Effects of normalization errors on size distributions obtained from dynamic light scattering data

Horst Ruf
Max-Planck-Institut für Biophysik, D-6000 Frankfurt 70, Federal Republic of Germany

ABSTRACT This paper presents a study of the influence of normalization errors on size distributions obtained from the analysis of intensity fluctuations by photon correlation spectroscopy. The effects of these errors are demonstrated by means of computer-generated autocorrelation functions simulating light scattered from a monomodal Schulz distribution of small, spherical, unilamellar lipid vesicles. The calculations show that even small errors in the baseline, modifying the data upon normalization systematically, will cause serious errors in the estimated size distribution. As it turns out this is due to the peculiar characteristics of normalization errors in data of the first order autocorrelation function. The errors introduced there are described in parts by functions of the delay time having positive exponents. Such components are not considered in the integral equations commonly used to analyze the measured data. The error's property to be a function of delay time in turn enables us to obtain the relative baseline error from the inversion of the data. The new method for its determination is described in some detail. Here, it has been realized with a modified version of the size distribution algorithm CONTIN.

INTRODUCTION

The methods of intensity fluctuation spectroscopy in laser light scattering experiments are widely applied to determine sizes or size distributions of macromolecules or macromolecular assemblies in solution (for review, see Pusey and Vaughan, 1975). To obtain from data of polydisperse samples the distribution of diffusion coefficients or hydrodynamic radii, one has to solve systems of integral equations of the Laplace transform type (Koppel, 1972). Yet the analytical inversion formula cannot be applied, because the experimental data are always noisy to a certain extent. As has been shown, even for low noise there exist many different solutions, which can have large deviations from each other, all satisfying the data to within experimental error (Phillips, 1962). The extraction of useful information about the actual distribution from the inversion of those equations, however, became possible with developments introducing tools for the selection of meaningful solutions from the existent variety into the evaluation procedure. One of these methods uses regularizors that allow, for instance, restriction of curvature of the solution to give a smoothed or parsimonious distribution (Phillips, 1962; Twomey, 1963; Provencher, 1978, 1979; Wiff, 1973), whereas a different approach uses truncated expansions of the integrals in the eigenfunctions of the Laplace transform operator to give a filtered transform, from which those parts of information of the underlying distribution are omitted that cannot be extracted from the noise (McWhirter and Pike, 1978; Ostrowsky et al., 1981). Imposing nonnegativity constraints to the results of the inversions and choosing the degree of smoothing by a criterion involving the Fisher test greatly enhanced the efficiency of the regularization method (Provencher, 1978, 1979).

Although some of the new methods proved very successful in reducing the variety of possible solutions to a small number of meaningful ones, the reliability of the results ultimately depends on the accuracy of the data, of course. Errors in the data may be due to imperfections of the experimental set-up as misalignments, laser instabilities, unwanted stray light, dust, etc. On the other hand, random errors resulting from both the statistical nature of the photon detection process and of the intensity fluctuations are practically always present. The magnitudes of both statistical errors depend on the number of fluctuations being analyzed (Jakeman et al., 1971; Degiorgio and Lastkova, 1971), but otherwise the characteristics are different (Saleh and Cardoso, 1973). The correlations of the errors in the readings of the various channels due to the statistical nature of the intensity fluctuations may result in artifact components of the distribution (Kojro, Z., manuscript in preparation). There are additional errors which ultimately originate from any of the error sources just mentioned. These errors are created when the data are normalized (Jakeman et al., 1971). The experimental parameters used for that are subject to the same types of errors as the other data. However, normalization by an erroneous parameter will affect the data in a systematic way, and its very special effects on the results

0006-3495/89/07/67/12 $2.00

obtained from the inversion of the data of polydisperse samples will be the subject of this paper.

When intensity fluctuations are analyzed by calculating correlation functions, the value of the baseline is used to normalize the data. Experimentally, this one can be determined in different ways, and accordingly different methods of normalization exist (Koppel, 1974; Oliver, 1978). The errors introduced upon normalization do not alter the decay characteristics of the intensity- or photocount autocorrelation function (Hughes et al., 1973). Moreover, the bias due to statistical errors in the data was determined to be only of order $1/N$ (Jakeman et al., 1971), where N is the number of intensity measurements performed during an experiment. Accordingly, it was assumed that this type of error will not have appreciable effects on the results of the analysis. In spite of that, the importance of the accuracy of the baseline value has been pointed out repeatedly (Oliver, 1981; Chu, 1983). Recently, Weiner and Tscharnuter (1987) analyzed calculated autocorrelation functions simulating the light scattered by a sample of monodisperse particles by the cumulant technique. They demonstrated that the baseline must be correctly established within at least 0.1%, otherwise suprisingly large errors occur in the parameters that describe the width of the distribution. The authors also mentioned that spurious peaks will occur when one of the size distribution algorithms is used to evaluate these data.

To find the cause for that apparent discrepancy between expectations and findings, the characteristics of the normalization errors in the data of the first order autocorrelation functions were investigated here, because it is this function that is actually used to analyze data of polydisperse samples by size distribution algorithms. In the first section of results, some of the basic relations used in photon correlation spectroscopy are reviewed, and the characteristics of the error on normalization in data of the second order autocorrelation function are described. The peculiar characteristics of this error in data of the first order autocorrelation function are presented in the next section. There, also the effects on the results obtained from the inversion of polydisperse data are shown. This is demonstrated using computer-generated correlation functions, whose data points have been normalized by erroneous baseline values of magnitudes typical of common experimental situations. The data were analyzed by means of the regularization method for inverting Fredholm integral equations of the first kind (CONTIN; Provencher, 1982*a–c*). In the last section, finally, a new method is described by which the value of the baseline error, or at least an estimate of it, can be determined from data analysis.

