T \boldsymbol{\psi}(q_n)/\sigma_n + I_{\text{bkg.}}/\sigma_n - I_n/\sigma_n| < \epsilon$. Due to Eq. (355),(356) it follows that a_n^* and $\hat{a}_n^*$ are zero. Accordingly, only data points lying outside the ϵ-tube can yield $a_n^*, \hat{a}_n^* \neq 0$, and therefore contribute to the optimal solution $\mathbf{w}^*$ (by means of Eq. (349)). Data points that contribute to the solution are called support vectors.

7.3.2.1. Evaluations

Fig. 61 shows the results obtained for SVR in comparison with the original (Tikhonov) method (Eq. (288), respectively Eq. (294) and using 10-folded cross-validation) for two different $r_{\max}$ values (35 and 50 nm), different ϵ values (0.5, 1, and 2), and for the two different methods of treating the background intensity (background subtracted automatically and background subtracted before SVR is applied). Following aspects are observed:

1. Comparing the SVR results obtained via 10-folded cross-validation or the true error with the original Tikhonov method using 10-folded cross-validation, one sees that these solutions are very similar. However, outside the domain of the true PDDF the oscillatory behavior is visually slightly less for SVR with $\epsilon = 2$ (background subtracted automatically) than the one obtained from the original Tikhonov method, but the average misfit error $\mathcal{E}$ is higher. The higher $\mathcal{E}$ can be seen as a result of a slight systematic mismatch within the physically relevant domain of the true PDDF.

2. On average, the oscillatory behavior slightly decreases with increasing ϵ, but at the same time there is a slightly increasing systematic mismatch (also indicated by $\mathcal{E}$).

3. The oscillatory behavior outside the physically relevant domains of the true PDDFs is much stronger in comparison with the results obtained from the RVM, cf. Fig. 58.

7.3.2.2. Results

The oscillatory behavior of a PDDF solution can only be slightly adjusted by ϵ, but adjusting ϵ appropriatly is an instance of the bias-variance tradeoff, cf. footnote 63 on page 129. However, all PDDF solutions found by SVR and via the various ϵ's are very similar to the original Tikhonov method. Hence, at a first glance SVR does not seem to be interesting, but in practice the method can be still valuable, since the loss function is linear for residuals larger than ϵ, which means that outliers do not have as much impact on the empirical error term as it would be the case with a squared loss function. Therefore, the method is more robust against outlier points than the Tikhonov method, which makes SVR interesting in practical applications where outliers usually occur.

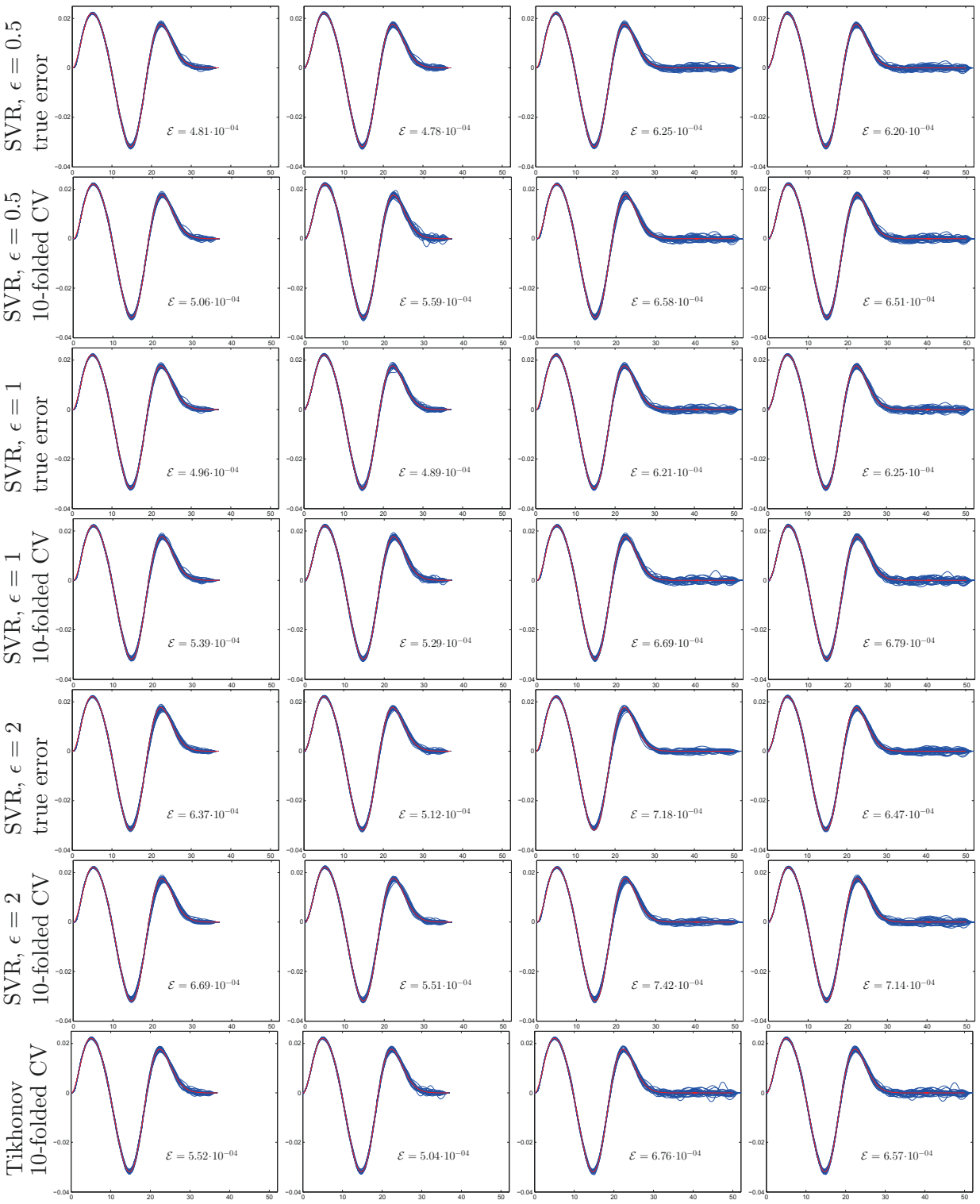


Figure 61: Obtained PDDFs ($p(r)/(\mathrm{cm}^{-1}\mathrm{nm}^{-1})$ vs. r/nm) (blue curves) for the SVR approach (rows 1-6) using different ϵ values and using 10-folded Cross-Validation (CV) or the true error objective function. Moreover, for comparison reasons, PDDFs (blue curves) obtained from the original method (i.e., minimizing Eq. (288), respectively Eq. (294) using 10-folded CV are plotted in row 7. 1st column: $r_{\max} = 35\,\mathrm{nm}$ and background is determined automatically. 2nd column: $r_{\max} = 35\,\mathrm{nm}$ and background is subtracted a priori. 3rd column: $r_{\max} = 50\,\mathrm{nm}$ and background is determined automatically. 4th column: $r_{\max} = 50\,\mathrm{nm}$ and background is subtracted a priori. Red curves: true PDDFs. $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.

7.3.3. Least Absolute Shrinkage and Selection Operator (LASSO)

In the LASSO [155, 156] method linear regression is performed subject to a 1-norm $\|\cdot\|_1$ constraint[77]

$$\arg\min_{\mathbf{w}} \left\{ \frac{1}{2} \left\| \mathbf{\Sigma}^{-1/2}(\mathbf{I} - \mathbf{\Psi}\mathbf{w}) \right\|^2 \right\} \text{ subject to } \|\mathbf{L}\mathbf{w}\|_1 \leq t, \tag{362}$$

where t is an upper bound (tuning parameter) on the norm, and here, in addition to the original method, the regularization matrix $\mathbf{L}$ is included in the 1-norm term, in order to impose additional smoothness on the solution. Assuming $\mathbf{L}$ is regular and using Eq. (332) and (335) the optimization problem[78] is rewritten as

$$\arg\min_{\tilde{\mathbf{w}}} \left\{ \frac{1}{2} \left\| \mathbf{\Sigma}^{-1/2}(\mathbf{I} - \mathbf{\Psi}_\mathbf{L}\tilde{\mathbf{w}}) \right\|^2 \right\} \text{ subject to } \|\tilde{\mathbf{w}}\|_1 \leq t. \tag{363}$$

The appropriate Lagrangian function reads as

$$\arg\min_{\tilde{\mathbf{w}}} \left\{ \frac{1}{2} \left\| \mathbf{\Sigma}^{-1/2}(\mathbf{I} - \mathbf{\Psi}_\mathbf{L}\tilde{\mathbf{w}}) \right\|^2 + \gamma\big(\|\tilde{\mathbf{w}}\|_1 - t \big) \right\}, \tag{364}$$

or equivalently

$$\arg\min_{\tilde{\mathbf{w}}} \left\{ \frac{1}{2} \left\| \mathbf{\Sigma}^{-1/2}(\mathbf{I} - \mathbf{\Psi}_\mathbf{L}\tilde{\mathbf{w}}) \right\|^2 + \gamma \|\tilde{\mathbf{w}}\|_1 \right\}. \tag{365}$$

With increasing γ, respectively with decreasing t, the solution becomes sparse. The reason for this is explained in Fig. 62. The solution of Eq. (365) can be interpreted as the MAP solution in a Bayesian approach, where the 1-norm regularization term can be considered as an a priori distribution that consists of a product of Laplacian distributions[79]

$$\mathrm{pdf}(\tilde{\mathbf{w}}|\gamma) = \prod_{m=1}^{M} \mathrm{Lap}(\tilde{w}_m|0, 1/\gamma) \propto \prod_{m=1}^{M} \exp\big(-\gamma|\tilde{w}_m| \big). \tag{366}$$

Unfortunately, calculating the evidence is analytically not possible and deterministic numerical integrations are not tractable because of the high dimensionality of $\tilde{\mathbf{w}}$. Hence, the problem is considered in a frequentist framework. It is noted that the objective function in Eq. (365) is non-differentiable with respect to $\tilde{\mathbf{w}}$ whenever any $\tilde{w}_m$ is zero in the 1-norm (e.g., there is a kink at $\tilde{w}_1 = 0$ in the LASSO constraint region of Fig. 62), hence is a non-smooth optimization problem [111, §13.3.2] (i.e., simple gradient based optimization routines cannot be applied). On the contrary, the problem in Eq. (363) describes a quadratic optimization problem with linear inequality constraints, which can be solved by a quadratic program [70]. In doing so, $\tilde{\mathbf{w}}$ is decomposed as [155, 177]

$$\tilde{\mathbf{w}} = \tilde{\mathbf{w}}^+ - \tilde{\mathbf{w}}^-, \tag{367}$$

where

$$\tilde{w}_m^+, \tilde{w}_m^- \geq 0 \quad 1 \leq m \leq M, \tag{368}$$

which reduces the number of inequality constraints[80] in the quadratic program from 2^M to $2M$. In order to select t (i.e., the model complexity) 10-folded cross-validation is used as objective function.

[77] The 1-norm of a vector $\mathbf{w}$ reads $\|\mathbf{w}\|_1 = \sum_{m=1}^{M} |w_m|$.

[78] The solution will be unique if $\mathbf{\Psi}_\mathbf{L}$ has rank equal to the number of basis function[157].

[79] $\mathrm{Lap}(x|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b} \right)$

[80] There is another condition that must hold for the decomposition of $\tilde{w}_m$ in to $\tilde{w}_m^+ - \tilde{w}_m^-$, which is $\tilde{w}_m^+ \cdot \tilde{w}_m^- = 0$, but this condition can be shown to be redundant, see, e.g., [177].

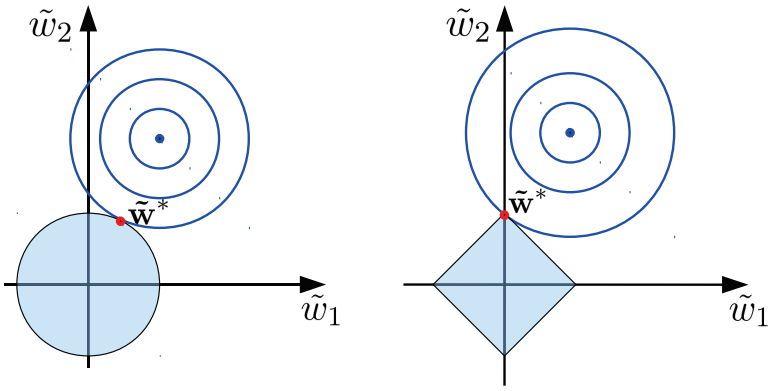


Figure 62: Schematic illustration of the LASSO method based on [155], respectively [65, §3.4.3], in two dimensions. Contours of the empirical error are shown in blue. For the Tikhonov case (left) the constraint region (light blue) is given by $\tilde{w}_1^2 + \tilde{w}_2^2 \leq t^2$, and for the LASSO case (right) by $|w_1| + |w_2| \leq t$. $\tilde{\mathbf{w}}^*$ is the optimum (red dot), which for the LASSO case is solely given by $\tilde{w}_2^*$, while the other parameter is zero; hence the solution is sparse. Note, in general the contours might have any shape.

In case the intensity data contain an unknown constant background intensity, the intensity vector $\mathbf{I}$ and the basis function matrix $\mathbf{\Psi_L}$ must first be centered as described in section 7.2.2. Subsequently, the LASSO method can be applied on the centered quantities.

7.3.3.1. Evaluations

Fig. 63 show the results obtained for the LASSO method in comparison with the original (Tikhonov) method (Eq. (288), respectively Eq. (294)) using 10-folded cross-validation) for two different $r_{\max}$ values (35 and 50 nm), for two different methods of treating the background intensity (background subtracted automatically and background subtracted before the LASSO method is applied), and for different objective functions (true error objective function, 10-folded cross-validation, and manually selecting a kink point — see below).

Following aspects are observed:

1. In average, for $r_{\max} = 50$ nm the PDDFs obtained via the LASSO method (and using the true error or 10-folded cross-validation) show a less strongly oscillatory behavior as the one of the PDDFs obtained from the original method. For $r_{\max} = 35$ nm the observation is similar. However, the average misfit error is in a direct comparison slightly larger for the LASSO method, which can be seen as a result of a slight systematic mismatch.

