The Sampling Theorem and Small-Angle Scattering*

BY THOMAS GERBER AND PAUL W. SCHMIDT†
Sektion Physik der Wilhelm-Pieck Universität Rostock, DDR-2500 Rostock, Universitätsplatz 3, German Democratic Republic

(Received 14 January 1982; accepted 10 May 1983)

Abstract
All information about the scattering sample obtainable from small-angle scattering data is shown to be contained in the discrete measured points of a scattering curve distorted by arbitrary collimation conditions, provided that the interval $\Delta h$ between adjacent measured points fulfills the condition $\Delta h < \pi/L$, where $L$ is the largest correlation distance in the sample; $h = 4\pi\lambda^{-1}\sin\theta$; $\lambda$ is the scattered wavelength; and $2\theta$ is the scattering angle. A simple technique has been developed for separation of part of the noise. It is shown that if the total time for measuring a scattering curve is held constant, a scattering curve recorded with $\Delta h$ equal to the sampling-point interval $n/L$ gives results like those obtained for a scattering curve recorded with a measured point distance smaller than that sampling-point interval. Therefore, $\Delta h$ should be chosen to be small enough to guarantee that $\Delta h < \pi/L$. Furthermore, a technique has been developed to calculate missing data points from the measured intensities. The condition $\Delta h < \pi/L$ has been found to be important for this calculation.

1. Introduction
Several factors must be considered in the analysis of small-angle scattering data. Firstly, the curve is recorded only at a series of discrete points $h_k = 4\pi\lambda^{-1}\sin\theta_k$ at a finite number of scattering angles $2\theta_k$ for an X-ray wavelength $\lambda$. Secondly, the measured curve is distorted by collimation effects, which result from the fact that the slits employed to define the incident and scattered beams have finite dimensions, so that the measured scattering $S(h_k)$ is an average over an interval of scattering angles, rather than being simply the intensity scattered at a single scattering angle. Finally, random intensity fluctuations, or noise, will be superimposed on the measured scattering. The effect of all these factors can be expressed by the equation (Müller & Damaschun, 1979)

$$S(h_k) = \delta(h_k) G_1(h) [G_2(h) E(h)] + N(h_k),$$

(1)

where $E(h)$ is the intensity distorted only by slit-length effects, i.e.

$$E(h) = \int P(t) I[(h^2 + t^2)^{1/2}] \, dt.$$  

Here $P(t)$ is the slit-length weighting function (Hendricks & Schmidt, 1967); $I(h)$ is the undistorted scattered intensity; $G_2(h)$ is the slit-width weighting function (Hendricks & Schmidt, 1967); the asterisk (*) denotes convolution; $G_1(h)$ is a function which describes termination effects and is defined to equal 1 throughout the entire interval in which data are recorded and which is set equal to zero outside this interval; $N(h_k)$ is the noise at point $h_k$; and

$$\delta(h_k) = \begin{cases} 1 & \text{for } h = h_k, \\ 0 & \text{for } h \neq h_k. \end{cases}$$

The measured point distance $\Delta h$, which is defined to be the interval between two adjacent measured points, is assumed to be constant. Usually small-angle-scattering experiments can be designed to satisfy this condition. The noise function $N(h_k)$ and the measured intensity $S(h_k)$ are defined only at the measured points $h_k$.

Most discussions of small-angle scattering theory consider the ideal perfect-collimation intensity $I(h)$, and most methods for data analysis assume that this corrected intensity is known. Thus (1) must be inverted to give $I(h)$ before many of the usual procedures of data analysis can be carried out.

However, in practice, an exact expression for $I(h)$ can rarely be obtained from (1). Numerical approximations are therefore necessary. Many techniques have been developed for interpolation and smoothing of scattering data. Schmidt (1976) has reviewed a number of collimation-correction methods.

Some results from information theory have opened a new path to the analysis of scattering data (Damaschun, Müller & Pürschel, 1967; Damaschun, Müller, Damaschun, Pürschel, Walter & Kranold, 1974; Müller, Schmidt, Damaschun & Walter, 1980; Müller & Damaschun, 1979). As we explain in §§ 2, 3, and 4,
we have employed several concepts from information theory in an investigation of how the analysis of scattering data is affected by the distance between the measured points and by the noise. Furthermore, we have improved the frequency-filtering techniques developed by Damaschun, Müller & Pürschel (1971).