METHODS

Simulations

First order, or field autocorrelation functions, simulating light scattered from unilamellar spherical lipid vesicles, were calculated for a given size distribution according to the integral equation (Koppel, 1972),

$$|g^{(1)}(\tau)| = \int_0^\infty S(\Gamma)e^{-\Gamma\tau}d\Gamma, \tag{1}$$

where $\Gamma = Dq^2$ is the decay constant for particles of diffusion coefficient D, τ is the delay time, and $S(\Gamma)$ is the distribution of particles with respect to Γ, $q = (4\pi n_s/\lambda)\sin(\Theta/2)$ denotes the magnitude of the scattering vector with n_s being the refractive index of the solution, λ the wavelength of the incident light in vacuo, and Θ the scattering angle. For noninteracting spheres in dilute solutions, the diffusion constant is related to the hydrodynamic radius of a vesicle, r, by the Stokes-Einstein equation,

$$D = kT/(6\pi\eta r), \tag{2}$$

where k is Boltzmann's constant, T the absolute temperature, and η the viscosity of the medium. The distribution function $S(\Gamma)$ then can be expressed in terms of radii by (Goll and Stock, 1977; Flamberg and Pecora, 1984)

$$S(\Gamma) = n(r)I(r, \mathbf{q})(dr/d\Gamma), \tag{3}$$

where $n(r)dr$ is the number fraction of vesicles having radii between r and $r + dr$, and $I(r, \mathbf{q})$ is the intensity scattered by a vesicle of that size in direction of the detector specified by the scattering vector $\mathbf{q}$. The latter is proportional to the square of the vesicles' polarizability, and hence to the square of the volume of the spherical shell of thickness d, that is

$$I(r, \mathbf{q}) \sim \{(4\pi/3)[r^3 - (r - d)^3]\}^2 P(r, \mathbf{q}), \tag{4}$$

where $P(r, \mathbf{q})$ is the scattering form factor of spherical shells, for which the relation given by Pecora and Aragon, 1974, was used. The size distribution selected here is a two-parameter unimodal Schulz distribution (Schulz, 1935; Aragon and Pecora, 1976; Bott, 1983),

$$n(r) = (1/z!)[(z + 1)/\bar{r}]^{z+1} r^z \exp[-(z + 1)r/\bar{r}], \tag{5}$$

which has a number average radius of 11 nm and a Schulz z parameter of 169 that corresponds to a standard deviation of ±0.844 nm. This distribution of fractional number density, together with the corresponding density functions of mass $m(r) \sim n(r)(4\pi/3)[r^3 - (r - d)^3]$ and of scattered intensity $n(r)I(r, \mathbf{q})$ are shown in Fig. 1. Size and width selected here are typical of distributions of small vesicles prepared according to the procedure of Brunner et al., 1976. Because these distributions are, in general, quite narrow (Stelzer et al., 1983), the three density functions of number, mass, and scattered intensity per size unit do not differ much, as illustrated in Fig. 1. Correspondingly, the average radii and standard deviations, weighted by numbers, mass (m), or mass squared times the form factor (zp), $\bar{r} = 11.00$ nm, $\bar{r}_m = 11.16$ nm, $\bar{r}_{zp} = 11.32$ nm, and $\bar{s} = \pm 0.844$ nm, $\bar{s}_m = \pm 0.848$ nm, $\bar{s}_{zp} = \pm 0.853$, differ only little. In order to not overcomplicate the situation, the distribution selected here approaches zero smoothly on both sides, and thus does not take into consideration the lower size limit of small vesicles due to packing constraints (Huang and Mason, 1978). Instead of using the simulation facilities of the program CONTIN, the data were generated by a separate program developed mainly for purpose of

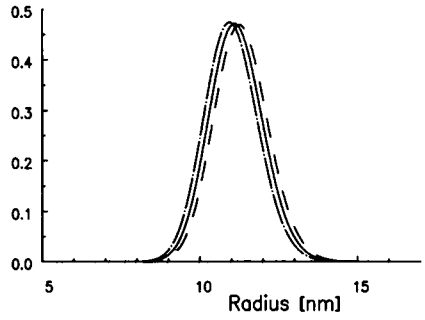

FIGURE 1 Unimodal Schulz distribution of hollow spheres used to calculate computer-generated autocorrelation functions simulating light scattered from unilamellar lipid vesicles. The distribution has a number average radius of 11 nm and Schulz z parameter of 169 that corresponds to a standard deviation of 0.844 nm. For comparison, number density function $n(r)$ (- · -), mass density function m(r) (——), and density function of scattered light intensity $n(r)I(r, \mathbf{q})$ (---) normalized to total number, total mass, and total scattering intensity, respectively, are depicted. The hydrodynamic radius is on the abscissa; density of number, mass, or scattered light intensity is on the ordinate. The intensity function gives here the distribution of light scattered at an angle of 90°.

convenient graphic representation. There, numerical integration of Eq. 1 is performed by means of an automatic adaptive routine based on the 8-panel Newton-Cotes rule (Forsythe et al., 1977). Values of the autocorrelation function were calculated for 136 equidistant delay times (multiples of sampling interval $\delta t = 2\mu s$) covering about two correlation times associated with the mean radius. The physical parameters used to simulate these functions were: absolute temperature $T = 295.6$°K, viscosity $\eta = 0.9436$ cP, wavelength of incidence light in vacuo $\lambda = 632.8$ nm, refractive index of the suspension $n_s = 1.333$, scattering angle $\Theta = 90$°, and thickness of the spherical shell $d = 4.5$ nm.

Normalization errors were introduced into the data of the simulated autocorrelation function as follows. The calculated, exact values of the normalized first order autocorrelation function were squared, and 1 was added to give the corresponding values of the second order function. Then these values were multiplied by the exact baseline value B and divided by the erroneous baseline value $\hat{B}$, which gives the erroneous values of the second order function. In practice, those two operations amount to a simple multiplication by the factor $(1 - \Delta B/\hat{B})$, where $\Delta B/\hat{B} = (\hat{B} - B)/\hat{B}$ is referred to as the relative baseline error. The values of the corresponding first order autocorrelation functions in turn were obtained by subtracting 1 from the erroneous data and extracting the square root.

Data analysis

Inversion of the data was carried out by the program CONTIN (Provencher, 1982*a–c*), which was kindly provided by Dr. S. W. Provencher. If not specified, the default values of the program's control parameters were used. The values of the normalized second order autocorrelation function were taken as input data, which were of six-figure accuracy. A certain level of noise is imposed to the data due to rounding of the seventh figure. Because this type of noise does not conform to the characteristics of noise of real data, unweighted analysis was performed throughout. For the quadrature according to Simpson's rule, a logarithmic grid of 31 grid points was chosen to cover the size range from 7 to 15 nm. The regularizor that minimizes the curvature of the distribution was used throughout. The amplitude values of the resulting solutions were constrained to be positive, and to be zero at two extra grid points outside of each of the two integration limits. The kernel of the Fredholm integral equation was specified to include the volume and the scattering form factor of the spherical shell, and hence the mass density function $m(r)$ and the mass fraction radius $\bar{r}_m$ were obtained from the analysis. For applying the normalization-error function in data analysis, two subroutines of CONTIN were modified. According to these modifications, the coefficients of this function either can be calculated from the input data inside of the program, or can be entered as additional input data (details may be obtained from the author). All computations, i.e., simulations and data analysis, were carried out on a minicomputer HP 1000 (Hewlett Packard, D-6380 Bad Homburg), to which the program CONTIN has been adapted in its double-precision version (Stelzer, 1982).