2. In the performed evaluation it has been noted that the solution is quite sensitive with respect to adjusting t. Accordingly, the determination of the optimal model complexity, i.e., t^*, via 10-folded cross-validation can easily yield PDDFs that slightly oscillate. However, by inspecting the 10-folded cross-validation curve a kink point can be detected in the curve, see Fig. 64, which is close to the minimum, but it corresponds to a slightly simpler model. This kink point has been determined in all 10-folded cross-validation curves of the LASSO method and the appropriate PDDFs were calculated for them and are also shown in Fig. 63. These PDDFs clearly show that the oscillatory behavior outside of the physically relevant domains of the true PDDFs is strongly suppressed. On the other hand, the PDDFs are more edged and slightly deviate systematically from the true PDDFs. Consequently, the average misfit errors are increased, but principles features of the true PDDFs are retained.

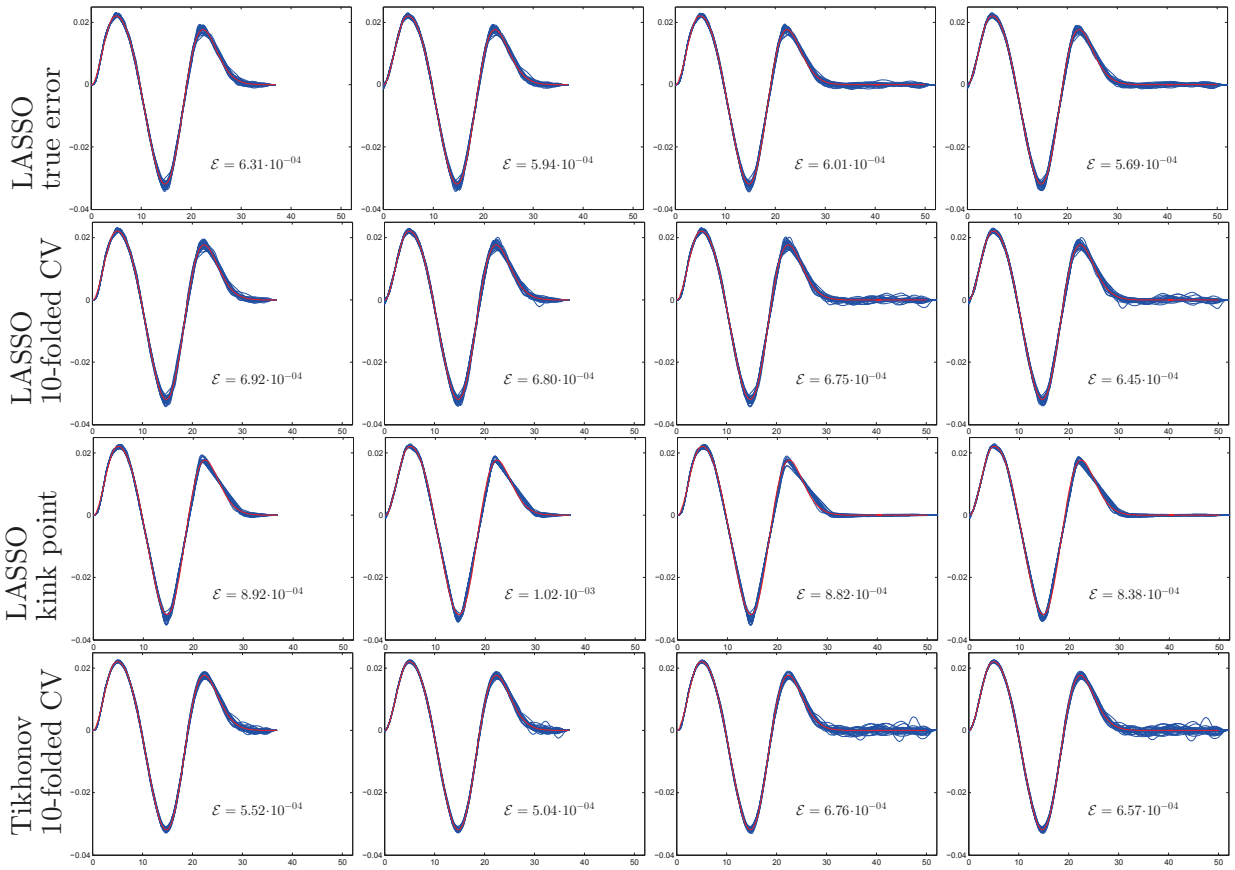


Figure 63: Obtained PDDFs ($p(r)/(\mathrm{cm}^{-1} \cdot \mathrm{nm}^{-1})$ vs. r/nm) (blue curves) for the LASSO method (rows 1-3) using different objective functions. Row 1: true-error objective function is used. Row 2: 10-folded Cross-Validation (CV) is used. Row 3: manual selection of the kink point in the 10-folded CV curve (cf. Fig. 64). Moreover, for comparison reasons, PDDFs (blue curves) obtained from the original method (i.e., minimizing Eq. (288), respectively Eq. (294)) using 10-folded CV are plotted in row 4. 1st column: $r_{\max} = 35\,\mathrm{nm}$ and background is determined automatically. 2nd column: $r_{\max} = 35\,\mathrm{nm}$ and background is subtracted a priori. 3rd column: $r_{\max} = 50\,\mathrm{nm}$ and background is determined automatically. 4th column: $r_{\max} = 50\,\mathrm{nm}$ and background is subtracted a priori. Red curves: true PDDFs. $\mathcal{E}$: average misfit error. Details within the plots are visible in the electronic version of this document.

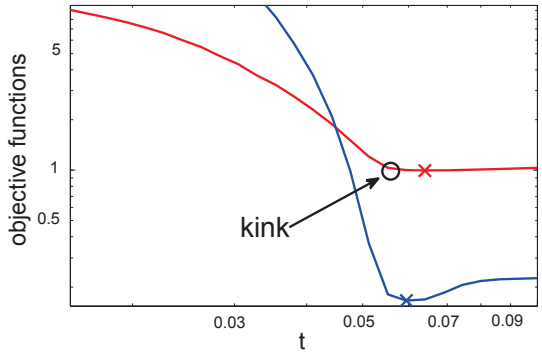


Figure 64: The kink point method for determining an optimal model. Blue curve: true error curve. Red curve: 10-folded cross-validation curve. In order to get a kink point from the 10-folded cross-validation curve, the curve is inspected in the regime left from its most left minimum (there is only one (local) minimum in the displayed curve, but in general there can be more than one minimum) for a relatively strong change of the slope, which then defines the optimal t, i.e., model complexity.

7.3.3.2. Results

The LASSO method gives good results if the true error is used as objective function. In practice 10-folded cross-validation can be used as a substitute, which produces slightly less oscillations than the original (Tikhonov) method. A minor problem with the LASSO method is that it is quite sensitive with respect to adjusting t appropriately, while at the same time the optimal model complexity estimation by means of 10-folded cross-validation is not accurate enough. Consequently, the oscillatory behavior of the PDDFs obtained from 10-folded cross-validation is higher than the one obtained from the true error. However, t can be re-adjusted according to the kink point method described above. Solutions obtained by this method are slightly edged and deviate systematically, but on the other hand oscillatory artifacts are strongly suppressed while principle features of the true PDDFs are retained. In practice the true solution can be considered to lie in between the solution of the kink point method and the one obtained from 10-folded cross-validation.

7.4. Conclusion

The inverse Fourier transform is an ill-posed problem, hence it needs regularization / a priori knowledge in order to get a reasonable solution, which is implemented in the IFT method. However, adjusting the model complexity requires an objective function in order to perform objectively model comparison and to get a unique solution. Previously, only the evidence method of the Bayesian inference approach could accomplish this [64]. In this work now, the model selection problem has been discussed in a complementary frequentist probability approach, and it has been shown that an objective function is available by means of 5- or 10-folded cross-validation — as a substitute for the generalization error as the fundamental objective function in the frequentist approach. The obtained results show that the cross-validation method performs equally well or often even better than the evidence method.

Moreover, the loss function, respectively the regularization term / a priori distribution, have been changed in the original Tikhonov approach in order to get a more stable solution. The considered changes result in the RVM, SVR, and the LASSO method. For the RVM only the evidence can be efficiently calculated, while for SVR and LASSO it is the cross-validation method. SVR with a small ϵ gives similarly good solutions as the ones obtained from the original Tikhonov method. Nevertheless, SVR can be very valuable in practice in order to decrease the influence of outlier points on the solution (the linear loss function does not penalize outliers as much as the squared loss function) — hence in this case the method can be more robust. The RVM and the LASSO method have the potency to suppress undesired oscillations. Unfortunately, they also distort the PDDFs to more edged curves, but the methods retain principles features. Comparing the RVM with the LASSO method, the LASSO method seems to be more attractive, since it allows to easily adjust the degree of sparseness with a single parameter.

In practice it may be valuable to have the RVM, SVR and LASSO as alternatives for the inter-model comparison approach using the original Tikhonov formula, e.g., for the case when the original Tikhonov approach fails and in order to perform cross-checking of a Tikhonov solution.

As an outlook, some possible variations and extensions of the given methods are given here. E.g., the LASSO method can be extended by a quadratic regularization term (hence is a mixture of Tikhonov and LASSO regularization, which is known as elastic net [111, §13.5.3]) in order to get a smoother solution as one would obtain with the LASSO method. However, it requires the determination of another regularization parameter, thus needs more computational power.

The SVR approach can be modified by replacing the linear ϵ-insensitive loss function by a quadratic ϵ-insensitive loss function [164, §6.1], which would be more in agreement with a Gaussian noise model, but it would then lose the property of being a robust estimator. One can also consider a mixture of SVR and the LASSO method (i.e., using an ϵ-insensitive loss function together with a 1-norm regularization term) or a mixture of SVR and an elastic net, with the goal to have a robust estimator that can also give sparse solutions. However, these are all typical frequentist ad-hoc approaches, which are not built on fundamental principles. Nevertheless, the principle idea of performing model selection based on the KLD, respectively the generalization error, remains[81].

Complementary to the frequentist approach the Bayesian one gives a clear inference procedure, which is of a hierarchical structure if different a priori terms and likelihood functions are

[81] Note, the KLD approach fails if the solution is not unique. In this case additional knowledge or principles (e.g., Occam's razor) are required in order to perform model selection.

used, cf. section 2.2. However, practical implementations would usually require deterministic [16, §10] or Monte-Carlo [16, §11] approximate inference schemes, which makes the complete inference process non-trivial, and which also include approximation errors. Conversely, here, the RVM has been presented as a tractable Bayesian approach, and it is noted that the RVM can be easily extended to allow for an integration over an unknown scaling of the covariance matrix [161].

The RVM uses an a priori distribution in which each parameter has its own unknown standard deviation, hence it is the complementary extreme case of the basic Tikhonov approach, in which all parameters share the same standard deviation. Considering this, an intermediate approach could group parameters, where each parameter group has its own associated standard deviation. Considering only two parameter groups, having different associated standard deviations, and such that one group of parameters is associated to $r \leq \hat{r}'_{\max}$ and the other group of parameters is associated to $r > \hat{r}'_{\max}$, and if the desired goal is to push parameters from the last group to zero (by a standard deviation that goes to relatively large values) and to adjust $\hat{r}'_{\max} \approx \hat{r}_{\max}$, one essentially performs inter-model comparison, but the described scheme is more complex than the evidence approach for inter-model comparison given in section 7.2.

With the discussed inference approaches a model-free analysis of SAS data via IFT can be performed objectively, automatically, and online. This is of high importance due to the development of large scale facilities receiving users from a broad community of scientists from academic research or industry, who might have little understanding of the technique and theory, but who need the IFT analysis results. To date, the requirement of rather deep knowledge in scattering theory and expertise in data analysis is keeping back the technique from spreading and be used routinely. Last but not least, the different IFT methods presented in this work allow to compare their results which each other and hence increasing the reliability of outcomes.

8. Overall Conclusion and Outlook

In this thesis different technical and scientific aspects of Small-Angle Scattering (SAS) experiments have been discussed.

SAS has been interpreted and discussed within the Bayesian and frequentist inference frameworks. This discussion has shown the importance of a priori knowledge about the sample system and scattering set-up in order to infer structural information within a physical modeling approach. Moreover, the two inference frameworks have been used to derive objective functions for the Indirect Fourier Transform (IFT), and additionally to interpret different other new regularization approaches for the IFT in order to get a more stable solution.

A newly developed program, called SASET, has been presented that allows to handle and efficiently evaluate comprehensive 1- and 2-dimensional SAS data series, a crucial point that was the bottleneck in the analysis process up to now. An analysis of an IPEC contrast variation data set has been done with SASET, showing the importance of a priori knowledge and of the information content within a scattering experiment.

MC simulations have been performed to model the scattering of some complex cluster structures. The simulations show the power and resolution limit of SAS experiments, and aiming at clarifying the structure of some real data. Furthermore, successful model simplifications were made to get an analytical model for the scattering of vesicles that are decorated with particles.

Simulations of coalescence have been done in order to model the creation of vesicles by a disk-disk-coalescence process, and to determine the scattering intensity. The simulations show which parameters can be inferred from a time series of scattering intensities. Moreover, the coalescence model has been fitted to experimental data, showing that the mechanism of the considered system is indeed a coalescence one.

The evaluation of 2-dimensional scattering images has been considered by means of different methods of quantifying the anisotropy, and two new methods (entropy, PCA) have been introduced. Moreover, it has been shown that the orientational distribution function should be calculated from determined order parameters by additionally using the maximum entropy principle, hence yielding a more robust solution.

In summary, determining the scattering length density profile of a SAS (SANS or SAXS) experiment is an inverse problem that is ill-posed, hence a certain amount of (physical) a priori knowledge and distinctive features in the scattering data are required in order to perform reasonably well inference. The theoretical inference discussion has emphasized the important point of using a priori knowledge for fitting models to SAS data, and consequently there is no 'mathematical trick' to overcome a lack of knowledge in the inference process. However, inferring physical models is supported and enhanced by sophisticated software: e.g., if the software allows to evaluate simultaneously contrast variation data sets/series (i.e., the amount of available and processed information is rather high), if it can sequentially fit models to SAS data of continuously changing systems (i.e., the model changes are constrained to change continuously) or if it can fit simultaneously the overall process with a single model (i.e., a lot of information is processed to determine a few parameters reliably), if the software allows to incorporate a priori knowledge via the maximum entropy method, etc. The procedure of inferring free-form solutions is supported by software that allows to determine objectively solutions, and that gives different regularization methods at hand to get the most robust solution. Last but not least, increasing interest of analyzing complex hierarchical systems requires sophisticated software in order to simulate and fit these systems, and to understand which influence parameters have on the features of scattering intensities.