This paper is the first of a series which uses the sampling theorem and other results from information theory to extract information from small-angle scattering data. The next paper (Lembke & Gerber, 1983) discusses the optimum strategy in the measurements. A third paper (Gerber, Walter & Schmidt, 1983) will show how the new technique can be employed for analysis of data for a collimation system with slits which have finite dimensions.

2. Theory

Given a continuous scattering curve $S(h)$ represented by a discrete set of data points $S(h_k)$ and recorded for arbitrary noise and collimation conditions, we want to extract all obtainable information and make an optimal separation of this information from the noise. For this analysis, we will use the frequency function $F(r)$, which is defined to be the Fourier cosine transform of $S(h)$ (Damaschun et al., 1974). According to this definition

$$ F(r) = \int_0^\infty S(h) \cos(hr) \, dh $$

with the inversion

$$ S(h) = \frac{2}{\pi} \int_0^\infty F(r) \cos(hr) \, dr. $$

Since $F(r)$ is the Fourier cosine transform of $S(h)$, all information present in the scattering data is contained in the frequency function.

In this section, we investigate the properties of the frequency function and develop a way to calculate it. We then show how the frequency function can be useful for extracting information from the data.

Müller et al. (1980) have shown that the correlation function $C(r)$ (Debye & Bueche, 1949) and the frequency function $F(r)$ are related by the equation

$$ F(r) = 2\pi^2 \int_0^\infty C[(r^2 + z^2)^{1/2}] W(z) \, dz, $$

where

$$ W(z) = z \int_0^\pi P(t) J_0(zt) \, dt. $$

and $J_0(z)$ is the Bessel function of zero order of the first kind.

The correlation function $C(r)$ has the property (Debye & Bueche, 1949) that

$$ C(r) = 0 \quad \text{for } r \geq L, $$

where $L$ is the largest distance at which there is a correlation between two scattering centers. In a dilute system, $L$ is the largest distance between two points in a particle.

From (4) and (5),

$$ F(r) = 0 \quad \text{for } r \geq L. $$

As the same value of $L$ is obtained from (6) as from (5), the upper limiting frequency $L$ is not affected by the distortion of the scattering curve resulting from the length of the incident beam. Since the effect of the width of the collimating slits can be expressed as a convolution (Guinier, Fournet, Walker & Yudowitch, 1955; Hendricks & Schmidt, 1967), the width of the slits will not affect the value of $L$ calculated from the frequency function. Relations (4) and (6) thus imply that $L$ can be determined from the measured intensity data before collimation corrections are made. The upper limiting frequency $L$ is a quantity which is often useful in the analysis of the scattering data. The fact that $L$ can be calculated from the measured intensity is an important property of the frequency function.*

Although the integral in (2) cannot be evaluated analytically, the frequency function can be calculated from the measured data points $S(h_k)$ which represent the values of the continuous function $S(h)$ (Gerber, Walter & Kranold, 1982).

Because of (6), (3) can be written

$$ S(h) = \frac{2}{\pi} \int_0^\infty F(r) \cos(hr) \, dr, $$

where $X \geq L$.

The change of variable $r = (X/\pi)y$ gives

$$ S(h) = \frac{2X}{\pi^2} \int_0^\pi f(y) \cos(hXy/\pi) \, dy, $$

where $f(y) = F(Xy/\pi)$. If $f(y)$ is expanded as a series of the orthogonal set of functions

$$ (1/\pi)^{1/2} \cdot (2/\pi)^{1/2} \cos(ky) \quad \text{with } k = 1, 2, 3, .... $$

we obtain the frequency function in the form

$$ f(y) = (1/\pi)^{1/2} a_0 + \sum_{k=1} \frac{(2/\pi)^{1/2} a_k \cos(ky)}. $$

The coefficients $a_0$ and $a_k$ are defined by the integrals

---

*In spite of the name 'frequency function' and the term 'upper limiting frequency', the argument of the frequency function and thus the upper limiting frequency $L$ have the dimensions of length, as can be seen from (2).
\[ a_k = (2/\pi)^{1/2} \int_{0}^{\pi} f(y) \cos(ky) \, dy \]

and

\[ a_0 = (1/\pi)^{1/2} \int_{0}^{\pi} f(y) \, dy. \]  \hfill (9)

Comparison of (7a) and (9) leads to the relations

\[ a_k = \frac{\pi^2 S(h_k)}{(2\pi)^{1/2} X}, \]  \hfill (10)

where \( k = 1, 2, 3, \ldots \);

\[ a_0 = \frac{\pi^2 S(0)}{\pi^{1/2} 2X}; \]

and

\[ h_k = \frac{k\pi}{X}. \]

According to (10), the coefficients \( a_k \) are proportional to the scattered intensity at the discrete points \( h_k \), and the increment of \( h \) is

\[ \Delta h = \pi/X \quad \text{with} \quad X \geq L. \]  \hfill (11)

This relation always satisfies the condition \( \Delta h \leq \pi/L \) imposed by the sampling-point theorem (Damaschun, Müller & Pürschel, 1971, p. 18). The \( h \) increment \( \pi/L \) will be called the sampling-point interval.