RESULTS

Analysis of fluctuations of scattered light: characteristics of normalization errors in second order autocorrelation functions

The temporal fluctuations of light scattered by macromolecules in solution are most frequently analyzed by calculating correlation functions. When the direct intensity fluctuation method is applied, one calculates from the signal the second order or intensity autocorrelation function. At the relatively low scattering intensities typical for most macromolecular solutions, it is common to use photon counting techniques to improve the signal-to-noise ratio. There, one counts the number of photons registered by the detector during short sampling intervals of length δt, and calculates, for a discrete set of delay times τ_k, the photocount autocorrelation function,

$$\frac{C(\tau_k)}{\bar{n}^2} = \frac{\langle n(\tau_k)n(0)\rangle}{\bar{n}^2}, \qquad (6)$$

where $n(0)$ and $n(\tau_k)$ denote the number of photons counted during two sampling intervals separated by the delay time τ_k. The mean number of photons per sampling interval, $\bar{n}$, represents the baseline of this function, which is used to normalize the data (for convenience, here the notation $\bar{n}$ instead of $\langle n\rangle$ is used). In the limit of an infinite number of samples, normalized photocount- and normalized intensity autocorrelation function are equal, except for the 'zeroth' channel $\tau_k = 0$ (Jakeman and Pike, 1969), despite the fact that the number of photons sampled in the time interval δt is not simply proportional to the intensity of the incident light, but is related to it by Poisson statistics (Mandel, 1959).

The correlation function that is directly related to the parameters of particle motion is the first order or field autocorrelation function (Eq. 1). For the case of Gaussian light, the values of that function can be obtained from those of the second order autocorrelation function by means of Siegert's relation, which reads

$$\frac{C(\tau_k)}{\bar{n}^2} = [1 + f(A)|g^{(1)}(\tau_k)|^2], \quad (7)$$

when the photocount autocorrelation function is involved (Jakeman and Pike, 1969; Jakeman, 1974). There, $f(A)$ is a spatial coherence factor that depends on the number of coherence areas viewed and on the sample time δt. That relation is always included when diffusion coefficients or sizes are determined from dynamic light scattering data.

In real experiments, the values of the photocount autocorrelation functions are calculated from a finite number of samples, N. Due to the statistical nature of both the fluctuations and the photon detection process, one obtains estimators which usually deviate from their corresponding expectation values. Hard-wired correlators calculate, before formation of the average, the sum of products. After N samples were taken, the content of the kth counter can be written as

$$\hat{C}(\tau_k) = NC(\tau_k) + E_k, \quad (8)$$

where $C(\tau_k)$ represents the N-fold estimator for the expectation value of the photocount autocorrelation function $C(\tau_k)$, and E_k denotes the deviation from this one. E_k comprises both types of statistical errors and also the errors which may arise from other sources. Dividing Eq. 8 by the value of the experimental baseline, $\hat{B} = N\hat{n}^2$, which can be determined from the average number of photons measured per sampling interval, $\hat{n}$, yields the experimental, normalized photocount autocorrelation function. As mentioned before, the value of this baseline is subject to errors in the same way as the other data, and thus usually deviates from its expectation value $B = N\bar{n}^2$. Although the error in baseline may be of random nature, the errors introduced upon normalization of the photocount autocorrelation function are systematic ones because all data points are divided by the same erroneous value. The functional characteristics of normalization errors can easily be visualized. Using the abbreviation $\Delta B = \hat{B} - B$ and the identity $B/\hat{B} = 1 - (\Delta B/\hat{B})$, one obtains from Eqs. 6 and 8 for the values of the experimental normalized photocount autocorrelation function the relation,

$$Y_k = \frac{\hat{C}(\tau_k)}{\hat{B}} = \frac{C(\tau_k)}{\bar{n}^2} - \frac{C(\tau_k)\Delta B}{\bar{n}^2\hat{B}} + \frac{E_k}{\hat{B}}, \quad (9)$$

where the noise term $E_k/\hat{B}$ represents the normalized sum of the unknown statistical and nonstatistical errors. The term $-[C(\tau_k)/\bar{n}^2](\Delta B/\hat{B})$ represents the normalization error, being a function of delay time because it is weighted by the expectation values of the photocount autocorrelation function. Combining the terms of the same dependence on delay time, introducing Eq. 7 and omitting the noise term, Eq. 9 becomes identical with the expression derived by Hughes et al., 1973 (Eq. 7 therein). As it follows from that relation, the error on normalization causes merely the amplitude of the squared first order autocorrelation function to be changed by a factor $(1 - \Delta B/\hat{B})$, and a constant amount $\Delta B/\hat{B}$ to be added to each of its values. Yet it does not alter the second order correlation function's decay characteristics, containing the information about the size distribution. Thus it affects only two parameters, which anyhow are treated as adjustable ones in the various fitting procedures.

Characteristics of normalization errors in first order autocorrelation functions

The corresponding values y_k of the first order autocorrelation function are obtained from Eq. 9 applying Siegert's relation (Eq. 7), which gives

$$y_k = [Y_k - 1]^{1/2} = [(1 - \Delta B/\hat{B})f(A)|g^{(1)}(\tau_k)|^2] - (\Delta B/\hat{B} + E_k/\hat{B})^{1/2} \quad (10)$$

This transformation does not change the autocorrelation function's general decay characteristics, but it alters considerably those of the normalization errors as illustrated in Fig. 2. There, calculated exact and erroneous first-and second order autocorrelation functions, normalized by baseline values containing errors of $\pm 2.5 \times 10^{-3}$, and the corresponding differences are depicted (to demonstrate the effect of normalization errors only, the errors E_k have been set to zero). Unlike with second order correlation functions (Fig. 2, *A* and *B*), the normalization errors in the first order correlation functions increase as the values of those functions decrease. This means that upon normalization by an erroneous baseline value, functional components are introduced to the data of the first order autocorrelation function that rise exponentially with delay time. Because such components are not included in Eq. 1, the evaluation procedures, based on the inversion of this equation, will have great difficulties in finding optimal fits to the data.