Since SANS and SAXS are indispensable techniques for determining nano-structured mate-

rials, it is foreseeable that a future interest will be the development of even more sophisticated software in order to support users in their inference procedure. As a desirable target, it is conceivable that a software gets some physical information about a system and the scattering data, and the software automatically suggests possible structures. In any case, it is mandatory that the evolution of more sophisticated SAS software requires the symbiosis of expertises from fields as computer sciences, physics, chemistry, engineering, statistics, machine learning, etc.

Appendices

A. Appendix

A.1. Probability Density Transformation

Let x be a random variable and $\mathrm{pdf}_x(x)$ its probability density function. Now, consider the transformation $x = g(y)$. Since the probability in the infinitesimal small interval $[x, x + \mathrm{d}x]$ is $|\mathrm{pdf}_x(x)\,\mathrm{d}x|$, which must be equal [16, chapter 1.2.1] to the probability of its transformed density pdf_y in $[y, y + \mathrm{d}y]$, i.e.,

$$|\mathrm{pdf}_y(y)\ \mathrm{d}y| = |\mathrm{pdf}_x(x)\ \mathrm{d}x|, \tag{369}$$

it follows

$$\begin{aligned} \mathrm{pdf}_y(y) &= \mathrm{pdf}_x(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right| \\ &= \mathrm{pdf}_x(g(y)) \left| \frac{\mathrm{d}g(y)}{\mathrm{d}y} \right|. \end{aligned} \tag{370}$$

As an example, consider the lognormal distribution, which is an often used distribution in the implemented scattering models in this work. A random variable y is said to be lognormally distributed if $\ln(y) = x$ is normally distributed, i.e.,

$$\mathrm{pdf}_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right), \tag{371}$$

where μ is the mean of the distribution and σ is its standard deviation. Using the transformation $x = \ln(y)$, then yields the transformed density

$$\begin{aligned} \mathrm{pdf}_y(y) &= \mathrm{pdf}_x(\ln(y)) \left| \frac{\mathrm{d}\ln(y)}{\mathrm{d}y} \right| \\ &= \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2} \right), \end{aligned} \tag{372}$$

where $\mathrm{d}(\ln(y))/\mathrm{d}y = 1/y$ has been used. Note, herein μ and σ are not the mean and standard deviation anymore.

A.2. Inversion Sampling

Let x be a random variable with cumulative distribution function F_x and let u a uniformly distributed random variable over the interval $[0, 1]$, i.e., $u \sim U(0, 1)$. Then, due to [129, chapter 7.3.2],

$$x = F_x^{-1}(u), \tag{373}$$

where F_x^{-1} is the inverse function to F_x.

This means that if the cumulative distribution function F_x is calculated from the probability density function of the random variable x, pdf_x, the random variable $x \sim \mathrm{pdf}_x$ can be generated by calculating the inverse cumulative distribution function of F_x, F_x^{-1}, and applying it to the uniformly distributed variable u.

A.3. Uniformly Distributed Random Points from a Ball

In section 4.2 a method is given for sampling random points uniformly from a ball. Here, a second method is explained that was implemented firstly by using the description from [2] that is based on the theorem 1 in [10]. As it has been discovered, the description in [2] is wrong: theorem 1 in [10] involves some random distributions of the form $1/(2\Gamma(1+1/p))e^{-|t|^p}$ $(t \in \mathbb{R})$, where Γ is the gamma function and $p > 0$ defines some p-norm. However, in [2] it is stated that for creating random numbers uniformly from a ball in three dimensions (using the 2-norm), the random distribution is a standard normal distribution. This is not correct, but by applying the probability transformation from appendix A.1 the standard normal distribution can be used:

Due to theorem 1 in [10] uniformly distributed random points in the unit three dimensional ball in the Euclidean space (i.e., taking the 2-norm, $p = 2$) can be created as follows: Let g_1, g_2 and g_3 be i.i.d. random variables with density

$$\mathrm{pdf}_{g_i}(g_i) = \frac{\exp(-g_i^2)}{2\Gamma\left(\frac{3}{2}\right)} \qquad g \in \mathbb{R},\ i = 1, 2, 3, \tag{374}$$

and let y be an exponential random variable with density

$$\mathrm{pdf}_y(y) = \exp(-y) \qquad y \geq 0, \tag{375}$$

and independent of g_1, g_2 and g_3. Then the random variable

$$\frac{(g_1, g_2, g_3)^T}{\sqrt{g_1^2 + g_2^2 + g_3^2 + y}} \qquad \in \mathbb{R}^3 \tag{376}$$

is uniformly distributed over the unit ball in three dimensions.

Now, consider the change of variable

$$g_i = u_i(x_i) := x_i/\sqrt{2} \qquad , i = 1, 2, 3. \tag{377}$$

Then, the probability density transformation method yields

$$\left|\mathrm{pdf}_{x_i}(x_i)\ \mathrm{d}x_i\right| = \left|\mathrm{pdf}_{g_i}(g_i)\ \mathrm{d}g_i\right| \tag{378}$$

$$\begin{aligned} \Rightarrow \quad \mathrm{pdf}_{x_i}(x_i) &= \mathrm{pdf}_{g_i}\big(u_i(x_i)\big) \left|\frac{\mathrm{d}u_i(x_i)}{\mathrm{d}x_i}\right| \\ &= \frac{\exp(-\frac{x_i^2}{2})}{2\Gamma\left(\frac{3}{2}\right)} \frac{1}{\sqrt{2}} \\ &= \frac{1}{\sqrt{2\pi}} \exp(-\frac{x_i^2}{2}), \end{aligned} \tag{379}$$

showing x_i is standard normally distributed. Accordingly, if x_1, x_2 and x_3 are independent standard normally distributed variables, and using the transformation in Eq. (377), it follows that

$$\frac{(x_1, x_2, x_3)^T/\sqrt{2}}{\sqrt{\frac{x_1^2}{2} + \frac{x_2^2}{2} + \frac{x_3^2}{2} + y}} \qquad \in \mathbb{R}^3 \tag{380}$$

is uniformly distributed over the unit ball in three dimensions.

Note: An exponential distribution can be obtained from an uniformly distributed random variable z in $[0,1]$ [129, 7.3] ($z \sim U(0,1)$) as follows: Let $z \sim U(0,1)$, i.e., the probability of z lying in $[z, z + \mathrm{d}z]$ reads

$$\mathrm{pdf}_z(z)\ \mathrm{d}z = \begin{cases} \mathrm{d}z & 0 \leq z \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{381}$$

Defining the transformation

$$z = u(y) := \exp(-y) \tag{382}$$

and using

$$\mathrm{pdf}_y(y) = \mathrm{pdf}_z(z) \left| \frac{\mathrm{d}z}{\mathrm{d}y} \right| \tag{383}$$

yields

$$\mathrm{pdf}_y(y) = \exp(-y), \tag{384}$$

showing that the transformation in Eq. (382) yields an exponentially distributed random variable y if $z \sim U(0,1)$. Hence, if using $y = u^{-1}(z) = -\ln(z)$, Eq. (380) can also be written as

$$\frac{(x_1, x_2, x_3)^T / \sqrt{2}}{\sqrt{\frac{x_1^2}{2} + \frac{x_2^2}{2} + \frac{x_3^2}{2} - \ln(z)}} \qquad \in \mathbb{R}^3, \tag{385}$$

where $x_1, x_2, x_3 \sim \mathcal{N}(x)$ and $z \sim U(0,1)$.

The method of random point picking from a ball described here, as well as the one from section 4.2, page 70, have been both implemented. No significant differences in the execution speed of the two methods (for the considered complex cluster systems described in section 4) were observed.

B. Appendix

B.1. Random Particle Placing

In section 4.2 it is mentioned (page 75) that the sequential particle setting procedure does not create configurations that are samples from the desired hard sphere equilibrium distribution.

In order to explain this issue consider the following: Let $\mathbf{r}$ a vector describing the discrete configuration of particles and $E(\mathbf{r})$ the appropriate energy of the system. Then, the hard sphere interaction potential energy yields

$$E(\mathbf{r}) = \begin{cases} \infty & \text{particles overlap} \\ 0 & \text{otherwise.} \end{cases} \tag{386}$$

Let N_0 the number of all configurations and N the number of non-overlapping particle configurations. The goal is to sample from the Boltzmann distribution

$$\begin{aligned} \mathrm{pdf}(\mathbf{r}) &= \frac{\exp\left(-\frac{E(\mathbf{r})}{k_B T}\right)}{\sum_{i=1}^{N_0} \exp\left(-\frac{E(\mathbf{r}_i)}{k_B T}\right)} = \frac{\exp\left(-\frac{E(\mathbf{r})}{k_B T}\right)}{N} \\ &= \begin{cases} 0 & \text{particles overlap} \\ 1/N & \text{otherwise} \end{cases} \end{aligned} \tag{387}$$

showing that each non-overlapping configuration has to be created with the same probability $1/N$. A simultaneous particle setting procedure creates each non-overlapping configuration with probability $1/N$, however this is not the case for a sequential particle setting procedure as it is illustrated by the following 1-dimensional examples: Consider two particles, which are represented by line segments of length 2, and which are to be randomly placed on a 1-dimension discrete grid (particles can be placed at -1.5, -0.5, 0.5, 1.5) such that there is no overlap of the particles, see Fig. 65. The grid is bounded, hence this case can be associated with the case of placing particles within a vesicle. As shown in Fig. 65, the simultaneous particle setting procedure produces each non-overlapping configuration with identical probability, as required by Eq. (387), but in case of the sequential particle setting procedure configurations are created with different probabilities.

The case of placing particles on a vesicle, i.e., there are no bounds, can be considered by means of circularly closing a 1-dimensional grid. Then, the sequential particle setting procedure will also not give the correct solution: Consider three particles, then by virtue of setting the first particle artificial bounds are introduced, and thus the other two particles are influenced in their setting as explained previously.

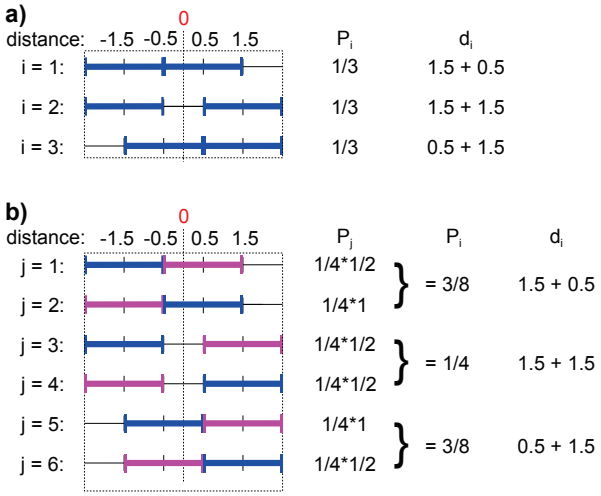


Figure 65: Setting two particles (represented by line segments) on a discrete grid. a) Simultaneous particle setting: There are three possible non-overlapping configurations, each occurring with the same probability. The probability of configuration i is P_i and the total distance of both particles to the origin is d_i. Thus, the average distance is $\sum_{i=1}^{3} P_i d_i = \frac{7}{3} = 2.\overline{3}$. b) Sequential particle setting: There are six possible ways to place the particles. The blue particle in each line is considered to be set at first, and subsequently the pink particle. However, since the particles are indistinguishable there are only three different configurations, which are occurring with different probability. The average distance is $\sum_{i=1}^{3} P_i d_i = \frac{9}{4} = 2.25$, and thus smaller as if the particles are set simultaneously.

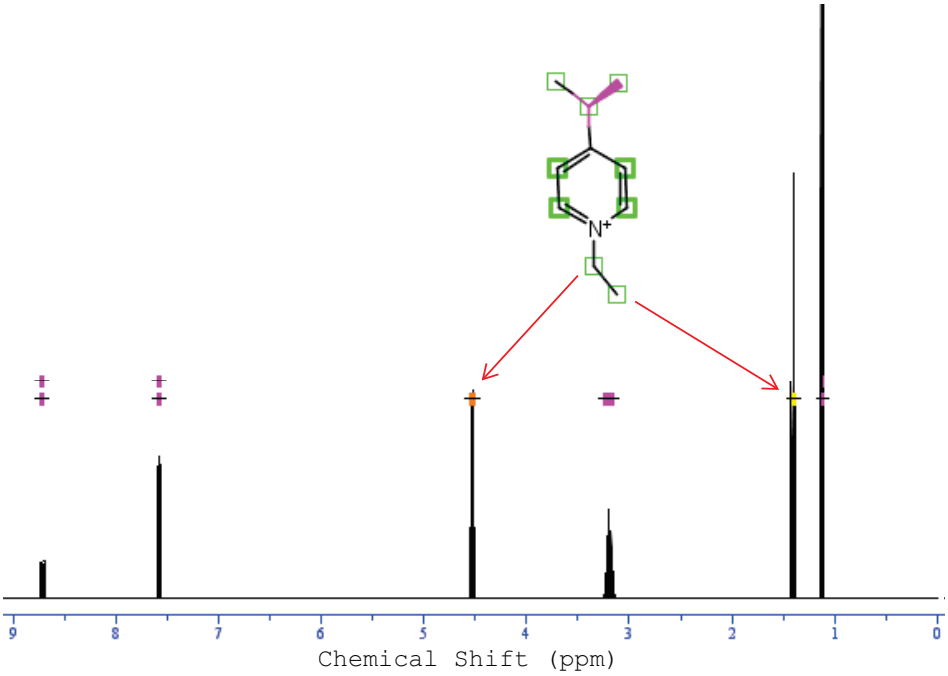


Figure 66: NMR spectrum simulation by means of [3]. There is an expected peak at ca. 4.5 ppm associated with the doublet at the ligand of the aromatic compound. There is another peak at ca. 1.4 ppm corresponding to the triplet of the ligand.

C. Appendix

C.1. NMR of P4VPQBr

Fig. 67 depicts the measured spectrum of ca. 90 % quaternized P4VPQBr (synthesized by Michaela Dzionara), and Fig. 66 depicts the simulated spectrum of the similar aromatic compound 1-ethyl-4-(propan-2-yl)pyridinium. In both spectra are peaks at ca. 7.6 ppm and at ca. 8.7 ppm, were each peak corresponds to two H atoms at the aromatic compound. Hence, the area of the peak at 8.7 ppm in the measured spectrum is set to two, and is used to define the reference area of one H atom.

The peak at ca. 4.5 ppm in the simulated spectrum corresponds to the two H atoms at the first C atom of the ligand (attached to the nitrogen atom of the aromatic compound). For a 90 % quaternization of the aromatic compounds with C_2H_5 ligands, the peak area should be 90 % of two times the reference area, i.e., $0.9 \cdot 2 = 1.8$. The corresponding peak in the measured spectrum has an area of 1.82, hence it can be considered that the chemical reaction (synthesis of 90 % quaternized P4VPQBr) was successful.