With the coefficients (10) and the change of variable from \( y \) to \( r \), (8) becomes

\[ F(r) = \frac{\Delta h S(0)}{2} + \sum_{k=1}^{\pi} \Delta h \cdot S(k\Delta h) \cos(k \Delta h \cdot r). \]  \hfill (12)

With (12), the frequency given by the integral (2) is expressed as a Fourier cosine series, the coefficients of which are determined from the discrete measured points. In information theory, the result that the frequency function can be obtained from the measured data points is known as the discrete Fourier transform (Brillouin, 1956). Note that (12) and the sampling-point theorem (Damaschun, Müller & Pürschel, 1971, p. 18) are valid for all collimation conditions.

Although (12) is an infinite series, in practice only a finite number of terms can be evaluated, because the scattering curve is measured only over a finite interval of scattering angles. The calculation of the frequency function is meaningful only if the series (12) converges. The convergence can be improved by extrapolation of the measured data. An extrapolation of the form \( Kh^{-n} \), where \( K \) is a constant and \( 3 \leq n \leq 4 \), is often useful. From (2), the part \( F_d(r) \) of the frequency function obtained by this extrapolation is

\[ F_d(r) = K \int_{h_e}^{r} \frac{\cos(hr)}{h^n} \, dh, \]  \hfill (13)

where \( h_e = h_{\max} + \Delta h/2 \). The lower limit of integration is determined by the condition that, in (12), the last point, at \( h_{\max} \), gives the contribution from the interval

\[ (h_{\max} - \Delta h/2) \leq h < (h_{\max} + \Delta h/2). \]

With the change of variable \( z = hr \),

\[ F_d(r) = Kr^{-1} \int_{h_e}^{r} \frac{\cos z}{z^n} \, dz. \]  \hfill (14)

As the integral in (14) converges rapidly and is a function only of the lower limit, it can be numerically evaluated without great difficulty.

If \( n \) is an integer, the integral in (14) can be expressed in terms of tabulated functions. For \( n = 3 \), for example,

\[ F_d(r) = \frac{Kr^2}{2} \left[ \frac{\cos w}{w^2} - \frac{\sin w}{w} + ci(w) \right], \]  \hfill (15)

where

\[ ci(w) = - \int_{w}^{\infty} \frac{\cos x}{x} \, dx \]  \hfill (16)

and \( w = hr \).

An extrapolation such as (14) should normally be used, in order to reduce the 'termination effect' which causes the curve calculated from (12) to differ from the 'true' curve.

All Fourier cosine series are periodic. According to (11), the period in (12) is \( 2X \) and thus is inversely proportional to the distance between the measured points. If this interval is chosen so that \( \Delta h < \pi/L \),

\[ X = \frac{\pi}{\Delta h} < L, \]

and, since the conditions \( \Delta h \leq \pi/L \) is not satisfied, so that the coefficients of the series (12) are not proportional to the measured intensities, a part of the information in the measured scattering curve is lost. In this case, reliable information usually cannot be extracted from the measured scattering curve.

Fig. 1 shows an example of how the calculated frequency function depends on the interval between the measured points. For this calculation, a theoretical curve for a two-phase sphere (Walter, Kranold, Damaschun & Müller, 1974) is used. The discrete measured points with the increments \( \Delta h \leq \pi/L \) contain all the information obtainable from the scattering curve, because a frequency function calculated from the discrete points contains all of this information. In the calculations discussed in §§ 3 and 4, which test the effects of the noise and of missing data points, the condition that \( \Delta h < \pi/L \) becomes very important (Damaschun, Müller & Pürschel, 1968).
3. Use of the frequency function for noise filtering

The frequency function calculated from (12) is subject to the effects of all of the errors always present in scattering experiments, such as the statistical error caused by quantum noise and the error resulting from the fact that all scattering curves are recorded only in a finite interval of scattering angles. The statistical error, which is present because the intensity is measured at discrete points, can always be calculated with (12).