That is clearly demonstrated by the results obtained from the analysis of the three autocorrelation functions shown in Fig. 2 *C* by means of the size distribution algorithm CONTIN. This general purpose program for the inversion of data represented by linear algebraic or

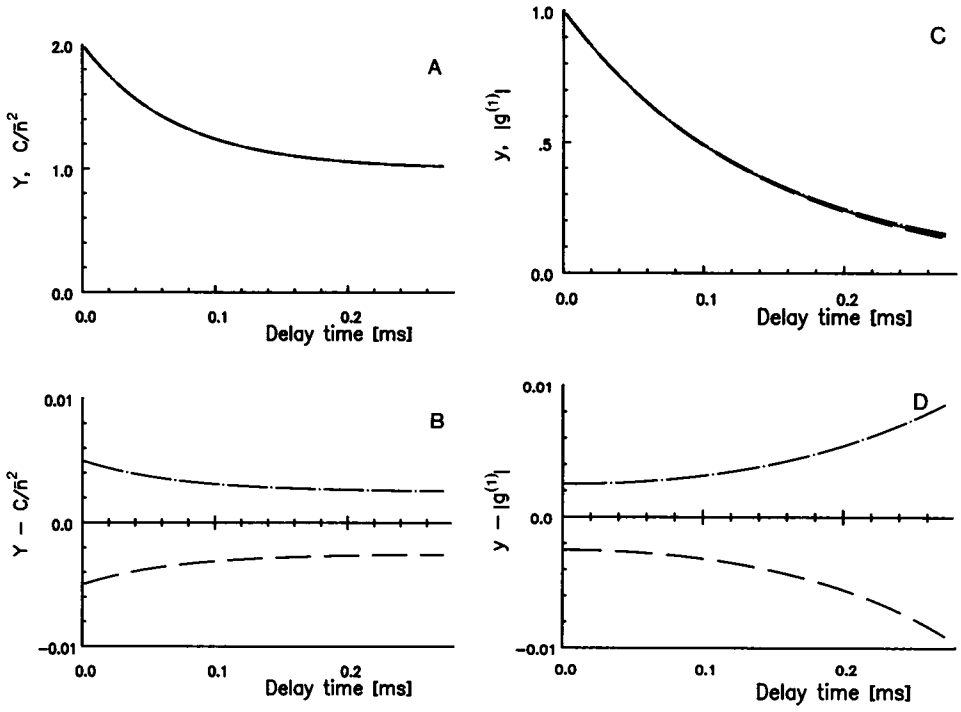

FIGURE 2 (*A*) Computer-generated normalized photocount autocorrelation functions calculated from the Schulz distribution of Fig. 1. The exact function $C(\tau)/\bar{n}^2$ (—), and the erroneous functions $Y(\tau)$, including normalization errors due to relative baseline errors of $+2.5 \times 10^{-3}$ (---) and of -2.5×10^{-3} (- · -), respectively, are plotted against the delay time τ. (*B*) Differences between the erroneous photocount autocorrelation functions and the exact one. (*C*) Corresponding exact and erroneous first order autocorrelation functions, $|g^{(1)}(\tau)|$ and $y(\tau)$. (*D*) Differences between the erroneous first order autocorrelation functions and the exact one.

integral equations here is used to solve equations of the type

$$y_k = \int_0^\infty S(\Gamma)e^{-\Gamma\tau_k}d\Gamma + \epsilon_k + \sum_{i=1}^{N_L} L_i(\tau_k)\beta_i, \qquad (11)$$

where ϵ_k represents the noise. The sum describes N_L functions of known coefficients $L_i(\tau_k)$ that may be included, in addition. The values of the adjustable parameters β_i are determined, together with the distribution $S(\Gamma)$, from the inversion of the data. Setting $N_L = 1$ and all the $L_1(\tau_k) = 1$ allows to count for a constant background, which is a standard option of that program. As will be shown in the next section, the characteristics of the normalization errors can be approximately described by such a linear function, and hence included in the evaluation procedure as well.

The inversions of the examples given here were performed with and without taking a constant background into account. The size distributions depicted in Figs. 3–5 are those selected by the program according to the Fisher test. For comparison, the distribution retrieved from the exact data is shown in Fig. 3, indicating an excellent agreement with the original one (similarly to the results obtained by Bott, 1983). The small deviations observed here are mainly due to the limited number of grid points ($N_g = 31$) used for the quadrature. That can be demonstrated by increasing that number, or by using a slightly narrower integration range. The quality of the solution is expressed in the character of the residuals of the best fit solution to the data of the first order autocorrelation function too (Fig. 3 *B*), which are random and just of order of the data's accuracy ($\pm 5 \times 10^{-7}$). In contrast, the distributions obtained from the erroneous data differ considerably from the original one, independent on whether a constant background is considered or not (Figs. 4 and 5). There, one retrieves either extremely narrow monomodal distributions, or spurious bimodal distributions, whose peaks are positioned above and below the original one. The occurrence of spurious peaks is dependent on the sign of the relative normalization error, as well as on whether a constant background is considered or not. Although the mean sizes are retrieved reasonably well in case of monomodal solutions, the corresponding widths are by far too small (see Table 1). The solutions depicted in Figs. 4 and 5 are also very unstable. The widths of the distributions, and in case of bimodal solu-

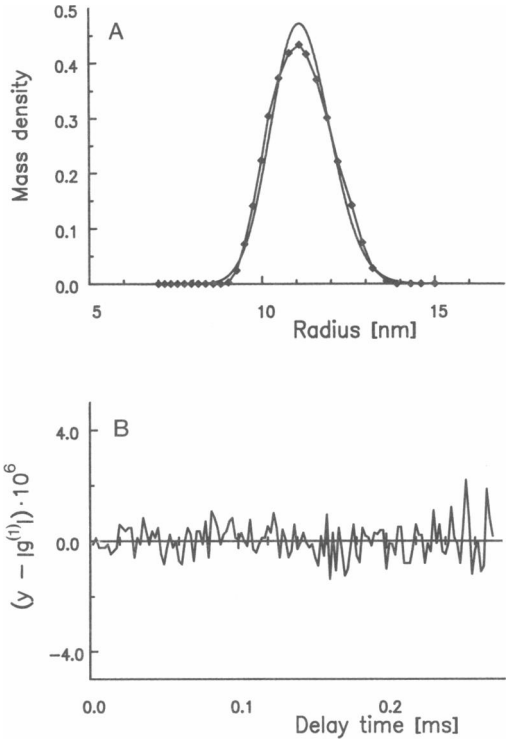

FIGURE 3 (*A*) Mass density function $m(r)$ (♦) obtained from the inversion of the data of the exact first order autocorrelation function shown in Fig. 2 *C* is plotted against the hydrodynamic radius r. This distribution, showing an excellent agreement with the original one (—), was selected by CONTIN as "chosen solution." The analysis was performed without a constant background. Other program parameters are specified in the text. Mass fraction average radius and width of distribution are given in Table 1. (*B*) Residuals of the corresponding best fit first order autocorrelation function plotted against the delay time. They are randomly distributed and in magnitude of order of the data's accuracy.