Another check can be considered for the three H atoms at the methyl group of the ligand. The peak of the methyl group is in the simulation at ca. 1.4 ppm, and has an area of three. Hence, the peak in the measured spectrum should have an area of $0.9 \cdot 3 = 2.7$, and the observed area is 3.1. The difference can easily be explained by the fact that the peak is superimposed with the rather broad peak on the l.h.s. of it.

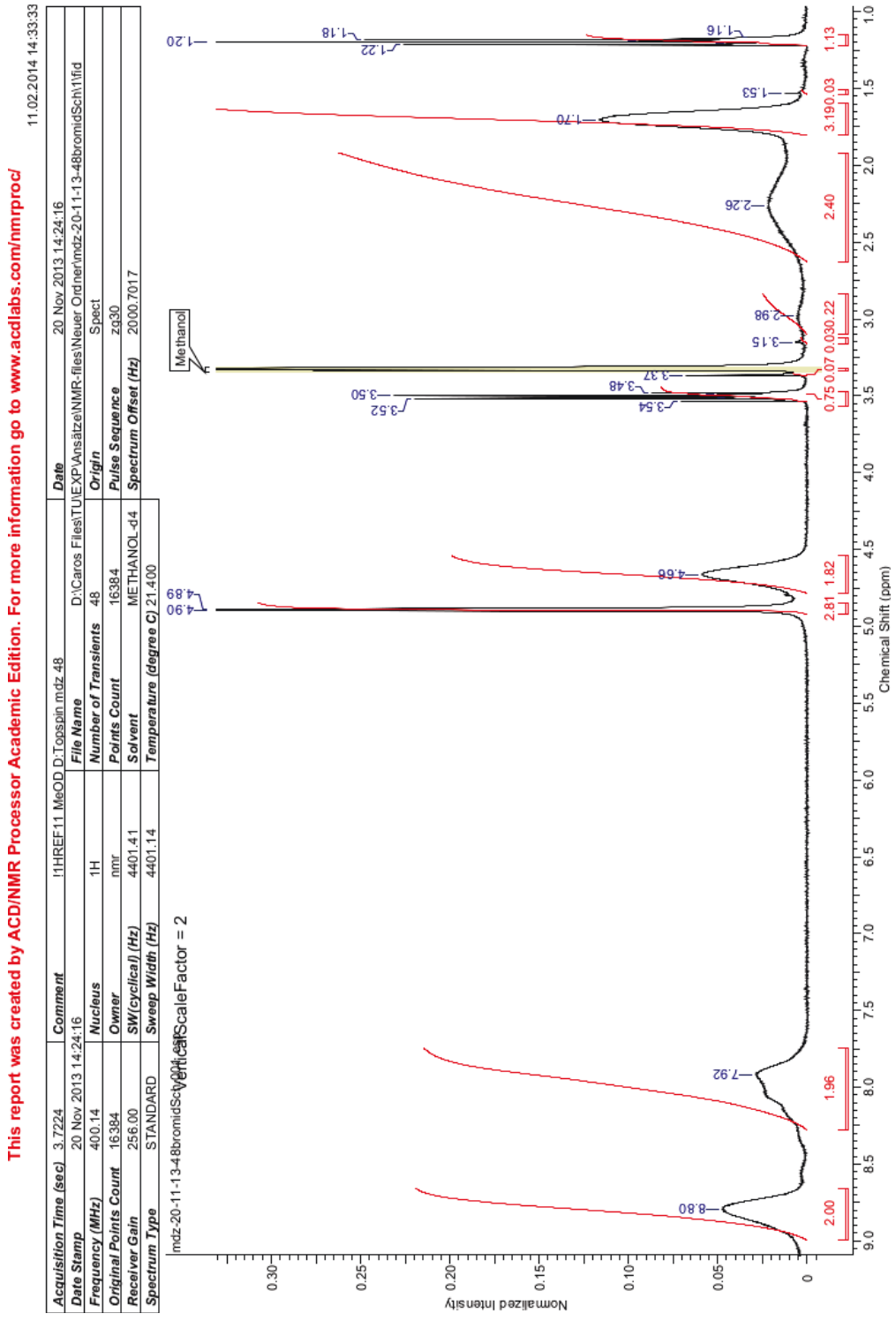


Figure 67: Measured NMR spectrum of P4VPQBr. P4VPQBr has been synthesized by Michaela Dzionara. NMR measurement has been done by Michaela Dzionara and Sven Riemer. Evaluation of the peaks in the NMR spectrum have been performed by Carolin Ganas.

C.2. Material Equations

The composition equations within the IPEC models of section 3.3 are given here.

In these equations, the apparent density is

$$\rho/(\mathrm{g{\cdot}cm^{-3}}) = \frac{M_w/(\mathrm{g{\cdot}mol^{-1}})}{v/\mathrm{nm^3}} \frac{10^{21}/(\mathrm{nm^3{\cdot}cm^{-3}})}{N_A/\mathrm{mol^{-1}}}, \tag{388}$$

and the scattering length density reads

$$\begin{aligned} \mathrm{SLD}/\mathrm{nm^{-2}} &= \frac{\mathrm{SL}/\mathrm{nm}}{v/\mathrm{nm^3}} \\ &= \frac{\mathrm{SL}/\mathrm{nm} \cdot \rho/(\mathrm{g{\cdot}cm^{-3}}) \cdot N_A/\mathrm{mol^{-1}}}{M_w/(\mathrm{g{\cdot}mol^{-1}}) \cdot 10^{21}/(\mathrm{nm^3{\cdot}cm^{-3}})}. \end{aligned} \tag{389}$$

SL is the scattering length, v is the apparent molecular volume, M_w the molecular weight, and N_A is the Avogadro constant ($N_A \approx 6.022 \cdot 10^{23}\,\mathrm{mol^{-1}}$). Following symbols and acronyms are used in the following material equations: DP: degree of polymerization, m: mass, N: quantity, Q^-: number of negative charges, Q^+: number of positive charges, Z: charge ratio, n: number density, V_{tot}: total system volume, $N_{\mathrm{agg.}}$: aggregation number, $\mathrm{pdf}(N_{\mathrm{agg.}})$: aggregation number distribution, φ_h: hydration.

For PIB, of unit IB (C_4H_8), the following quantities are known: DP(PIB), SL(IB), M_w(IB), ρ(PIB), m(PIB-PMAA), M_w(PIB-PMAA).

Then, following quantities are calculated:

$$v(\mathrm{IB}) = \frac{M_w(\mathrm{IB})10^{21}}{\rho(\mathrm{PIB})N_A} \tag{390}$$

$$\mathrm{SLD(PIB)} = \mathrm{SL(IB)}/v(\mathrm{IB}) \tag{391}$$

$$M_w(\mathrm{PIB}) = \mathrm{DP(PIB)}M_w(\mathrm{IB}) \tag{392}$$

$$v(\mathrm{PIB}) = \mathrm{DP(PIB)}v(\mathrm{IB}) \tag{393}$$

$$\begin{aligned} N(\mathrm{PIB}) = N(\mathrm{PMAA}) &= N(\mathrm{PIB\text{-}PMAA}) \\ &= \frac{m(\mathrm{PIB\text{-}PMAA})}{M_w(\mathrm{PIB\text{-}PMAA})} \end{aligned} \tag{394}$$

$$N(\mathrm{IB}) = \mathrm{DP(PIB)}N(\mathrm{PIB}) \tag{395}$$

For PMAA and PMANa, with respective units MAA ($C_4H_6O_2$) and MANa ($C_4H_5O_2Na$), the following quantities are known: DP(PMAA) = DP(PMANa), SL(MANa), M_w(MANa), ρ(PMANa).

Then, following quantities are calculated:

$$v(\mathrm{MANa}) = \frac{M_w(\mathrm{MANa})10^{21}}{\rho(\mathrm{MANa})N_A} \tag{396}$$

$$\mathrm{SLD(PMANa)} = \mathrm{SL(MANa)}/v(\mathrm{MANa}) \tag{397}$$

$$M_w(\mathrm{PMANa}) = \mathrm{DP(PMAA)}M_w(\mathrm{MANa}) \tag{398}$$

$$v(\mathrm{PMANa}) = \mathrm{DP(PMAA)}v(\mathrm{MANa}) \tag{399}$$

$$N(\mathrm{MAA}) = N(\mathrm{MANa}) = \mathrm{DP(PMANa)}N(\mathrm{PMAA}) \tag{400}$$

$$Q^- = N(\mathrm{MAA}) \tag{401}$$

For d5-P4VPQBr$_\alpha$, α-percentage of the 4VP units (C_7H_7N) are quaternized with C_2D_5Br, yielding d5-4VPQBr$_\alpha$ ($C_{7+2\alpha}H_7D_{5\alpha}NBr_\alpha$), and the following quantities are known: α, m(d5-P4VPQBr$_\alpha$), M_w(d5-P4VPQBr$_\alpha$) , SL(d5-4VPQBr$_\alpha$)

Then, following quantities are calculated:

$$\begin{aligned} N(\text{d5-P4VPQBr}_\alpha) = N(\text{P4VP}) &= \frac{m(\text{d5-P4VPQBr}_\alpha)}{M_w(\text{d5-P4VPQBr}_\alpha)} \\ &= \frac{m(\text{d5-P4VPQBr}_\alpha)}{M_w(\text{P4VP}) + \alpha \text{DP(P4VP)} M_w(\text{C}_2\text{D}_5\text{Br})} \end{aligned} \tag{402}$$

$$\text{DP(P4VP)} = \frac{M_w(\text{P4VP})}{M_w(\text{4VP})} \tag{403}$$

$$N(4VP) = N(\text{P4VP})\text{DP(P4VP)} \tag{404}$$

$$Q^+ = \alpha N(\text{4VP}) \tag{405}$$

For all samples it is considered that d5-P4VPQBr$_\alpha$ builds a complex with PMANa, therefore the apparent density of d5-P4VPQBr$_\alpha$ is not relevant, only the apparent density of the (PMANa+d5-P4VPQBr$_\alpha$)-complex $\rho_{\text{complex}}(Z)$ is needed, which however is unknown and hence fitted.

There are two main model groups:

(i) **Core-shell or core-corona models:** The overall material of the complex (that is in the shell or in the corona) consists of the total PMANa and d5-P4VPQBr$_\alpha$, and the charge ratio is $Z = Q^+/Q^- = \frac{\alpha N(\text{4VP})}{N(\text{MANa})} \approx 0.4$. The overall fraction of MANa that is in the shell or in the corona is $f = 1$.

(ii) **Core-shell-corona models:** The overall fraction of MANa that is in the shell is

$$f = \frac{\alpha N(\text{4VP})}{N(\text{MANa})}, \tag{406}$$

and the charge ratio in the shell is $Z = 1$. The fraction of MANa that is in the corona is $1 - f$.

For these two cases following quantities are calculated:

$$\langle v_{\text{core}}(N_{\text{agg.}}) \rangle = v(\text{PIB}) \int N_{\text{agg.}} \, \text{pdf}(N_{\text{agg.}}) \, \text{d}N_{\text{agg.}} \tag{407}$$

$$n/\text{nm}^{-3} = \frac{v(\text{PIB})/\text{nm}^3}{\langle v_{\text{core}}(N_{\text{agg.}}) \rangle/\text{nm}^3} \frac{N(\text{PIB})/\text{mol}}{V_{\text{tot}}/\text{L}} \frac{N_A/\text{mol}^{-1}}{10^{24}/(\text{nm}^3\text{L}^{-1})} \tag{408}$$

$$M_w(\text{complex,Z}) = M_w(\text{MANa}) + Z \left(\frac{M_w(\text{4VP})}{\alpha} + M_w(\text{C}_2\text{D}_5\text{Br}) \right) \tag{409}$$

$$v(\text{complex,Z}) = \frac{M_w(\text{complex,Z})}{\rho(\text{complex,Z})} \frac{10^{21}}{N_A} \tag{410}$$

$$\text{SLD(complex,Z)} = \frac{\text{SL(MANa)} + Z \left(\frac{\text{SL(4VP)}}{\alpha} + \text{SL}(\text{C}_2\text{D}_5\text{Br}) \right)}{v(\text{complex,Z})} \tag{411}$$

$v_{\text{core}}(N_{\text{agg.}})$ is the core volume of one object with a PIB-PMANa aggregation number of $N_{\text{agg.}}$, and its appropriate shell and/or corona material volume (i.e., without hydration) is/are for the

(i) **core-shell or core-corona models:**

$$v_{(\text{shell or corona})}(N_{\text{agg.}}) = N_{\text{agg.}} v(\text{complex,Z})\text{DP}(\text{PMANa}) \tag{412}$$

(ii) **core-shell-corona models:**

$$v_{\text{shell}}(N_{\text{agg.}}) = N_{\text{agg.}} v(\text{complex,Z})\text{DP}(\text{PMANa}) f \tag{413}$$

$$v_{\text{corona}}(N_{\text{agg.}}) = N_{\text{agg.}} v(\text{MANa})\text{DP}(\text{PMANa})(1-f) \tag{414}$$

The shell thickness t_s is determined as follows (φ_h: is the hydration fraction of the shell, R_s: is the shell radius, $R_i = (3v_{\text{core}}(N_{\text{agg.}})/(4\pi))^{1/3}$: is the core radius):

$$R_s = \left(\frac{3}{4\pi}\left(\frac{v_{\text{shell}}(N_{\text{agg.}})}{1-\varphi_h} + v_{\text{core}}(N_{\text{agg.}})\right)\right)^{1/3} \tag{415}$$

$$\Rightarrow \qquad t_s = R_s - R_i, \tag{416}$$

and the corona thickness t_c is determined as follows:
Linear Corona: The thickness t_c of the linear corona, which is defined in the interval $[R_1, R_1 + t_c]$, is calculated from the total volume of material (complex or PMANa) in the corona and Eq. (85) ($\varphi'_{\text{int.}}$: interfacial volume fraction, see section 3.3.1):

$$\int_{R_1}^{R_1+t_c} \varphi_{\text{lin.}}(r) 4\pi r^2 \, \mathrm{d}r \stackrel{!}{=} v_{\text{corona}}(N_{\text{agg.}}) \tag{417}$$

$$\begin{aligned}
\Rightarrow \quad t_c = \frac{1}{6\pi h}\Bigg(& \big[352 R_1^3 \pi \varphi'_{\text{int.}} + 324 v_{\text{corona}}(N_{\text{agg.}}) + \\
& 36\sqrt{96 R_1^6 \pi^2 (\varphi'_{\text{int.}})^2 + 176 R_1^3 \pi \varphi'_{\text{int.}} v_{\text{corona}}(N_{\text{agg.}}) + 81 v_{\text{corona}}(N_{\text{agg.}})^2}\big] \pi^2 (\varphi'_{\text{int.}})^2 \Bigg)^{1/3} \\
& - R_s^2 \pi \varphi'_{\text{int.}} \Bigg(\big[352 R_1^3 \pi \varphi'_{\text{int.}} + 324 v_{\text{corona}}(N_{\text{agg.}}) \\
& + 36\sqrt{96 R_1^6 \pi^2 (\varphi'_{\text{int.}})^2 + 176 R_1^3 \pi \varphi'_{\text{int.}} v_{\text{corona}}(N_{\text{agg.}}) + 81 v_{\text{corona}}(N_{\text{agg.}})^2}\big] \pi^2 (\varphi'_{\text{int.}})^2 \Bigg)^{-1/3} \\
& - \frac{4}{3} R_1.
\end{aligned} \tag{418}$$

Algebraic Corona: The thickness t_c of the algebraic corona, defined in the interval $[R_1, R_c]$, is calculated from the total volume of material (complex or PMANa) in the corona and Eq. (89):

$$\int_{R_1}^{R_1+t_c} \varphi_{\text{alg.}}(r) 4\pi r^2 \, \mathrm{d}r \stackrel{!}{=} v_{\text{corona}}(N_{\text{agg.}}) \tag{419}$$

$$\Rightarrow \quad t_c = R_s \left(\left(\frac{4\varphi'_{\text{int.}} \pi R_1^3 - v_{\text{corona}}(N_{\text{agg.}})\alpha + 3 v_{\text{corona}}(N_{\text{agg.}})}{4\varphi'_{\text{int.}} \pi R_1^3} \right)^{\frac{1}{3-\alpha}} - 1 \right). \tag{420}$$

The last two integration results have been obtained via MAPLE.