From (1), we can conclude that the continuous scattering curve $S(h)$ may be divided into a part $\tilde{I}(h)$ containing the information about the structure of the scattering sample and a noise term $N(h)$. Thus one can write

$$S(h) = \tilde{I}(h) + N(h). \quad (17)$$

It is convenient to write the function $N(h)$ in the form

$$N(h) = f(h) n(h), \quad (18)$$

where $n(h)$ describes the effects of the statistics of the quantum noise, which is assumed to be normalized, while $f(h)$ is determined by the counting modes and therefore by the form of the scattering curve and the total time required for the measurements.

For the noise, the frequency function is given by

$$F_N(r) = \frac{1}{2} \int_{-\infty}^{\infty} N(h) \cos hr \, dh \approx \frac{1}{2} \int_{-\infty}^{\infty} n(h) \cos hr \, dh, \quad (19)$$

where $N(h)$ is defined to be $N(-h)$ when $h < 0$. By the convolution theorem for Fourier transforms, (19) can be written

$$F_N(r) = F_f(r) \ast F_n(r), \quad (20)$$

where $F_f(r)$ and $F_n(r)$ are the Fourier cosine transforms of $f(h)$ and $n(h)$, respectively, and where the asterisk denotes convolution.

The frequency function obtained from (12) for the noise $N(h)$ is

$$F_N(r) = \frac{\Delta h}{2} N(0) + \sum_{k=1}^{k_{\text{max}}} \Delta h N(k\Delta h) \cos(k\Delta h r). \quad (21)$$

Thus $F_N(r)$ is the sum of a series of cosine functions with amplitudes proportional to the noise $N(k\Delta h)$ at the data points and to the interval between the points. In the optimum counting mode, the absolute error is the same at every data point (Lembke & Gerber, 1983).

We now will investigate how the noise is affected by the choice of the distance between measured points. For this study we consider the function $n(h)$. Fig. 2 shows an example of a normalized noise curve and the corresponding frequency function. To study the effect of the distance between points, $\Delta h$ was reduced by a factor of five, while the value of $h_{\text{max}}$ was not changed, so that the number of data points was five times as great as in the previous calculation. The amplitude of every cosine function in (21) thus is reduced by five. The resulting frequency function is shown in Fig. 2. In this example, the values of the frequency function lie in a band with a width which has been decreased to nearly 1/3 of its former value. However, since the noise...
is decreased by a factor of three, rather than five, our calculations suggest that the sum in (21) depends less on the number of terms than on the amplitudes of the individual terms.

When the interval between the measured points is smaller than \( \pi/L \), part of the noise can also be removed from the structure-dependent part of the scattering curve by re-transformation of the frequency function calculated with (12). The integration in this re-transformation extends only from 0 to \( L \), since all of the information about the sample is contained in the interval \( 0 \leq r \leq L \). The interval from \( L \) to \( \pi/\Delta h \), the period of the calculated frequency function \( F(r) \), containing only information about noise, is not transformed into \( h \) space. In Fig. 2, the noise is smoothed by re-transformation of values of the frequency function only for \( 0 \leq r \leq L \), rather than for \( 0 \leq r \leq X \).

The determination of \( L \) is very important for noise filtering. Fig. 3 gives an example. The exact value of \( L \) cannot be determined, and it is possible only to estimate the place where the structure-dependent part of the frequency function vanishes within an error band around \( F(r) = 0 \). To estimate this error band, the frequency function must be calculated throughout an interval wider than that corresponding to the estimated \( L \) value.

The frequency function can be transformed to \( h \) space by direct inversion of (12) over an interval \( 0 \leq r \leq Y \). With (7), this transformation can be expressed as

\[
\hat{I}_{sm}(h) = \frac{\Delta h}{\pi} \frac{\sin(hY)}{h} S(0) + \sum_{k=1}^{\infty} \frac{2\Delta h}{\pi} S(k\Delta h) f(h, k\Delta h, Y),
\]

where \( \hat{I}_{sm}(h) \) is the smoothed intensity, and

\[
f(h, k\Delta h, Y) = \int_0^Y \cos(k\Delta h r) \cos(hr) \, dr.
\]

Thus

\[
f(h, k\Delta h, Y) = \frac{\sin[(h-k\Delta h)Y]}{2(h-k\Delta h)} + \frac{\sin[(h+k\Delta h)Y]}{2(h+k\Delta h)}
\]

for \( k\Delta h \neq h \), and

\[
f(h, k\Delta h, Y) = \frac{Y}{2} + \frac{\sin(2hY)}{4h}
\]

when \( k\Delta h = h \). For all \( k \), \( L \leq Y \leq \pi/\Delta h \).