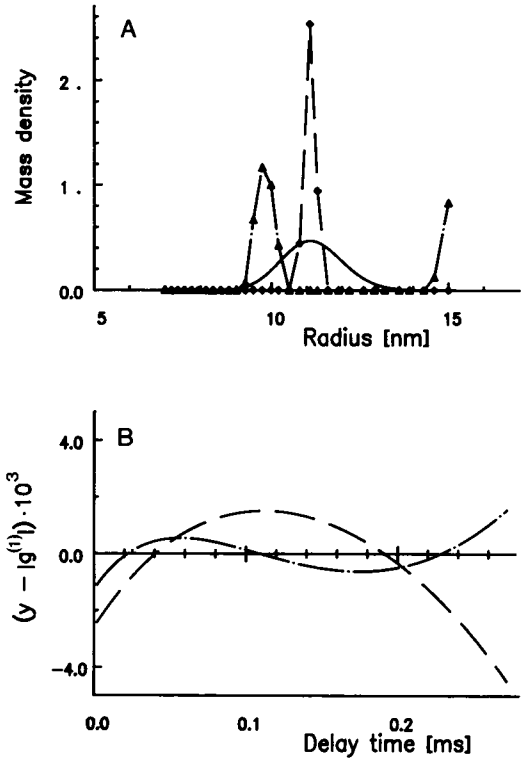

FIGURE 4 (*A*) Mass density function $m(r)$ derived from data of the erroneous first order autocorrelation function shown in Fig. 2 *C*. The inversion was performed as in Fig. 3. Here, the data normalized by a baseline value with positive error ($\Delta B/\hat{B} = +2.5 \times 10^{-3}$) yield a very narrow monomodal distribution (♦, ---), whereas those normalized with negative error ($\Delta B/\hat{B} = -2.5 \times 10^{-3}$) result in a bimodal distribution (▲, —·—). Both agree rather poorly with the original distribution (—) (see also Table 1). (*B*) Residuals of the corresponding best fit first order autocorrelation function plotted against the delay time; they exhibit strong correlations and are of the order of the baseline error.

tions also the positions of the peaks, are strongly dependent on the size range selected for the quadrature. Especially, it is the position of the right-hand peak, sometimes expressed as a shoulder only, which is practically always identical with the upper integration limit. The poor quality of those results is also expressed in the residuals of the corresponding best fit autocorrelation functions shown in Figs. 4 *B* and 5 *B*. These are highly correlated and almost just as large as the relative baseline error itself. The correlations in the residuals reflect the program's incapability to treat exponentially rising parts in the input data with its basic options. It is interesting to note that a slight improvement is achieved when a constant background is considered (Fig. 5). The heights and widths of the peaks are closer to those of the original distribution (see Table 1). That is expected because each value of the normalization error function includes a constant portion (Fig. 2 *D*). Yet the instability of the solutions is not removed by using that additional parameter, and spurious peaks may still occur.

The results of these investigations clearly demonstrate the severe effects of errors on normalization on size distributions obtained from the inversion of dynamic light scattering data, and confirm the findings of other investigations. They reveal that this is a consequence of the square root extraction carried out to transform the data into values of the first order autocorrelation function. To obtain reliable results, one has to determine thus the baseline with rather high accuracy. As is well known, the statistical parts of the baseline error can be reduced by increasing the duration of a measurement, or by averaging the results of repeated measurements on the same sample. An alternative, however, would be to determine the baseline error itself, and to correct the normalized

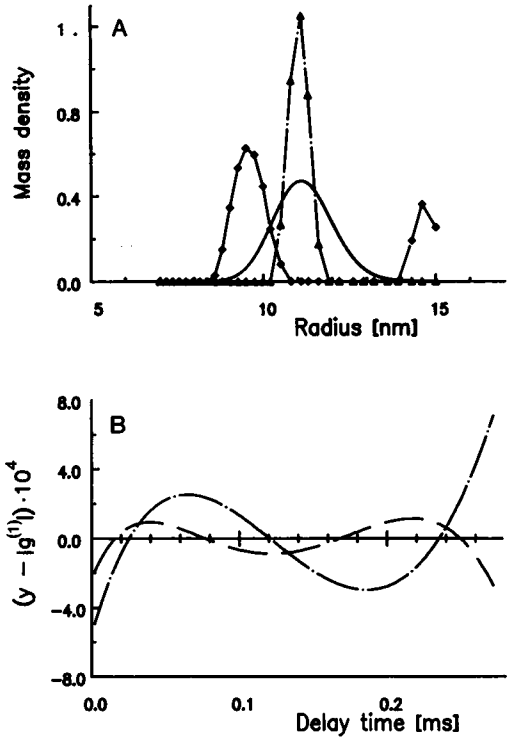

FIGURE 5 (*A*) As in Fig. 4, but taking a constant background into account. A bimodal distribution now appears where a monomodal one appeared in Fig. 4, and vice versa. Although a certain improvement is achieved (see also Table 1), both solutions disagree with the original distribution. (*B*) Residuals of corresponding best fit first order autocorrelation functions plotted against the delay time; they are smaller in magnitude but still very correlated.