D. Appendix

D.1. Monodisperse Hard Sphere Structure Factor in the PY Approach

In the following $S_1(q)$ is the structure factor formula for particles that interact via a hard sphere interaction potential. Consider a hard sphere radius r_{HS} and a particle number density n, then the appropriate volume fraction reads $\phi_{\mathrm{HS}} = n\frac{4\pi}{3}r_{\mathrm{HS}}^3$. Accordingly, for the Percus-Yevick closure approximation [154, 170, 85] the structure factor formula reads

$$S_1(q) = \left(1 + \frac{24\eta G(2r_{\mathrm{HS}}q)}{2r_{\mathrm{HS}}q}\right)^{-1}, \tag{421}$$

where

$$\begin{aligned} G(A) = \frac{\alpha(\sin(A) - A\cos(A))}{A^2} + \frac{\beta(2A\sin(A) + (2 - A^2)\cos(A) - 2)}{A^3} \\ + \frac{\gamma\left(-A^4\cos(A) + 4\big[(3A^2 - 6)\cos(A) + (A^3 - 6A)\sin(A) + 6\big]\right)}{A^5} \end{aligned} \tag{422}$$

and

$$\alpha = \frac{(1 + 2\phi_{\mathrm{HS}})^2}{(1 - \phi_{\mathrm{HS}})^4}, \tag{423}$$

$$\beta = \frac{-6\phi_{HS}(1 + \frac{\phi_{\mathrm{HS}}}{2})^2}{(1 - \phi_{\mathrm{HS}})^4}, \text{ and} \tag{424}$$

$$\gamma = \frac{\phi_{\mathrm{HS}}\alpha}{2}. \tag{425}$$

D.2. Partial Structure Functions for a Hard Sphere Potential in the PY Approach of a Polydispserse System

In the following, the partial structure functions are given for a polydisperse collection of particles that interact via hard sphere potentials. The solution is attained in the Percus-Yevick closure approximation. The partial structure function between two particles having hard sphere potential diameters σ_i and σ_j is [56, 17, 18]

$$H(q; \sigma_i, \sigma_j) = -2\frac{Z_2Z_3 + Z_1Z_4}{q^3(X^2 + Y^2)}, \tag{426}$$

where

$$Z_1 = Y\sin(q\sigma_{ij}) - X\cos(q\sigma_{ij}), \tag{427}$$

$$Z_2 = X\sin(q\sigma_{ij}) + Y\cos(q\sigma_{ij}), \tag{428}$$

$$Z_3 = Q'' - qR_3, \quad \text{and} \tag{429}$$

$$Z_4 = qQ'(\sigma_i, \sigma_j) + qR_4. \tag{430}$$

Therein, the quantities are

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}, \tag{431}$$

$$Q'' = \frac{2\pi}{\Delta}\left(1 + \frac{\xi_3 \pi}{2\Delta}\right), \tag{432}$$

$$Q'(\sigma_i, \sigma_j) = \frac{\pi}{\Delta}\left(\sigma_i + \sigma_j + \frac{\sigma_i \sigma_j \xi_2 \pi}{2\Delta}\right), \tag{433}$$

$$\Delta = 1 - \frac{\pi \xi_3}{6}, \tag{434}$$

$$\tag{435}$$

and ξ_i, for $i = 2, 3$, is the i-th moment of the used probability density function, $\mathrm{pdf}(\sigma)$, times the number density n, i.e.,

$$\xi_i = n \int_0^\infty \sigma^i \,\mathrm{pdf}(\sigma)\, \mathrm{d}\sigma. \tag{436}$$

The other quantities are

$$R_3 = n\left(\frac{\pi}{\Delta}\right)^2 \int_0^\infty (\sigma - \sigma_i)(\sigma - \sigma_j)\, y_1(\sigma)\, \mathrm{pdf}(\sigma)\, \mathrm{d}\sigma, \tag{437}$$

$$R_4 = n\left(\frac{\pi}{\Delta}\right)^2 \int_0^\infty (\sigma - \sigma_i)(\sigma - \sigma_j)\, x_1(\sigma)\, \mathrm{pdf}(\sigma)\, \mathrm{d}\sigma, \tag{438}$$

$$\begin{aligned} X = 1 &- nQ'' \int_0^\infty x_2(\sigma)\, \mathrm{pdf}(\sigma)\, \mathrm{d}\sigma - n\frac{2\pi}{\Delta} \int_0^\infty \sigma x_1(\sigma) \left(1 + \frac{\pi \xi_2 \sigma}{4\Delta}\right) \mathrm{pdf}(\sigma)\, \mathrm{d}\sigma \\ &- \frac{1}{2}\left(\frac{\pi n}{\Delta}\right)^2 \int_0^\infty \int_0^\infty \big[x_1(\sigma')x_1(\sigma'') - y_1(\sigma')y_1(\sigma'')\big](\sigma' - \sigma'')^2 \mathrm{pdf}(\sigma')\mathrm{pdf}(\sigma'')\, \mathrm{d}\sigma'\, \mathrm{d}\sigma'', \end{aligned} \tag{439}$$

and

$$\begin{aligned} Y = -nQ'' &\int_0^\infty y_2(\sigma)\, \mathrm{pdf}(\sigma)\, \mathrm{d}\sigma - n\frac{2\pi}{\Delta} \int_0^\infty \sigma y_1(\sigma) \left(1 + \frac{\pi \xi_2 \sigma}{4\Delta}\right) \mathrm{pdf}(\sigma)\, \mathrm{d}\sigma \\ &- \frac{1}{2}\left(\frac{\pi n}{\Delta}\right)^2 \int_0^\infty \int_0^\infty \big[x_1(\sigma')y_1(\sigma'') + y_1(\sigma')x_1(\sigma'')\big](\sigma' - \sigma'')^2 \mathrm{pdf}(\sigma')\mathrm{pdf}(\sigma'')\, \mathrm{d}\sigma'\, \mathrm{d}\sigma'', \end{aligned} \tag{440}$$

where

$$x_1(\sigma) = [\cos(q\sigma) - 1]/q^2, \tag{441}$$

$$y_1(\sigma) = [q\sigma - \sin(q\sigma)]/q^2, \tag{442}$$

$$x_2(\sigma) = [q\sigma - \sin(q\sigma)]/q^3, \tag{443}$$

$$y_2(\sigma) = -\big[\cos(q\sigma) + (q^2\sigma^2)/2 - 1\big]/q^3. \tag{444}$$

Scattering of Schulz(-Zimm) distribution Homogeneous Spheres In case of Schulz(-Zimm) distributed homogeneous spheres, the integrals involved in the scattering function can be solved analytically [56]. The Schulz probability density function reads as

$$\mathrm{pdf}(\sigma) = \mathrm{schulzpdf}(\sigma; b, c) = \left(\frac{\sigma}{b}\right)^{c-1} \frac{\exp(-\sigma/b)}{b\Gamma(c)}, \tag{445}$$

where b, c are some parameters. Then, the first three moments of the distribution are

$$\langle\sigma\rangle = bc, \tag{446}$$
$$\langle\sigma^2\rangle = b^2c(c+1), \quad \text{and} \tag{447}$$
$$\langle\sigma^3\rangle = b^3c(c+1)(c+2), \tag{448}$$

and the scattering intensity becomes

$$I(q) = I_1(q) + I_2(q), \tag{449}$$

where

$$I_1(q) = 8\pi^2(\Delta\text{SLD})^2 q^{-6}[1 - \chi - q\psi' + \frac{1}{4}q^2(\zeta'' + \chi'')], \tag{450}$$

and

$$\begin{aligned} I_2(q) = -2\{ & \\ & \Lambda\big[\Lambda(Y\delta_1 - X\delta_6) + \Lambda'(Y\delta_2 - X\delta_4) + M(X\delta_1 + Y\delta_6) + M'(X\delta_2 + Y\delta_4)\big] \\ & +\Lambda'\big[\Lambda(Y\delta_2 - X\delta_4) + \Lambda'(Y\delta_3 - X\delta_5) + M(X\delta_2 + Y\delta_4) + M'(X\delta_3 + Y\delta_5)\big] \\ & +M\big[\Lambda(X\delta_1 + Y\delta_6) + \Lambda'(X\delta_2 + Y\delta_4) + M(X\delta_6 - Y\delta_2) + M'(X\delta_4 - Y\delta_2)\big] \\ & +M'\big[\Lambda(X\delta_2 + Y\delta_4) + \Lambda'(X\delta_3 + Y\delta_5) + M(X\delta_4 - Y\delta_2) + M'(X\delta_5 - Y\delta_3)\big] \\ & \}/\big[q^3(X^2 + Y^2\big]. \end{aligned} \tag{451}$$

Therein, the quantities are

$$\Lambda = n2\pi\Delta\text{SLD}[\psi - q(\zeta' + \chi')/2]/q^3, \tag{452}$$
$$\Lambda' = n2\pi\Delta\text{SLD}[\psi' - q(\zeta'' + \chi'')/2]/q^3, \tag{453}$$
$$M = n2\pi\Delta\text{SLD}[1 - \psi - q\psi'/2]/q^3, \tag{454}$$
$$M' = n2\pi\Delta\text{SLD}[\zeta' - \psi' - q\psi''/2]/q^3, \tag{455}$$

$$\begin{aligned} X = 1 - nQ''(q\zeta' - \psi)/q^3 - n\frac{2\pi}{q^2\Delta}\left[(\psi' - \zeta') + \frac{\pi\xi_2}{4\Delta}(\chi'' - \zeta'')\right] \\ - \left(\frac{\pi}{\Delta}\right)^2\left(\frac{n}{q^2}\right)^2\left[(\chi - 1)(\chi'' - \zeta'') - (\chi' - \zeta')^2 - (q\zeta' - \psi)(q\zeta''' - \psi''') + (q\zeta'' - \psi')^2\right], \end{aligned} \tag{456}$$

$$\begin{aligned} Y = nQ''(q\zeta' - \psi)/q^3 - n\frac{2\pi}{q^2\Delta}\left[q\zeta'' - \psi' + \frac{\pi\xi_2}{4\Delta}(q\zeta''' - \psi'')\right] \\ - \left(\frac{\pi}{\Delta}\right)^2\left(\frac{n}{q^2}\right)^2\left[(q\zeta' - \psi)(\chi'' - \zeta'') - 2(q\zeta'' - \psi')(\chi' - \zeta') + (q\zeta''' - \psi'')(\psi - 1)\right], \end{aligned} \tag{457}$$

$g(\sigma)$	n	$\frac{1}{b^c\Gamma(c)}\int_0^\infty \sigma^n g(\sigma)\mathrm{pdf}(\sigma)\ \mathrm{d}\sigma$	Notation
1	0	1	1
1	1	bc	ζ'
1	2	$b^2c(c+1)$	ζ''
1	3	$b^3c(c+1)(c+2)$	ζ'''
$\sin(q\sigma)$	0	$v_1^{c/2}\sin[c\tan^{-1}(bq)]$	ψ
$\sin(q\sigma)$	1	$bcv_1^{(c+1)/2}\sin[(c+1)\tan^{-1}(bq)]$	ψ'
$\sin(q\sigma)$	2	$b^2c(c+1)v_1^{(c+2)/2}\sin[(c+2)\tan^{-1}(bq)]$	ψ''
$\sin(q\sigma/2)$	0	$2^c v_2^{c/2}\sin[c\tan^{-1}(bq/2)]$	μ
$\sin(q\sigma/2)$	1	$2^{c+1}bcv_2^{(c+1)/2}\sin[(c+1)\tan^{-1}(bq/2)]$	μ
$\cos(q\sigma)$	0	$v_1^{c/2}\cos[c\tan^{-1}(bq)]$	χ
$\cos(q\sigma)$	1	$bcv_1^{(c+1)/2}\cos[(c+1)\tan^{-1}(bq)]$	χ'
$\cos(q\sigma)$	2	$b^2c(c+1)v_1^{(c+2)/2}\cos[(c+2)\tan^{-1}(bq)]$	χ''
$\cos(q\sigma/2)$	0	$2^c v_2^{c/2}\cos[c\tan^{-1}(bq/2)]$	λ
$\cos(q\sigma/2)$	1	$2^{c+1}bcv_2^{(c+1)/2}\cos[(c+1)\tan^{-1}(bq/2)]$	λ'

Table 8: Integral notations from [56]. Here, $v_m = [m^2 + (qk)^2]^{-1}$, and $\mathrm{pdf}(\sigma)$ is the Schulz distribution.