The function \( \hat{I}_{sm}(h) \) is defined for all \( h \), rather than only at the \( h \) values for which the intensity is measured.

If \( Y \) is chosen so that \( Y = X = \pi/\Delta h \) and thus is one-half the period of the frequency function, we obtain the sampling interpolation (Püschel, 1970),

\[
\hat{f}(h) = S(0) \frac{\sin(hX)}{h} + 2 \sum_{k=1}^{\infty} (-1)^k S(k\Delta h) \frac{\sin(hX)}{hX - (k\pi)^2/hX}.
\]

(23)

The sampling interpolation (23) does not smooth the scattering curve, because the entire frequency function of the noise is transformed into \( h \) space. All values of

![Fig. 3. Frequency function for a theoretical scattering curve for a log-normal distribution of uniform spheres with random noise superimposed. The parameters of the distribution are \( \sigma = 1.3 \) nm and \( \mu = 10 \) nm. The scattering curve recorded with fixed-count timing for the same total measured time would have a statistical error of 5%. Although there are spheres with diameter larger than 44 nm in the diameter distribution, for \( r > 40 \) nm the frequency function vanishes within the error band around \( F(r) = 0 \). Therefore the correct value of \( L \) cannot be obtained from this \( F(r) \) curve.](image)

![Fig. 4. Theoretical discrete measured points of the scattering curve for a two-phase sphere with \( R_1 = 6 \) nm, \( R_2 = 10 \) nm, and \( \rho_1/\rho_2 = -1 \) with noise of 10% superimposed ( ). The entire scattering curve (including noise) calculated with the sampling interpolation (23) is shown by curve 1, and the smoothed scattering curve obtained with the modified sampling interpolation (22) is given by curve 2. Intensity is plotted in arbitrary units.](image)
The scattered intensity is given by $\hat{I}(h)$ and has nearly the same statistical error as the measured data points. If $Y$ is chosen to be nearly equal to $L$, the smoothed scattering curve can be calculated from (22). Fig. 4 shows an example.

The sum (22) is a convolution of the scattering curve and the function $\sin(hY)/hY$, because we have multiplied the frequency function by the step function (Müller & Damaschun, 1979)

$$\rho(r) = \begin{cases} 1 & \text{for } 0 \leq r \leq Y \\ 0 & \text{for } Y < r. \end{cases}$$

Thus, an ideal low-pass filter is used for noise filtering. Of course, this filtering is possible only if the structure-dependent part of the frequency function fulfills the condition

$$F(r) = 0 \quad \text{for } r \geq L$$

and the termination effect can be neglected.

Naturally, the noise reduction depends on the ratio of the measured point distance $\Delta h$ to the sampling point distance $\pi/L$. A scattering curve recorded with $\Delta h$ equal to the sampling point distance cannot be smoothed. With the help of the theory of digital filters, we now will investigate the degree of noise reduction obtained for several values of the ratio of the measured point distance to the sampling point distance. From the results of this calculation, some statements can be made about the choice of the optimum measured point distance.

1. The variance reduction factor $R = \Delta h/\Delta h_m$, where $\Delta h_m$ is the sampling point distance $\pi/L$, is a measure of the smoothing that can be achieved (Müller & Damaschun, 1979). The frequency function of the noise in real space is reduced by the factor $R^{1/2}$ when the measured point distance is decreased by $R$. For example, in the curve shown in Fig. 2, the frequency function of the noise calculated for an increment $\Delta h/5$ is $5^{1/2}$ times smaller than the function obtained for an increment $\Delta h$.

2. The effect of the noise on the information obtained about the structure decreases as the measured point distance becomes smaller.

3. An optimum counting mode can be determined only if the total data-measurement time is constant. The statistical error increases at every measured point if the measured point distance becomes smaller, because more points must be measured in the same time. If we reduce the increment by the factor two, the statistical error at every point increases by a factor $2^{1/2}$, because we are able to register only one-half the number of counts at every point, and the statistical error is proportional to $N^{1/2}$, where $N$ is the number of counts at the measured point.