TABLE 1 Mass fraction average radii and normalized standard deviations of distribution obtained from inversion of computer-generated autocorrelation functions simulating light scattered from a monomodal Schulz distribution of hollow spheres

		Peak 1		Peak 2		
	$\Delta B/\hat{B}$*	$\bar{r}_m$	$s_m/\bar{r}$	$\bar{r}_m$	$s_m/\bar{r}_m$	$B_c^{\ddagger}$
	$\times 10^{-3}$	*nm*	$\times 10^{-2}$	*nm*	$\times 10^{-2}$	$\times 10^{-2}$
Exact AC§ (NLINF = 0)¶	—	11.16	7.60	—	—	—
Erroneous AC¶ (NLINF = 0)	2.5	11.12	1.77	—	—	—
	−2.5	9.84	2.55	14.86	1.22	—
Erroneous AC** (NLINF = 1)	2.5	9.57	4.54	14.57	1.63	−3.10
	−2.5	11.05	2.55	—	—	1.56
Original‡‡ distribution‡‡	—	11.16	7.60	—	—	—

*Relative baseline error (β_2).
‡Background (β_1) returned from the analysis.
§See Fig. 3.
¶NLINF: program parameter of CONTIN (Provencher, 1982*b* and *c*) controlling whether (1) or not (0) a constant background is taken into account.
¶See Fig. 4.
**See Fig. 5.
‡‡See Fig. 1.

data for it. How this can be achieved will be described in the following section.

Determination of baseline error from data analysis

The method for correcting normalization errors from the data is based on the error's property to be a function of the expectation values of the correlation function, and thereby of the delay time, with the relative baseline error as unknown parameter. The program CONTIN was designed for allowing to take additional functions, with an unknown parameter each, into account (Eq. 11). Yet the normalization errors in the first order autocorrelation function $F_0(\tau_k)$, which are obtained by subtracting the corresponding expectation values $f^{1/2}(A)|g^{(1)}(\tau_k)|$from the data y_k corrected for noise, do not meet the requirements of those functions. For certain conditions, however, their dependence on delay time can be approximated by such a relationship. Since ΔB and the E_k are independent of one another, Eq. 6 can be expanded in a McLaurin series in $\Delta B/\hat{B}$ and $E_k/\hat{B}$ about the point $y_k = f^{1/2}(A)|g^{(1)}(\tau_k)|$ (the bias due to common division by $\hat{B}$ is neglected), which gives

$$y_k \simeq f^{1/2}(A)|g^{(1)}(\tau_k)| - \frac{1}{2}\left[\frac{1}{f^{1/2}(A)|g^{(1)}(\tau_k)|} + f^{1/2}(A)|g^{(1)}(\tau_k)|\right] \cdot \frac{\Delta B}{\hat{B}} + \frac{1}{f^{1/2}(A)|g^{(1)}(\tau_k)|}\frac{E_k}{\hat{B}} + \text{higher order terms.} \quad (12)$$

The second term of the series expansion, which is of the required form, represents the linear contribution of the normalization error, and thus will be referred to as the "linear" normalization error function,

$$F(\tau_k) = L_2(\tau_k)\Delta B/\hat{B}, \quad (13)$$

whose coefficients are given by

$$L_2(\tau_k) = -\frac{1}{2}\left[\frac{1}{f^{1/2}(A)|g^{(1)}(\tau_k)|} + f^{1/2}(A)|g^{(1)}(\tau_k)|\right]. \quad (14)$$

The special feature of the normalization error to increase with delay time can be recognized from that equation. It

is the reciprocal term $1/[f^{1/2}(A)|g^{(1)}(\tau_k)|]$, being present in all higher order contributions of the normalization error too, which is responsible for these characteristics. That term dominates at any delay time. When the correlation function can be described by a single decaying exponential, and $f(A)$ is equal to 1, the coefficients of the normalization error function (Eq. 14) will be described by the hyperbolic cosine function, -cosh ($\Gamma\tau_k$).

The expectation values of the first order autocorrelation function, required to calculate the coefficients of the error function, are not known before the final analysis. Yet one can use estimators instead. These can be obtained, for instance, from a minimum least square solution of the corresponding first order correlation function. In case of data of low noise, however, it is possible to use the experimental data directly. If one rearranges Eq. 6 to express $f^{1/2}(A)|g^{(1)}(\tau_k)|$ as a function of the data y_k, and performs the series expansion about the point y_k, one obtains a corresponding power series, but now with $\Delta B/B$ being the relative baseline error. That is related to the previously defined one by

$$\Delta B/B = \Delta B\hat{B}/(1 - \Delta B/\hat{B}). \tag{15}$$

The other linear approximation to $F_0(\tau_k)$ thus reads

$$\hat{F}(\tau_k) = \hat{L}_2(\tau_k)\Delta B/B, \tag{16}$$

whose coefficients,

$$\hat{L}_2(\tau_k) = -\frac{1}{2}\left[\frac{1}{y_k} + y_k\right], \tag{17}$$

are calculated from the experimental data. In practice, $\Delta B/B$ is mostly nearly equal to $\Delta B/\hat{B}$, and thus either approximation can be used.

The linear approximations F and $\hat{F}$ of the normalization error function F_0 are not only of the required form for being included into the inversion procedure, but also suffice to describe the actual normalization errors well enough in parts of the data. That is illustrated in Fig. 6, where the three different error functions calculated for two levels of baseline error, $\Delta B/\hat{B} = \pm 5 \times 10^{-3}$ and $\pm 2.5 \times 10^{-3}$, are depicted. Both F and $\hat{F}$, approaching the correct error course from either side, agree nicely with F_0 from the very beginning up to certain values, which depend on the magnitude of the baseline error. Above those points the deviations from F_0 become remarkable. That means that there the contributions of the higher order terms of the power series are no longer negligible. These limits are related to defined values of the correlation function and therewith to defined delay times. Thus for a given baseline error, the limit may be expressed in terms of the corresponding decay of the first order autocorrelation function, or in terms of the corresponding delay time. For our examples, the limits are ~0.5 or 0.1 ms in case of $\Delta B/\hat{B} = \pm 5 \times 10^{-3}$ and ~0.35 or 0.15 ms in case of $\Delta B/\hat{B} = \pm 2.5 \times 10^{-3}$. These numbers indicate that the range that is well described by the linear approximations extends as the baseline error decreases. In other words, they express that the higher order terms in Eq. 12 vanish more rapidly than the linear one when $\Delta B/\hat{B}$ approaches zero.