and

$$\delta_1 = (\pi/\Delta)\left(2 + \frac{\pi}{\Delta}\left[\xi_3 - \frac{n}{q}(q\zeta''' - \psi'')\right]\right), \tag{458}$$

$$\delta_2 = \left(\frac{\pi}{\Delta}\right)^2 \frac{n}{q}(q\zeta'' - \psi'), \tag{459}$$

$$\delta_3 = -\left(\frac{\pi}{\Delta}\right)^2 \frac{n}{q}(q\zeta' - \psi), \tag{460}$$

$$\delta_4 = \frac{\pi}{\Delta}\left[q - \frac{\pi n}{q\Delta}(\chi' - \zeta')\right], \tag{461}$$

$$\delta_5 = \left(\frac{\pi}{\Delta}\right)^2 \left[\frac{n}{q}(\chi - 1) + \frac{1}{2}q\xi_2\right], \tag{462}$$

$$\delta_6 = \left(\frac{\pi}{\Delta}\right)^2 \frac{n}{q}(\chi'' - \zeta''), \tag{463}$$

and the remaining parameters/expressions are defined in Table 8. Furthermore, the average structure factor can be written as [55]

$$\begin{aligned} S_{\mathrm{ave}}(q) &= n\int_0^\infty \mathrm{pdf}(\sigma)\ \mathrm{d}\sigma + n^2\int_0^\infty\int_0^\infty h(q,\sigma',\sigma'')\mathrm{pdf}(\sigma')\mathrm{pdf}(\sigma'')\ \mathrm{d}\sigma'\ \mathrm{d}\sigma'' \\ &= 1 + nh(q), \end{aligned} \tag{464}$$

where

$$
\begin{aligned}
h(q) = -2\big\{ & \\
& \lambda\big[\lambda(Y\delta_1 - X\delta_6) + \lambda'(Y\delta_2 - X\delta_4) + \mu(X\delta_1 + Y\delta_6) + \mu'(X\delta_2 + Y\delta_4)\big] \\
& +\lambda'\big[\lambda(Y\delta_2 - X\delta_4) + \lambda'(Y\delta_3 - X\delta_5) + \mu(X\delta_2 + Y\delta_4) + \mu'(X\delta_3 + Y\delta_5)\big] \\
& +\mu\big[\lambda(X\delta_1 + Y\delta_6) + \lambda'(X\delta_2 + Y\delta_4) + \mu(X\delta_6 - Y\delta_2) + \mu'(X\delta_4 - Y\delta_2)\big] \\
& +\mu'\big[\lambda(X\delta_2 + Y\delta_4) + \lambda'(X\delta_3 + Y\delta_5) + \mu(X\delta_4 - Y\delta_2) + \mu'(X\delta_5 - Y\delta_3)\big] \\
& \big\}/\big[q^3(X^2 + Y^2\big].
\end{aligned}
\tag{465}
$$

The implemented formula was counter-checked by means of the scattering plots obtained in this work and the structure factor plots in [55, 56].

D.3. Intensity Form Factor of a Randomly Oriented Disk

The amplitude form factor of a homogeneous, circular cylindrical disk can be calculated by considering the Fourier transform of the SLD profile of a disk in cylindrical coordinates (ϱ, φ, z) [133, §B.5], where the z axis coincident with the symmetry axis of the disk, ϱ being orthogonal to it, and φ is the polar angle. Then,

$$
A_{\text{disk}}(\mathbf{q}, R, t) = 2\pi\Delta_{\text{SLD}} \int_{-\frac{t}{2}}^{\frac{t}{2}} \int_0^R J_0(q_\varrho \varrho)\varrho\, e^{-iq_z z}\, \mathrm{d}\varrho\, \mathrm{d}z, \tag{466}
$$

where

$$
q_z = q\cos(\theta), \text{ and} \tag{467}
$$

$$
q_\varrho = q\sin(\theta) \tag{468}
$$

with $q = \|\mathbf{q}\|$, and θ is the angle enclosed between the scattering vector $\mathbf{q}$ and the z axis. Δ_{SLD} is the SLD difference between the homogeneous disk and the matrix, and J_0 is the zeroth order cylindrical Bessel function. The integration over z yields

$$
A_{\text{disk}}(q, \theta, R, t) = 2\pi\Delta_{\text{SLD}} \frac{2\sin(q_z t/2)}{q_z} \int_0^R J_0(q_\varrho \varrho)\varrho\, \mathrm{d}\varrho. \tag{469}
$$

Defining $\tilde{\varrho} = \varrho R$ and using the Bessel equality $[\zeta^m J_m(\zeta)]' = \zeta^m J_{m-1}(\zeta)$ [68, §6.2.3] yields

$$
\begin{aligned}
A_{\text{disk}}(q, \theta, R, t) &= 2\pi\Delta_{\text{SLD}} \frac{2\sin(q_z t/2)}{q_z} \frac{1}{q_\varrho^2} \int_0^{q_\varrho R} J_0(\tilde{\varrho})\tilde{\varrho}\, \mathrm{d}\tilde{\varrho} \\
&= 4\pi\Delta_{\text{SLD}} \frac{\sin(q_z t/2)}{q_z} \frac{1}{q_\varrho} R J_1(q_\varrho R) \\
&= 2\pi\Delta_{\text{SLD}} t \frac{\sin(q\cos(\theta)t/2)}{q\cos(\theta)t/2} \frac{1}{q\sin(\theta)} R J_1(qR\sin(\theta)) \\
&= (\pi R^2 t)\Delta_{\text{SLD}} \cdot \frac{\sin(q\cos(\theta)t/2)}{q\cos(\theta)t/2} \cdot \frac{2J_1(qR\sin(\theta))}{qR\sin(\theta)},
\end{aligned}
\tag{470}
$$

and the appropriate forward amplitude form factor reads

$$
\lim_{q\to 0} A_{\text{disk}}(q, \theta, R, t) = (\pi R^2 t) \cdot \Delta_{\text{SLD}}, \tag{471}
$$

since the limit of each of the last two terms in the last line of Eq. (470) converges to one. Squaring the amplitude form factor and performing an orientational average yields the intensity form factor of an orientationally averaged intensity form factor of a disk

$$I_{\text{disk}}(q, R, t) = \frac{1}{2} \int_0^{\pi} A^2_{\text{disk}}(q, \theta, R, t) \sin(\theta) \, \mathrm{d}\theta. \tag{472}$$

The integration cannot be solved analytically, hence requires a numerical integration.

E. Appendix

E.1. Marginal and Conditional Gaussians

In the following, lower case bold letters are vectors, and upper case bold letters are matrices. Then, following relations for marginal and conditional Gaussian distributions hold [16, §2.3.3]: Let

$$\text{pdf}(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}) \tag{473}$$

a marginal distribution for $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Lambda}^{-1}$, and let

$$\text{pdf}(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1}) \tag{474}$$

a conditional distribution for $\mathbf{y}$ given $\mathbf{x}$ with mean $\mathbf{A}\mathbf{x} + \mathbf{b}$ and covariance matrix $\mathbf{L}^{-1}$, then the marginal distribution for $\mathbf{y}$ reads

$$\text{pdf}(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^T) \tag{475}$$

and the conditional distribution for $\mathbf{x}$ given $\mathbf{y}$ is

$$\text{pdf}(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma}) \tag{476}$$

where

$$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^T\mathbf{L}\mathbf{A})^{-1}. \tag{477}$$

F. Appendix

F.1. Regularization Matrix

In the original work of Glatter [49, 26] the regularization term is defined by the matrix

$$\mathbf{L}_D^T\mathbf{L}_D := \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & 1 & & & & \\ & -1 & 2 & -1 & & & \\ & & & \ddots & & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 1 & \end{pmatrix}, \tag{478}$$

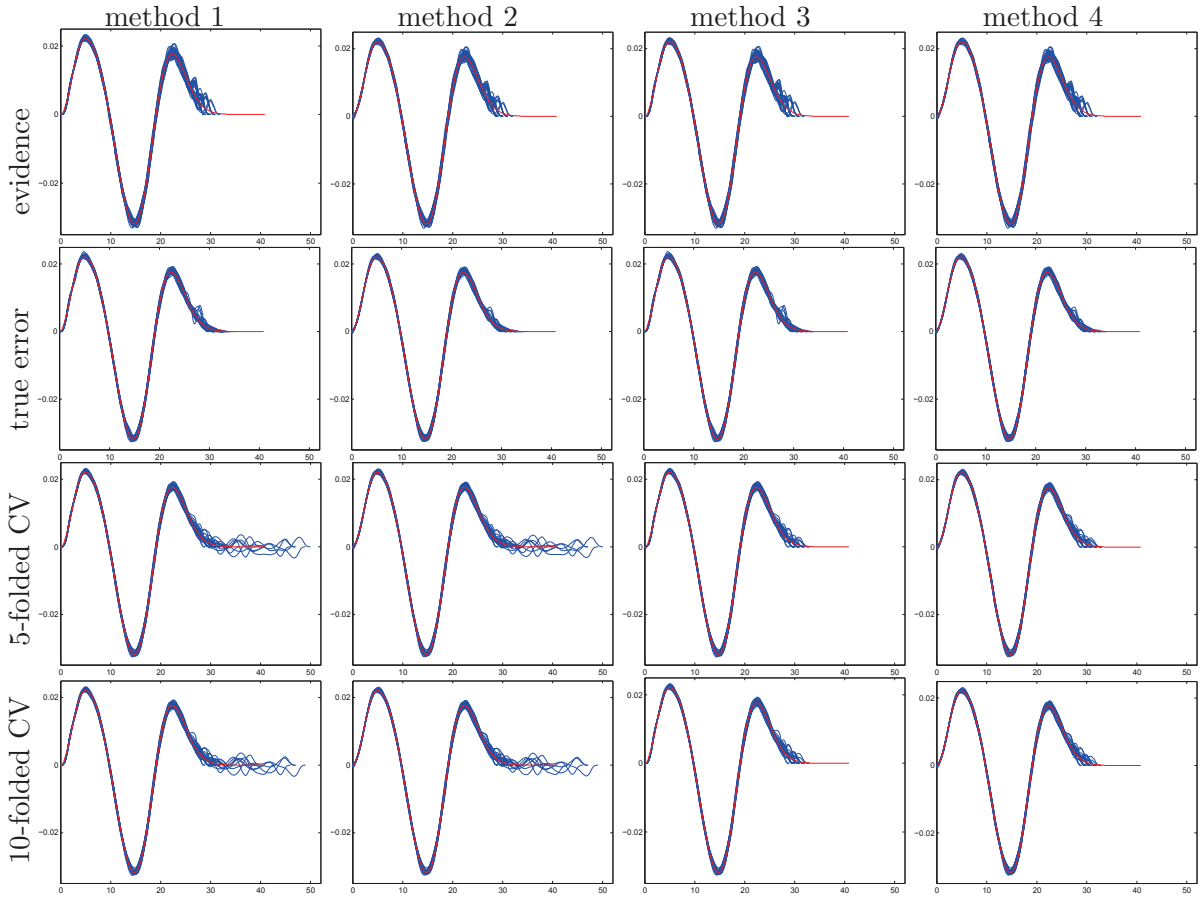
where

$$\mathbf{L}_D = \begin{pmatrix} -1 & 1 & & & & & \\ 0 & -1 & 1 & & & & \\ & 0 & -1 & 1 & & & \\ & & & \ddots & & & \\ & & & 0 & -1 & 1 & \\ & & & & 0 & 0 & \end{pmatrix} \tag{479}$$

is a discrete first order differential operator[82].

Fig. 68 shows the obtained PDDFs for the case that this regularization matrix is applied in the IFT method in the inter-model comparison approach described in section 7.2.3. Comparing these results with the ones shown in Fig. 55, where the second order derivative operator $\mathbf{L}_{D^2}$ is used, it is seen that the PDDFs obtained with the second order derivative operator $\mathbf{L}_{D^2}$ are better (smoother, less spreads and closer to the true PDDF) than the ones obtained by using the first order derivative operator $\mathbf{L}_D$.

[82] Usually, the last line in the derivative matrix $\mathbf{L}_D$ is removed, see, e.g., [63, §8.1]). Note, $\mathbf{L}_D$ does not have full rank.



	method 1	method 2	method 3	method 4
evidence	(27.8, 0.91) nm	(27.8, 0.91) nm	(27.8, 0.91) nm	(27.8, 0.91) nm
true error	(30.3, 1.40) nm	(30.2, 1.34) nm	(29.9, 1.27) nm	(29.9, 1.23) nm
5-folded cross-validation	(32.3, 6.42) nm	(32.3, 6.42) nm	(28.8, 1.09) nm	(28.8, 1.13) nm
10-folded cross-validation	(32.1, 6.06) nm	(31.9, 5.23) nm	(29.8, 1.02) nm	(28.7, 1.05) nm

Figure 68: Obtained PDDFs ($p(r)/(\mathrm{cm}^{-1}\cdot\mathrm{nm}^{-1})$ vs. r/nm) for the inter-model comparison approach using different objective function methods. The regularization term in Eq. (478) is used. For each method 32 i.i.d. scattering data sets are simulated for the zero average contrast condition case, and accordingly the 32 appropriate determined PDDFs are plotted in one diagram (blue curves). The red curves show the true PDDFs. 1st column (method 1): PDDFs are calculated for the global extrema of the objective functions and background is subtracted automatically. 2nd column (method 2): PDDFs are calculated for the global extrema of the objective functions and background is subtracted a priori. 3rd column (method 3): PDDFs are calculated for extrema in accordance with the principle of Occam's razor and background is subtracted automatically. 4rd column (method 4): PDDFs are calculated for extrema in accordance with the principle of Occam's razor and background is subtracted a priori. 1st row: PDDFs obtained via evidence method. 2nd row: PDDFs obtained via true error method. 3rd row: PDDFs obtained via 5-folded Cross-Validation (CV) method. 4th row: PDDFs obtained via 10-folded CV method. The table at the bottom shows the mean and standard deviation of the determined $r_{\max}$ values (notation: (mean, std. dev.)).

References

[1] B-splines, https://lp.uni-goettingen.de/get/text/1245.

[2] http://mathworld.wolfram.com/ballpointpicking.html, 2013.

[3] http://www.nmrdb.org/predictor, 2013.

[4] Coefficient of variation, http://en.wikipedia.org/wiki/coefficient_of_variation, 2014.

[5] SLD calculator, http://www.sld-calculator.appspot.com/, 2014.

[6] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, Inc., New York, 1970.

[7] M. P. Allen. *Computer simulation of liquids*. Oxford University Press, 1987.

[8] M. Ambuehl, F. Bangerter, P. L. Luisi, P. Skrabal, and H. J. Watzke. Configurational changes accompanying vesiculation of mixed single-chain amphiphiles. *Langmuir*, 9(1):36–38, Jan. 1993.