4. If we know $L$ and record a scattering curve with an increment equal to the sampling point distance $\Delta h_m$, we obtain the same result as if we had recorded the scattering curve with a measured point distance smaller than $\Delta h_m$ and a greater statistical error, which is determined by the condition that the total time of measurement for the scattering curve remains constant.

5. In general, the value of the largest correlation distance is not known before a scattering experiment is started. Therefore it is best to choose a measured point distance small enough to guarantee that the condition $\Delta \leq \pi/L$ is fulfilled. This choice of $\Delta h$ does not increase the error in the scattering curve.

6. The direct transformation (22) plays a significant role in the calculation of missing measured data points for small $h$ values, as is shown in the next section.

We begin our investigation of noise reduction by transforming the frequency function into $h$ space. This transformation is executed by a method analogous to the calculation of the frequency function (12). We assume that the smoothed scattering curve $\hat{I}(h)$ can be expressed as a series of cosine functions in the interval $0 \leq h \leq h_m$, with $h_m > h_{\max}$.

The coefficients are defined by the integral

$$b_j = \frac{(2\pi)^{1/2}}{h_p} \int_0^{h_p} \hat{I}(h) \cos\left(\frac{\pi j h}{h_p}\right) dh. \quad (24)$$

Comparison of (24) and (2) shows that the coefficients $b_j$ are nearly proportional to the discrete points $j\Delta r$ of the frequency function calculated with an increment $\Delta r = \pi/h_p$.

For the smoothed scattering curve, by a calculation analogous to our development of (12), we obtain

$$\hat{I}(h) = \frac{F(0)}{\pi} \Delta r + \sum_{j=1}^{j_{\max}} \frac{2\Delta r}{\pi} F(j\Delta r) \cos(jh\Delta r). \quad (25)$$

The error in the approximation (25) may be neglected, since it is appreciable only at $h$ values greater than $h_{\max}$. This error is due to the termination effect for large $h$ values in the calculation of the frequency function from (12). The value of $j_{\max}$ is determined by the condition $j_{\max} \Delta r \geq L$. In practice, $j_{\max}$ is usually chosen so that $F(j_{\max} \Delta r)$ is nearly zero. It is possible, especially for scattering curves registered with fixed-count timing, that the frequency function calculated from the measured data has no zero near $r = L$. In this case the frequency function of the noise is determined mainly by terms in (12) for which $k$ is large (Gerber & Lembke, 1983), and there is a maximum in the frequency function which must be transformed in $h$ space. Fig. 5 shows a frequency function with a maximum at $j_{\max} \Delta r = L_2$. The corresponding smoothed scattering curve is given in Fig. 6.

There is an oscillation in $h$ space for values larger than those which contribute to the noise term, and there is no smoothing in this range. But, on average, the statistical error is also reduced in this case because a digital filter of this form reduces the sum of the squares...
of the absolute errors. Nevertheless, for large $h$, there is no reduction of the noise, since the low scattered intensity will already have a small absolute error if the data are recorded for fixed-time counting. The error in this interval of $h$ can increase, even though a smaller value is obtained for the sum of the squares of the absolute errors over the entire curve.

If $Y$ is chosen to be a zero of the frequency function, there will be no oscillation in $h$ space. (See Figs. 5 and 6 with $j_{\text{max}} AR = L_1$.) When no zero can be found, a transfer function can be introduced, as in the theory of digital filters (Müller & Damaschun, 1979). The part of the frequency function for arguments greater than $L$ is multiplied by this function and therefore becomes zero. In this way the step in the frequency function is removed. The part of the frequency function containing the information about the structure must not be multiplied by the transfer function, in order that the information about the structure is not lost. This procedure need be employed only if the frequency function has no zero for large $r$ and normally is not necessary.

Damaschun, Müller & Pürschel (1971) employed a series of cosine functions nearly identical to (24) for frequency filtering. The scattering curve was continued periodically, and integrals were evaluated to obtain the coefficients. These coefficients are nearly proportional to the discrete points of the frequency function and therefore must be nearly zero for frequencies $r \geq L$. We have used this idea for numerical calculations of frequency filtering.

These discrete points of the frequency function corresponding to the coefficients have an increment $\Delta r = \pi/h_{\text{max}}$, where $h_{\text{max}}$ is the point where the scattering curve is continued periodically.

Our calculations show that in this case a zero of the frequency function is easy to find, since (Lembke & Gerber, 1983) the period of the dominant oscillation in the error band is proportional to $\pi/h_{\text{max}}$, the increment of the calculated coefficients. In general there will be a maximum in the frequency function. A decrease of the last coefficients reduces this maximum and smooths the corresponding oscillations in $h$ space.