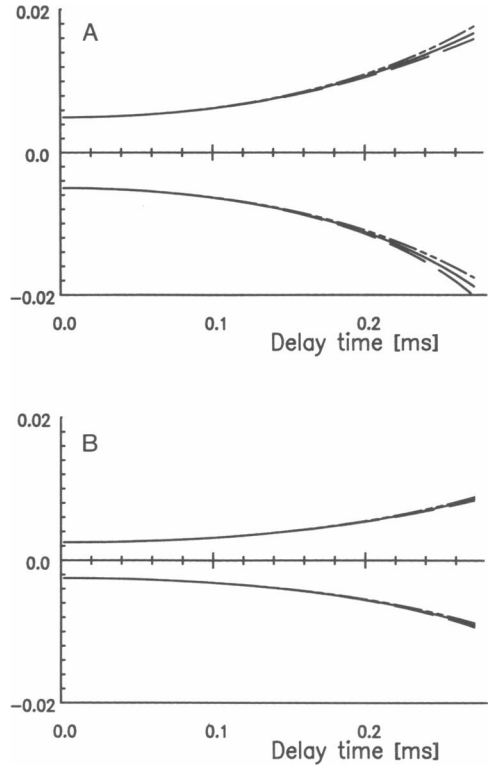

FIGURE 6 Normalization error functions plotted against delay time. The exact function F_0 (—) as well as the two approximations F (---) and $\hat{F}$ (—··—) are depicted for relative baseline errors of $\pm 5 \times 10^{-3}$ (A) and of $\pm 2.5 \times 10^{-3}$ (B). The values of $|g^{(1)}(\tau_k)|$ used to calculate these functions are the same as in Fig. 2. Both F and $\hat{F}$ agree very well with F_0 up to a certain delay time, which depends on the magnitude of the baseline error.

When analyzing the data of the erroneous autocorrelation function ($\Delta B/\hat{B} = +2.5 \times 10^{-3}$) by means of the modified version of the program CONTIN, which takes the "linear" normalization error function $\hat{F}(\tau_k)$ into account, one obtains the results depicted in Fig. 7. The inversions there were performed with successively reduced numbers of data points. When the complete set of 136 points is used, a spurious bimodal solution (Fig. 7 A) is obtained again, as for the case where a constant background was considered (Fig. 5). Yet the mean radius as well as the width of the main peak get closer to those of the original distribution (see Table 2). The corresponding residuals (Fig. 7 D) are smaller in magnitude as well, but

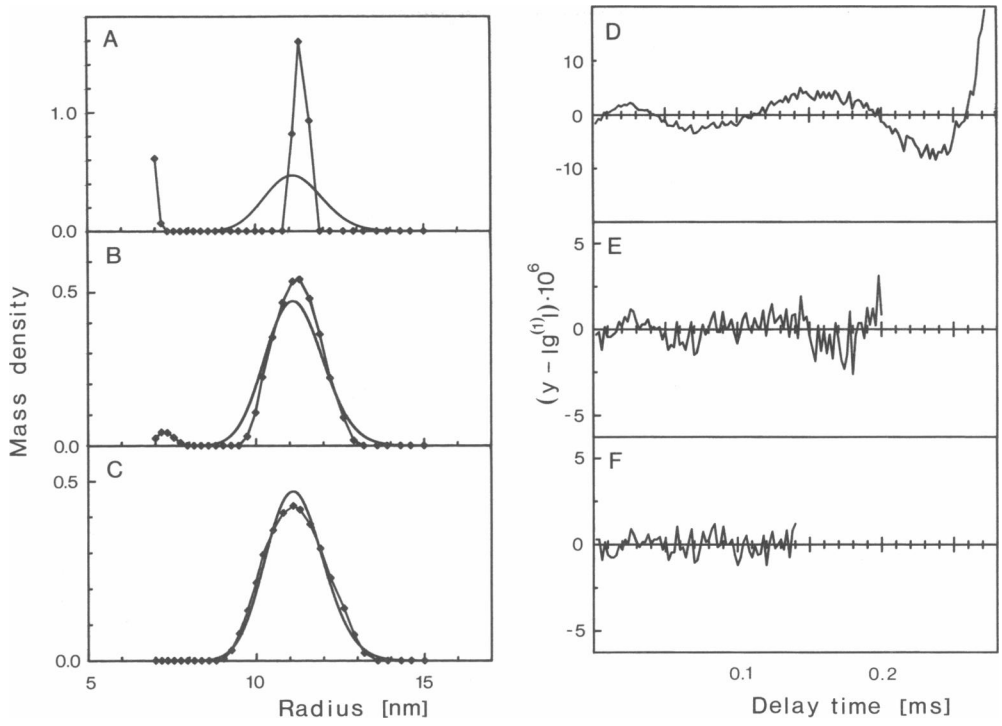

FIGURE 7 Inversions carried out by the modified version of CONTIN, which counts for the "linear" normalization error function $\hat{F}(\tau_k)$. Mass density function $m(r)$ derived from the first order autocorrelation function containing normalization errors due to $\Delta B/\hat{B} = 2.5 \times 10^{-3}$ when 136 (A), 100 (B), or 70 consecutive data points (C) were used (the characteristic parameters of these distributions together with the estimators of $\Delta B/\hat{B}$ are given in Table 2). The successive improvement of results (A–C) when data points are omitted from the range of long delay times is accompanied by a reduction of the corresponding residuals in magnitude as well as in correlation (D–F).

TABLE 2 Mass fraction average radii, normalized standard deviations of distribution, and errors on normalization obtained from inversion of a computer-generated autocorrelation function containing errors on normalization ($\Delta B/\hat{B} = +2.5 \times 10^{-3}$) when taking the normalization error function $\hat{F}(\tau)$ into account*

No. of data‡	Peak 1		Peak 2		
	$\bar{r}_m$	$s_m/\bar{r}_m$	$\bar{r}_m$	$s_m/\bar{r}_m$	$\Delta B/\hat{B}$§
	nm	$\times 10^{-2}$	*nm*	$\times 10^{-2}$	$\times 10^{-3}$
136	11.35	1.51	7.06	1.18	2.304
100	11.25	5.75	7.33	2.64	2.389
70	11.16	7.55	—	—	2.467
Original distribution‖	11.16	7.60	—	—	—

*See Fig. 7.
‡No. of consecutive data points used for the analysis.
§Relative baseline error returned from the analysis.
‖See Fig. 1.

they are still very correlated and especially large at long delay times. Successively better solutions are obtained when data points of that range are omitted, as shown for the case of 100 points (Fig. 7 B). Finally, when the first 70 points are used only, where the maximal deviation of $\hat{F}$ from F_0 at the rightmost point is merely $\sim 10^{-5}$, the original distribution is retrieved. The quality of this fit is expressed in the residuals too (Fig. 7 F), which now are just of order of the data's accuracy and randomly distributed again, as for the case of the exact function (Fig. 3). The successive improvement of the results can also be recognized from the values of the corresponding mean radii, mean widths, and the relative baseline errors obtained from those three evaluations, which are summarized in Table 2.