[9] A. H. Bahrami, M. Raatz, J. Agudo-Canalejo, R. Michel, E. M. Curtis, C. K. Hall, M. Gradzielski, R. Lipowsky, and T. R. Weikl. Wrapping of nanoparticles by membranes. *Adv. Colloid Interface Sci.*, Mar. 2014.

[10] F. Barthe, O. Guédon, S. Mendelson, and A. Naor. A probabilistic approach to the geometry of the ℓ_p^n-ball. *Ann. Prob.*, 33(2):480–513, Mar. 2005.

[11] D. Basak, S. Pal, and D. Patranabis. Support vector regression. *Neural Information Processing – Letters and Reviews*, 10:203–224, 2007.

[12] T. Baumgart, S. T. Hess, and W. W. Webb. Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature*, 425(6960):821–824, Oct. 2003.

[13] J.-F. Berret. Evidence of overcharging in the complexation between oppositely charged polymers and surfactants. *J. Chem. Phys.*, 123(16):164703, 2005.

[14] P. R. Bevington. *Data reduction and error analysis for the physical sciences*. McGraw-Hill, Boston, 3rd ed edition, 2003.

[15] C. M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, New York, 1995.

[16] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[17] L. Blum and G. Stell. Polydisperse systems. I. Scattering function for polydisperse fluids of hard or permeable spheres. *J. Chem. Phys.*, 71(1):42, 1979.

[18] L. Blum and G. Stell. Erratum: Polydisperse systems. I. Scattering function for polydisperse fluids of hard or permeable spheres. *J. Chem. Phys.*, 72:2212, 1980.

[19] O. Bousquet, S. Boucheron, and G. Lugosi. Introduction to statistical learning theory. In *Advanced Lectures on Machine Learning*. Springer Berlin Heidelberg, 2004.

[20] D. I. Bower. Orientation distribution functions for uniaxially oriented polymers. *J. Polym. Sci., Polym. Phys. Ed.*, 19(1):93–107, Jan. 1981.

[21] R. N. Bracewell. *The Fourier transform and its applications*. McGraw-Hill series in electrical and computer engineering. McGraw Hill, Boston, 3rd edition, 2000.

[22] K. Bressel, J. Gummel, T. Narayanan, and M. Gradzielski. Mechanistic insights into the polymer mediated size control of spontaneously formed surfactant vesicles. *in preparation*.

[23] K. Bressel, M. Muthig, S. Prévost, I. Grillo, and M. Gradzielski. Mesodynamics: watching vesicle formation in situ by small-angle neutron scattering. *Colloid Polym. Sci.*, 288(8):827–840, Apr. 2010.

[24] K. Bressel, M. Muthig, S. Prevost, J. Gummel, T. Narayanan, and M. Gradzielski. Shaping vesicles – controlling size and stability by admixture of amphiphilic copolymer. *ACS Nano*, 6(7):5858–5865, July 2012.

[25] K. Bressel, M. Muthig, S. Prévost, J. Gummel, T. Narayanan, and M. Gradzielski. Shaping vesicles by the presence of amphiphilic copolymer. *ESRF Highlights 2012*, pages 67–68, 2012.

[26] J. Brunner-Popela and O. Glatter. Small-Angle scattering of interacting particles. I. Basic principles of a global evaluation technique. *J. Appl. Crystallogr.*, 30(4):431–442, Aug. 1997.

[27] M. Burkhardt. *Micelles and Interpolyelectrolyte Complexes formed by Polyisobuthyline-block-([meth]acrylic acid) - Synthesis of Polymers and Characterization in Aqueous Solutions*. PhD thesis, Universität Bayreuth, 2007.

[28] J. Q. Candela. *Learning with Uncertainty – Gaussian Processes and Relevance Vector Machines*. Technical University of Denmark, 2004.

[29] G. Cevc. Lipid vesicles and other colloids as drug carriers on the skin. *Adv. Drug Delivery Rev.*, 56(5):675–711, Mar. 2004.

[30] J. Combet, P. Lorchat, and M. Rawiso. Salt-free aqueous solutions of polyelectrolytes: Small angle X-ray and neutron scattering characterization. *Eur. Phys. J. Special Topics*, 213(1):243–265, Dec. 2012.

[31] R. T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14:1, 1946.

[32] R. J. Davies. A new batch-processing data-reduction application for x-ray diffraction data. *J. Appl. Cryst.*, 39(2):267–272, Mar. 2006.

[33] P. J. Davis. *Methods of numerical integration*. Dover Publications, Mineola, N.Y, 2nd edition, 2007.

[34] H. D. Deas. The diffraction of X-rays by a random assemblage of molecules having partial alignment. *Acta Crystallogr.*, 5(4):542–546, July 1952.

[35] P. Debye. Zerstreuung von Röntgenstrahlen. *Ann. Phys. (Berlin)*, 351(6):809–823, 1915.

[36] C. Dewhurst. GRASP software package. Institute Laue-Langevin Grenoble, France, 2013.

[37] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys. (Berlin)*, 17:549–560, 1905.

[38] B. D. Ermi and E. J. Amis. Domain structures in low ionic strength polyelectrolyte solutions. *Macromolecules*, 31(21):7378–7384, Oct. 1998.

[39] J. I. Escalante, M. Gradzielski, H. Hoffmann, and K. Mortensen. Shear-induced transition of originally undisturbed lamellar phase to vesicle phase †. *Langmuir*, 16(23):8653–8663, Nov. 2000.

[40] J. Feder and T. Jøssang. Aggregation kinetics of immunoglobulins. *Phys. Scr.*, T13:33–42, Jan. 1986.

[41] L. Feigin and D. Svergun. *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. Plenum Press, 1987.

[42] F. Feroz, M. P. Hobson, and M. Bridges. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614, Oct. 2009.

[43] S. Fischer, O. Marti, T. Diesner, and B. Rieger. Small-angle x-ray scattering on melt-spun polypropylene fibers: Modeling and data reduction. *Macromolecules*, 43(11):5009–5015, June 2010.

[44] K. Fontell, L. Mandell, P. Ekwall, and O. Smidsrød. Some isotropic mesophases in systems containing amphiphilic compounds. *Acta Chem. Scand.*, 22:3209–3223, 1968.

[45] S. Förster, N. Hermsdorf, C. Böttcher, and P. Lindner. Structure of polyelectrolyte block copolymer micelles. *Macromolecules*, 35(10):4096–4105, May 2002.

[46] D. Frenkel. *Understanding molecular simulation: from algorithms to applications*. Number 1 in Computational science series. Academic Press, San Diego, 2nd edition, 2002.

[47] P. Fromherz. Lipid-vesicle structure: Size control by edge-active agents. *Chem. Phys. Lett.*, 94(3):259–266, Jan. 1983.

[48] O. Glatter. Data evaluation in small-angle scattering: calculation of the radial electron density distribution by means of indirect Fourier transformation. *Acta Physica Austriaca.*, 47:83–102, 1977.

[49] O. Glatter. A new method for the evaluation of small-angle scattering data. *J. Appl. Crystallogr.*, 10(5):415–421, Oct. 1977.

[50] O. Glatter. The interpretation of real-space information from small-angle scattering experiments. *J. Appl. Crystallogr.*, 12(2):166–175, Apr. 1979.

[51] O. Glatter. The inverse scattering problem in small-angle scattering. In P. Lindner and T. Zemb, editors, *Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter*, chapter 4. Elsevier Science B.V., first edition, Nov. 2002.

[52] O. Glatter and O. Kratky. *Small Angle X-ray Scattering*. Academic Press, 1982.

[53] M. Gradzielski, M. Müller, M. Bergmeier, H. Hoffmann, and E. Hoinkis. Structural and macroscopic characterization of a gel phase of densely packed monodisperse, unilamellar vesicles. *J. Phys. Chem. B*, 103(9):1416–1424, Mar. 1999.

[54] P. Gregory. *Bayesian logical data analysis for the physical sciences*. Cambridge University Press Cambridge, 2005.

[55] W. Griffith, R. Triolo, and A. Compere. Analytical structure function of a polydisperse Percus-Yevick fluid with Schulz-(γ-) distributed diameters. *Phys. Rev. A*, 33(3):2197–2200, Mar. 1986.

[56] W. Griffith, R. Triolo, and A. Compere. Analytical scattering function of a polydisperse Percus-Yevick fluid with Schulz-(γ-) distributed diameters. *Phys. Rev. A*, 35(5):2200–2206, Mar. 1987.

[57] I. Grillo, E. I. Kats, and A. R. Muratov. Formation and growth of anionic vesicles followed by small-angle neutron scattering. *Langmuir*, 19(11):4573–4581, May 2003.

[58] A. Guinier and G. Fournet. *Small-Angle Scattering of X-rays*. John Wiley & Sons, Inc., 1955.

[59] M. Gutting. Inverse problems. Technical report, TU Kaiserslautern - FB Mathematik, 2009.

[60] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.

[61] A. P. Hammersley, S. O. Svensson, M. Hanfland, A. N. Fitch, and D. Hausermann. Two-dimensional detector software: From real detector to idealised image or two-theta scan. *High Pressure Res.*, 14(4-6):235–248, 1996.

[62] B. Hammouda. *Probing nanoscale structures – the SANS toolbox*. National Institute of Standards and Technology, Center for Neutron Research, Gaithersburg, MD 20899-6102, 2013.

[63] P. C. Hansen. *Discrete inverse problems: insight and algorithms*. Fundamentals of algorithms. Society for Industrial and Applied Mathematics, Philadelphia, 2010.

[64] S. Hansen. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. *J. Appl. Crystallogr.*, 33(6):1415–1421, Dec. 2000.

[65] T. Hastie, R. J. Tibshirani, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York, NY, 2013.

[66] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97, Apr. 1970.

[67] S. Haykin. *Neural Networks a Comprehensive Foundation*. Prentice Hall, 1999.

[68] H. Henke. *Elektromagnetische Felder: Theorie und Anwendung*. Springer, Berlin, 2011.

[69] P. Heunemann, S. Prévost, I. Grillo, C. M. Marino, J. Meyer, and M. Gradzielski. Formation and structure of slightly anionically charged nanoemulsions obtained by the phase inversion concentration (PIC) method. *Soft Matter*, 7(12):5697, 2011.

[70] H. Hindi. A tutorial on convex optimization II: duality and interior point methods. In *American Control Conference*, page 11 pp. IEEE, 2006.

[71] A. Hoerl. Application of ridge analysis to regression problems. *Chem. Eng. Prog.*, 58:54–59, 1962.

[72] B. Hofmann. *Mathematik inverser Probleme*. Teubner, 1999.

[73] D. J. Iampietro and E. W. Kaler. Phase behavior and microstructure of aqueous mixtures of cetyltrimethylammonium bromide and sodium perfluorohexanoate. *Langmuir*, 15(25):8590–8601, Dec. 1999.

[74] J. Idier. *Bayesian approach to inverse problems*. ISTE ; Wiley, London; Hoboken, NJ, 2008.

[75] J. N. Israelachvili, D. J. Mitchell, and B. W. Ninham. Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers. *J. Chem. Soc., Faraday Trans. 2*, 72:1525–1568, 1976.

[76] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106(4):620–630, May 1957.

[77] E. T. Jaynes. *Probability theory: the logic of science*. Cambridge University Press, 2003.

[78] B. Jerman, M. Breznik, K. Kogej, and S. Paoletti. Osmotic and volume properties of stereoregular poly(methacrylic acids) in aqueous solution: Role of intermolecular association. *J. Phys. Chem. B*, 111(29):8435–8443, July 2007.

[79] H. Jung, S. Lee, E. Kaler, B. Coldren, and J. A. Zasadzinski. Gaussian curvature and the equilibrium among bilayer cylinders, spheres, and discs. *Proc. Natl. Acad. Sci. U.S.A.*, 99(24):15318–15322, Nov. 2002.

[80] A. Jusufi, A. Kohlmeyer, M. Sztucki, T. Narayanan, and M. Ballauff. Self-assembly of charged surfactants: Full comparison of molecular simulations and scattering experiments. *Langmuir*, 28(51):17632–17641, Dec. 2012.

[81] E. W. Kaler, K. L. Herrington, A. K. Murthy, and J. A. N. Zasadzinski. Phase behavior and structures of mixtures of anionic and cationic surfactants. *J. Phys. Chem.*, 96(16):6698–6707, Aug. 1992.

[82] N. Kamenka, M. Chorro, Y. Talmon, and R. Zana. Study of mixed aggregates in aqueous solutions of sodium dodecyl sulfate and dodecyltrimethylammonium bromide. *Colloids Surf.*, 67:213–222, Nov. 1992.

[83] U. Keiderling. The new 'BerSANS-PC' software for reduction and treatment of small angle neutron scattering data. *Appl. Phys. A: Mater. Sci. Process.*, 74(0):s1455–s1457, Dec. 2002.

[84] H. Kerner and W. v. Wahl. *Mathematik für Physiker*. Springer, Berlin; Heidelberg; New York, 2006.

[85] D. J. Kinning and E. L. Thomas. Hard-sphere interactions between spherical domains in diblock copolymers. *Macromolecules*, 17(9):1712–1718, Sept. 1984.

[86] S. R. Kline. Reduction and analysis of SANS and USANS data using IGOR pro. *J. Appl. Crystallogr.*, 39(6):895–900, Nov. 2006.

[87] U. Kockelkorn. Einführung in die Statistik. Script, 2003, TU Berlin.

[88] J. Kohlbrecher and I. Bressler. SASfit software package for fitting small-angle scattering curves, `https://kur.web.psi.ch/sans1/SANSSoft/sasfit.html`, 2013.

[89] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[90] D. Lasic. *Liposomes in gene delivery*. CRC Press, Boca Raton FL, 1997.

[91] O. Le Bihan, P. Bonnafous, L. Marak, T. Bickel, S. Trépout, S. Mornet, F. De Haas, H. Talbot, J.-C. Taveau, and O. Lambert. Cryo-electron tomography of nanoparticle transmigration into liposome. *J. Struct. Biol.*, 168(3):419–425, Dec. 2009.

[92] J. Leng, S. Egelhaaf, and M. Cates. Kinetics of the micelle-to-vesicle transition: Aqueous lecithin-bile salt mixtures. *Biophys. J.*, 85(3):1624–1646, Sept. 2003.

[93] J. Leng, S. U. Egelhaaf, and M. E. Cates. Kinetic pathway of spontaneous vesicle formation. *Europhysics Letters (EPL)*, 59(2):311–317, July 2002.

[94] P. Lindner and T. Zemb. *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*. Elsevier, 2002.