The calculation of a smoothed scattering curve from (25) is less complicated and is faster than with (12). Moreover, the use of the extrapolation (14) improves the noise filtering.

4. Calculation of missing measured points at small $h$ values

In the preceding sections we have shown that all the observable information in a scattering curve is contained in the discrete measured points if the increment fulfills the condition $\Delta h \leq \pi/L$. We also considered the termination effect caused by the missing measured points at large $h$. Missing measured points at small $h$ can also complicate the calculation of structure functions. An extrapolation to small $h$ is thus indispensable. In this section we will show that it is possible to calculate the missing measured points (and also the missing sampling points) from the other points if they are registered with an increment smaller than the sampling-point distance $\pi/L$. 

---

Fig. 5. Frequency function for the scattering curve shown in Fig. 6. The quantities $L_1$ and $L_2$ are the upper frequencies used for calculating the smoothed curves. The curve for large $r$ has been multiplied by a factor of 100 with respect to the inner portion of the curve.

Fig. 6. Theoretical smeared scattering curve (in arbitrary intensity units) for a uniform sphere with $R = 10$ nm (curve 1) with 10% noise superimposed (points). The curve has been smoothed with equation (25). The upper frequencies $L_1$ and $L_2$ shown in the frequency function plot in Fig. 5 have been used for calculating the smoothed intensity curves 2 and 3, respectively. Note that, in the latter curve, a step which produces oscillations in $h$ space has been transformed. The lower part of the figure shows the relative error in the smoothed curve 2. The smoothing reduces the sum of the absolute errors in the whole scattering curve. The relative error is reduced only where the intensity is high.
Equation (22) gives the entire smoothed scattering curve. We now will show that if the points \( \hat{I}(0) \) and \( \hat{I}(kAh) \) for \( k = 1, \ldots, n \) are missing, we can obtain them from the points which have been measured. Since these missing points must be calculated, we will move the unknown values to the left side of every equation. In this way, from (22) we obtain the system of algebraic equations

\[
\hat{I}(0) \left( 1 - \frac{YAh}{\pi} \right) - \sum_{k=1}^{n} \frac{2Ah}{\pi} \hat{I}(kAh) f(0, kAh, Y) \\
= \sum_{k=n+1}^{\infty} \frac{2Ah}{\pi} \hat{I}(kAh) f(0, kAh, Y)
\]

(26)

and

\[
\hat{I}(jAh) - \hat{I}(0) \frac{\sin(jAh Y)}{j\pi} - \sum_{k=1}^{n} \frac{2Ah}{\pi} \hat{I}(kAh) f(kAh, kAh, Y) \\
= \sum_{k=n+1}^{\infty} \frac{2Ah}{\pi} \hat{I}(kAh) f(jAh, kAh, Y)
\]

with \( j = 1, 2, \ldots, n \).

The function \( f(jAh, kAh, Y) \) is the same as in (22). We can write (26) as the matrix

\[
I_k M_{kj} = N_j,
\]

where \( k = 1, \ldots, n; \ j = 1, \ldots, n; \)

\[
M_{kj} = 1 - YAh/\pi \quad j = k = 0
\]

\[
M_{kj} = 1 - \frac{2Ah}{\pi} \left[ \frac{Y}{2} + \frac{1}{4jAh} \sin(2jAh Y) \right] \quad j = k \neq 0
\]

\[
M_{kj} = - \left\{ \frac{\sin[(j-k)Ah Y]}{(j-k)\pi} + \frac{\sin[(j+k)Ah Y]}{(j+k)\pi} \right\} \quad j \neq k
\]

and

\[
N_j = \sum_{k=n+1}^{\infty} \frac{2Ah}{\pi} \hat{I}(kAh) f(jAh, kAh, Y).
\]

This equation system can be solved by calculation of the inverse of the matrix \( M \). The quality of the solution depends on the convergence of the infinite series (27), on the ratio of \( Y \) to \( \pi/L \), and on the noise term.

Of course, a solution exists only when \( Y < \pi/Ah \). Therefore, there are restrictions on the use of this method. Also, measured data are not available for large \( h \), and we can improve the convergence only by an extrapolation. In general, however, the convergence is adequate with most measured curves.