[95] R. Lipowsky. Budding of membranes induced by intramembrane domains. *J. Phys. II*, 2:1825–1840, 1992.

[96] E. Loizou, L. Porcar, P. Schexnailder, G. Schmidt, and P. Butler. Shear-induced nanometer and micrometer structural responses in nanocomposite hydrogels. *Macromolecules*, 43(2):1041–1049, Jan. 2010.

[97] D. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., 1969.

[98] D. J. MacKay. Bayesian interpolation. *Neural Comput.*, 4:415–447, 1991.

[99] W. Maier and A. Saupe. Eine einfache molekular-statistische Theorie der nematischen kristallinflüssigen Phase 2. *Z. Naturforsch., A*, 15(4):287–292, 1960.

[100] G. S. Manning. Counterion condensation theory of attraction between like charges in the absence of multivalent counterions. *Eur. Phys. J. E*, 34(12), Dec. 2011.

[101] H. Matsuoka, D. Schwahn, and N. Ise. Observation of cluster formation in polyelectrolyte solutions by small-angle neutron scattering. 1. A steep upturn of the scattering curves from solutions of sodium poly(styrenesulfonate) at scattering vectors below 0.01 1/Å. *Macromolecules*, 24(14):4227–4228, July 1991.

[102] N. A. Mazer, G. B. Benedek, and M. C. Carey. Quasielastic light-scattering studies of aqueous biliary lipid systems. Mixed micelle formation in bile salt-lecithin solutions. *Biochemistry (Mosc.)*, 19(4):601–615, Feb. 1980.

[103] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087, 1953.

[104] R. Michel. *Liposomes in Contact and Interacting with Silica Nanoparticles: From Decorated Vesicles to Internalized Particles*. PhD thesis, TU Berlin, 2013.

[105] R. Michel and M. Gradzielski. Experimental aspects of colloidal interactions in mixed systems of liposome and inorganic nanoparticle and their applications. *Int. J. Mol. Sci.*, 13(12):11610–11642, Sept. 2012.

[106] R. Michel, E. Kesselman, T. Plostica, D. Danino, and M. Gradzielski. Internalization of silica nanoparticles into fluid liposomes — a new type of hybrid material. *In preparation*, 2014.

[107] R. Michel, T. Plostica, L. Abezgauz, D. Danino, and M. Gradzielski. Control of the stability and structure of liposomes by means of nanoparticles. *Soft Matter*, 9(16):4167, 2013.

[108] G. R. Mitchell. X-ray scattering from non-crystalline and liquid crystalline polymers. In *Comprehensive Polymer Science and Supplements*, pages 687–729. Elsevier, 1989.

[109] D. Miyajima, F. Araoka, H. Takezoe, J. Kim, K. Kato, M. Takata, and T. Aida. Electric-field-responsive handle for large-area orientation of discotic liquid-crystalline molecules in millimeter-thick films. *Angewandte Chemie*, 123(34):8011–8015, 2011.

[110] K. Morigaki. Thermodynamic and kinetic stability. Properties of micelles and vesicles formed by the decanoic acid/decanoate system. *Colloids Surf., A: Physicochemical and Engineering Aspects*, 213(1):37–44, Feb. 2003.

[111] K. P. Murphy. *Machine learning: a probabilistic perspective*. Adaptive computation and machine learning series. MIT Press, Cambridge, MA, 2012.

[112] M. Muthig. SASET: https://www.chemie.tu-berlin.de/gradzielski/, 2014.

[113] M. Muthig, S. Prévost, R. Orglmeister, and M. Gradzielski. *SASET*: a program for series analysis of small-angle scattering data. *J. Appl. Crystallogr.*, 46(4):1187–1195, July 2013.

[114] G. Nägele. The physics of colloidal soft matter, lecture notes. Centre of Excellence for Advanced Materials and Structures, 2004.

[115] G. Nalbantov, J. Bioch, and P. Groenen. Support vector regression basics. *Medium Econometrische Toepassingen (MET)*, 13:16–19, 2005.

[116] R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, Canada, 1993.

[117] F. Nettesheim, J. Zipfel, U. Olsson, F. Renth, P. Lindner, and W. Richtering. Pathway of the shear-induced transition between planar lamellae and multilamellar vesicles as studied by time-resolved scattering techniques. *Langmuir*, 19(9):3603–3618, Apr. 2003.

[118] J. R. Norris. Smoluchowski's coagulation equation: uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent. *Ann. Appl. Probab.*, 9(1):78–109, Feb. 1999.

[119] S.-L. Nyeo and R. R. Ansari. Sparse Bayesian learning for the Laplace transform inversion in dynamic light scattering. *Journal of Computational and Applied Mathematics*, 235(8):2861–2872, Feb. 2011.

[120] A. J. O'Connor, T. A. Hatton, and A. Bose. Dynamics of micelle-vesicle transitions in aqueous anionic/cationic surfactant mixtures. *Langmuir*, 13(26):6931–6940, Dec. 1997.

[121] B. R. Pauw. Everything SAXS: small-angle scattering pattern collection and correction. *J. Phys.: Condens. Matter*, 25(38):383201, Sept. 2013.

[122] Y. Pawitan. *In all likelihood: statistical modelling and inference using likelihood*. Clarendon Press; Oxford University Press, Oxford: New York, 2001.

[123] J. Pedersen. Analysis of small-angle scattering data from polymeric and colloidal systems: Modelling and least-squares fitting. *Adv. Colloid Interface Sci.*, 70:171–201, 1997.

[124] J. Percus and G. Yevick. Analysis of classical statistical mechanics by means of collective coordinates. *Phys. Rev.*, 110(1):1–13, Apr. 1958.

[125] D. V. Pergushov, E. V. Remizova, M. Gradzielski, P. Lindner, J. Feldthusen, A. B. Zezin, A. H. Müller, and V. A. Kabanov. Micelles of polyisobutylene-block-poly(methacrylic acid) diblock copolymers and their water-soluble interpolyelectrolyte complexes formed with quaternized poly(4-vinylpyridine). *Polymer*, 45(2):367–378, Jan. 2004.

[126] K. B. Petersen and M. S. Pedersen. The matrix cookbook. *http://matrixcookbook.com*, 2012.

[127] V. Pipich, I. Vasilief, R. Gadiou, and K. Franke. QtiKWS software package, Jülich Forschungszentrum, Germany, 2007.

[128] M. Prato. Regularization methods for the solution of inverse problems in solar X-ray and imaging spectroscopy. *Archives of Computational Methods in Engineering*, 16(2):109–160, Feb. 2009.

[129] W. H. Press. *Numerical recipes: the art of scientific computing*. Cambridge Univ. Press, Cambridge, 2007.

[130] S. Prévost, L. Wattebled, A. Laschewsky, and M. Gradzielski. Formation of monodisperse charged vesicles in mixtures of cationic gemini surfactants and anionic SDS. *Langmuir*, 27(2):582–591, Jan. 2011.

[131] C. E. Rasmussen. *Gaussian Processes for Machine Learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006.

[132] D. Richard, M. Ferrand, and G. J. Kearley. Lamp, the large array manipulation program. *J. Neutron Research*, 4:33–39, 1996.

[133] R.-J. Roe. *Methods of X-ray and Neutron Scattering in Polymer Science*. Oxford University Press, Inc., 2000.

[134] F. Rogers and D. Young. New, thermodynamically consistent, integral equation for simple fluids. *Phys. Rev. A*, 30(2):999–1007, Aug. 1984.

[135] J. S. Rowlinson. The equation of state of dense systems. *Rep. Prog. Phys.*, 28(1):169–199, July 1965.

[136] S. Schmölzer, D. Gräbner, M. Gradzielski, and T. Narayanan. Millisecond-range time-resolved small-angle X-ray scattering studies of micellar transformations. *Phys. Rev. Lett.*, 88(25), June 2002.

[137] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2002.

[138] V. F. Sears. Neutron scattering lengths and cross sections. *Neutron News*, 3(3):26–37, Jan. 1992.

[139] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, July 1948.

[140] A. Shioi and T. A. Hatton. Model for formation and growth of vesicles in mixed anionic/cationic (SOS/CTAB) surfactant systems. *Langmuir*, 18(20):7341–7348, Oct. 2002.

[141] D. S. Sivia and J. Skilling. *Data analysis: A Bayesian tutorial*. Oxford University Press, Oxford; New York, 2006.

[142] J. Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833–859, Dec. 2006.

[143] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, Aug. 2004.

[144] G. Squires. *Introduction to the Theory of Thermal Neutron Scattering*. Dover Publications, Inc., 1978.

[145] M. Stone. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. R. Stat. Soc., Series B*, pages 44–47, 1977.

[146] P. Strunz, J. Šaroun, U. Keiderling, A. Wiedenmann, and R. Przenioslo. General formula for determination of cross-section from measured SANS intensities. *J. Appl. Crystallogr.*, 33(3):829–833, June 2000.

[147] T. Suzuki, H. Endo, and M. Shibayama. Analysis of surface structure and Hydrogen/Deuterium exchange of colloidal silica suspension by contrast-variation small-angle neutron scattering. *Langmuir*, 24(9):4537–4543, May 2008.

[148] D. I. Svergun. *Structure analysis by small-angle X-ray and neutron scattering*. Plenum Press, New York, 1987.

[149] D. I. Svergun. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, 25(4):495–503, Aug. 1992.

[150] M. Sztucki and T. Narayanan. Development of an ultra-small-angle x-ray scattering instrument for probing the microstructure and the dynamics of soft matter. *J. Appl. Crystallogr.*, 40(1):459–462, 2006.

[151] M. Tanaka and E. Sackmann. Polymer-supported membranes as models of the cell surface. *Nature*, 437(7059):656–663, Sept. 2005.

[152] The MathWorks, Inc. ode113 function help of MATLAB (R2012b), 2012.

[153] The MathWorks, Inc. MATLAB help, 2012b.

[154] E. Thiele. Equation of state for hard spheres. *J. Chem. Phys.*, 39(2):474, 1963.

[155] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc., Series B*, 58:267–288, 1996.

[156] R. Tibshirani. Regression shrinkage and selection via the lasso: a retrospective. *J. R. Stat. Soc., Series B*, 73(3):273–282, June 2011.

[157] R. J. Tibshirani. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7(0):1456–1490, 2013.

[158] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics – Doklady*, 4:1035–1038, 1963.

[159] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. Scripta series in mathematics. Winston; distributed solely by Halsted Press, Washington: New York, 1977.

[160] M. Tipping. The relevance vector machine. *Microsoft Research*, 2000.

[161] M. Tipping. Sparse Bayesian learning and the relevance vector machine. *JMLR*, 1:211–244, 2001.

[162] M. Tipping and A. Faul. Fast marginal likelihood maximisation for sparse Bayesian models. *http://www.miketipping.com/papers.htm*, pages 3–6, 2006.

[163] R. van Haasteren. Bayesian evidence: can we beat MultiNest using traditional methods? *arXiv:0911.2150v1*, 2009.

[164] V. N. Vapnik. *Statistical learning theory*. Adaptive and learning systems for signal processing, communications, and control. Wiley, New York, 1998.

[165] M. von Smoluchowski. Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. *Phys. Z.*, 17:585–599, 1916.

[166] M. von Smoluchowski. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Z. Phys. Chem.*, 92:129–168, 1917.

[167] A. Vrij. Light scattering of a concentrated multicomponent system of hard spheres in the Percus–Yevick approximation. *J. Chem. Phys.*, 69(4):1742, 1978.

[168] A. Vrij. Mixtures of hard spheres in the Percus–Yevick approximation. light scattering at finite angles. *J. Chem. Phys.*, 71(8):3267, 1979.

[169] L. M. Walker and N. J. Wagner. SANS analysis of the molecular order in poly(γ-benzyl l -glutamate)/Deuterated dimethylformamide (PBLG/d-DMF) under shear and during relaxation. *Macromolecules*, 29(6):2298–2301, Jan. 1996.

[170] M. Wertheim. Exact solution of the Percus-Yevick integral equation for hard spheres. *Phys. Rev. Lett.*, 10(8):321–323, Apr. 1963.

[171] A. Wiedenmann, R. Gähler, C. D. Dewhurst, U. Keiderling, S. Prévost, and J. Kohlbrecher. Relaxation mechanisms in magnetic colloids studied by stroboscopic spin-polarized small-angle neutron scattering. *Phys. Rev. B*, 84(21), 2011.

[172] Wikipedia. http://en.wikipedia.org/wiki/speedup, February 2014.

[173] Wikipedia. http://en.wikipedia.org/wiki/theta_solvent, February 2014.

[174] W. Wilke and K. Göttlicher. Small angle scattering of partially oriented polymers. *Colloid Polym. Sci.*, 259(6):596–601, June 1981.

[175] C. Wolf, K. Bressel, M. Drechsler, and M. Gradzielski. Comparison of vesicle formation in zwitanionic and catanionic mixtures of hydrocarbon and fluorocarbon surfactants: Phase behavior and structural progression. *Langmuir*, 25(19):11358–11366, Oct. 2009.

[176] WolframAlpha, computational knowledge engine. www.wolframalpha.com, 2013.

[177] T. Z. Wu, Y. Chu, and Y. Yu. On solving lq-penalized regressions. *Journal of Applied Mathematics and Decision Sciences*, 2007:1–13, 2007.

[178] Y. Xia, I. Goldmints, P. W. Johnson, T. A. Hatton, and A. Bose. Temporal evolution of microstructures in aqueous CTAB/SOS and CTAB/HDBS solutions. *Langmuir*, 18(10):3822–3828, May 2002.

[179] J. Yan, P. D. Adams, R. J. Angel, N. L. Ross, M. Rivers, J. B. Parise, and S. M. Clark. The development of an automated data analysis system for high-pressure powder diffraction data collected using an area detector. *High Press. Res.*, 28(3):293–298, Sept. 2008.

[180] G. Yin. *Clinical trial design: Bayesian and frequentist adaptive methods*. Wiley series in probability and statistics. John Wiley & Sons, Hoboken, N.J, 2012.

[181] L. Zhai, J. Zhang, Q. Shi, W. Chen, and M. Zhao. Transition from micelle to vesicle in aqueous mixtures of anionic/zwitterionic surfactants studied by fluorescence, conductivity, and turbidity methods. *J. Colloid Interface Sci.*, 284(2):698–703, Apr. 2005.

[182] P.-W. Zhu and G. Edward. Orientational distribution of parent–daughter structure of isotactic polypropylene: a study using simultaneous synchrotron WAXS and SAXS. *J. Mater. Sci.*, 43(19):6459–6467, Sept. 2008.