The determination of \( Y \) for use with (26) and (27) can often be difficult. For a good-quality solution, \( Y \) must be chosen to be nearly equal to \( L \), but the limiting frequency \( L \) can be found only by use of the missing points. Nevertheless, the method gives good results with an estimated \( Y \) value, as is illustrated by the curves in Fig. 7.

5. Conclusions

We have shown that all of the information about the sample obtainable from the scattering data is contained in the discrete measured points if the measured point distance fulfils the conditions \( Ah \leq \pi/L \) (Damaschun, M"uller & P"urschel, 1971). The sampling point distance \( \pi/L \) is independent of collimation effects. The frequency function introduced by Damaschun et al. (1974) is a great help in extracting information from the scattering data. Equation (12) expresses this function as a Fourier cosine series in which the coefficients are obtained directly from the data points. Problems with numerical integration are therefore avoided. The series converges if the experimental curve decays by two or three orders of magnitude. In general, an extrapolation such as (14) improves the results.

The smoothed scattering curve can be calculated from (22) or (25). It is important to choose the measured point distance to be smaller than the sampling point distance if the scattering curve is distorted by errors caused by the noise, and there are missing measured points at small \( h \) values.

Since the statistical error is related to the fact that \( S(h) \) is sampled at discrete points, the influence of the error on the structure function can be investigated with (12) – and thus, by the discrete transformation. The structure-dependent part of the scattering curve can be separated from the noise if the measured point distance is smaller than the sampling point distance. The discrete transformation has been shown to be related to the frequency filtering introduced by

![Fig. 7. Theoretical discrete measured points of the scattering curve for a uniform sphere with \( R = 10 \text{nm} \) and with 10% noise superimposed. The curve has been smeared with an infinitely long primary beam. The measured point distance is \( Ah \text{,omp} = 12 \). The measured points of the first fourth of the measured curve are missing and have been calculated with different \( Y \) values (curve 1: \( Y = 21 \text{nm} \); curve 2: \( Y = 25 \text{nm} \); curve 4: \( Y = 30 \text{nm} \)). Curve 3 is the theoretical intensity function. For \( h \) greater than 1 \text{nm}\(^{-1} \), an extrapolation \( k^3 \text{h} \) has been used. Intensities plotted are in arbitrary units.](image)
Damaschun, Müller & Pürschel (1971). This transformation can simplify and improve the numerical calculations in frequency filtering. An ideal low-pass filter can in general be used for the filtering. The relation between the measured point distance, the amplitude of the noise, and the smoothing has been investigated for a scattering curve recorded for a constant total measured time.

When the increment $\Delta h$ is decreased to $1/n$ of its former value, the number of measured points is $n$ times greater. Therefore we can record only $1/n$ times as many counts at every point, and the statistical error thus is $n^{1/2}$ times larger, since the statistical error is proportional to the square root of the number of registered counts.

The extent by which the noise is reduced by frequency filtering depends on the increment $\Delta h$. If this increment is decreased by a factor $1/n$, the noise is reduced by approximately $1/n^{1/2}$. Thus, there is no difference if we measure a scattering curve with a large measured point distance (but with maximum sampling point distance) and a small statistical error or with a smaller measured point distance and greater noise, corresponding to the same total measuring time. Since the largest correlation length and therefore the sampling point distance are not known before a curve is measured, a small measured point distance should be used, at least at first, because the condition $\Delta h < \pi/L$ then will certainly be fulfilled, and the accuracy of the results will not be affected.

Furthermore, a method has been developed for calculating missing measured points and missing sampling points at small $h$ values. Equation (26) permits the calculation of missing points if the measured point distance is smaller than the sampling point distance. The noise in the scattering curve and the termination error at large $h$ can affect the results. An estimate of $L$ is needed to record and analyze the scattering curve and also to calculate missing points, because the measured point distance must be determined from $L$. Examples have been presented which illustrate that these calculations are possible for scattering curves similar to those likely to be encountered in small-angle scattering experiments.

This paper is the first of a series. A second paper is concerned with collimation conditions (Gerber, Walter & Schmidt, 1983). Furthermore, the techniques described in the preceding sections make it possible to investigate the optimum counting modes (Gerber & Lembke, 1983).

We would like to thank R. Kranold and G. Walter (Rostock) and J. J. Müller (Berlin-Buch) for useful advice and suggestions. In addition, we wish to express our appreciation to the US National Academy of Sciences and to the Akademie der Wissenschaften der DDR for making it possible for one of us (PWS) to visit the German Democratic Republic during the course of this investigation.

References