The results of the inversions demonstrate that the data analysis performed in that way enables us to retrieve the correct solution from data that contain normalization errors. However, one loses thereby parts of the experimental data and hence the information included therein. This disadvantage can be avoided when applying a weighted procedure, as suggested by S. W. Provencher (Max-Planck-Institut für Biophysikalische Chemie, D-

3400 Göttingen, FRG, personal communication). At present, this involves repeating inversions on data that have been corrected for portions of their normalization errors by means of the baseline error obtained from the previous analysis. (How to correct the data is described in the Appendix.) Provided that the estimator lies not too far from the true value of baseline error, the correction will reduce the normalization errors in the data, and consequently will extend the range that is well covered by the linear approximations F or $\hat{F}$. In some instances, a second inversion may already give the optimal solution, together with the remaining relative baseline error. This in fact is the case for our example. The first analysis performed with all 136 data points (Fig. 7 *A*, Table 2) yields a relative baseline error of 2.3×10^{-3}, that is ~80% of the actual error. Taking this value for correcting the data, and carrying out a second analysis brings us the original distribution back, and provides the remaining baseline error with rather high accuracy (Fig. 8).

The effect of normalization errors on size distributions

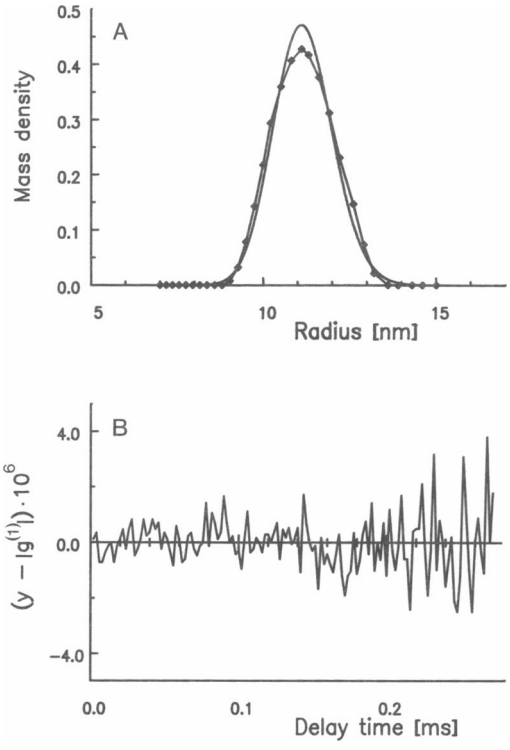

FIGURE 8 Inversion performed on the complete data set of 136 points, which was corrected by the estimator of the relative baseline error obtained from the analysis shown in Fig. 7 *A* being 2.3×10^{-3}. (*A*) Mass density function $m(r)$. (*B*) Residuals, showing that the best fit is almost as good as that obtained from the corresponding exact first order autocorrelation function (Fig. 3). The inversion retrieves the remaining relative baseline error of 2×10^{-4} with an accuracy of 99%.

obtained from the analysis of dynamic light scattering data has been demonstrated up to here with simulated data that do not contain noise of magnitudes typical of real data. Noise will have its own influence on the results of the inversions, of course, but it will also affect the procedure of correcting data for normalization errors. If one calculates from noisy data the coefficients of the normalization error function according to Eq. 17, one obtains a noisy normalization error function. The spread of that noise has the same tendencies as the error function itself. It is large where the magnitude of the actual normalization error is large, and where the agreement between exact and "linear" error functions is poor. This corresponds to the range of small values of the autocorrelation function again, that is, to the range of long delay times. Thus, when using the experimental data directly, one may introduce considerable errors into the values of the coefficients, which, in addition, are asymmetric. The magnitudes of these errors can be such to invalidate the iterative procedure of correcting the data. Therefore, in the case of noisy data one should employ the coefficients of a smoothed normalization error function that averages the random errors to certain extents. The coefficients of such a smoothed function can be calculated from a minimum residual mean square solution of the first order autocorrelation function, which in turn can be obtained from each of the evaluation methods based on Eq. 1. In the Appendix we describe how this can be achieved using the size distribution algorithm CONTIN.

DISCUSSION

The investigations of simulated intensity fluctuations reported in this paper, along with the calculations of Weiner and Tscharnuter (1985), show that even small inaccuracies in the value of the baseline can severely distort the results obtained from the inversion of the data. Accordingly, this error causes either too narrow size distributions to be returned, or distributions with spurious peaks, whose average sizes considerably deviate from those of the actual distribution. The solution obtained from these data are very unstable with regard to variations of the integration limits, and the residuals of the corresponding best fit correlation function are very correlated. These two features can be used in turn to check whether appreciable normalization errors are present in the data.

The effects of errors on normalization were found to be a consequence of the square root extraction included in Siegert's relation. This one transforms the error so as to grow exponentially with delay time, and thus introduces functional components, which are not considered by the

integral equations commonly employed to perform the inversion of the data. Because it is possible to describe this type of error as a function of the delay time, it can be included in the inversion procedure, as described here. The capability of the new method to retrieve reliable size distributions from the inversion of the data has been demonstrated, and a procedure is suggested to deal with noisy data. At present, the technique of removing normalization errors from experimental data, containing a certain amount of noise, is somewhat tedious yet. It requires performing the inversion of the data at least twice. However, for future developments, weighting the data to account for the normalization error as well as the statistical noise could speed up and simplify the procedure.

The effects of normalization errors were demonstrated here on a relatively narrow size distribution, typical of lipid vesicles prepared according to the procedure cited in Methods. Many systems of interest exhibit much broader, and often more complex, size distributions. To determine these, one has to measure the correlation function for a wider range of delay times. Occasionally, that would require using different sampling intervals and splicing the measured sections of the correlation function together, for instance, as in Flamberg and Pecora (1984). In these cases, it is quite likely that the temporal integration affected by the detection process (Jakeman, 1974) must also be considered. Concerning the effects of normalization errors, however, there are no principal differences to the size distribution used here, because the normalization error is primarily a function of the correlation function's expectation values, and only through this a function of the delay time. The method for correcting normalization errors has been realized with the algorithm CONTIN. Here the correction can be included with ease. Whether it can be introduced to other size distribution algorithms has not been checked yet, but in principle this should be possible.

APPENDIX

Correcting data for errors in normalization

The photocount autocorrelation function data can be corrected for normalization errors using the estimators of the relative baseline error obtained from the inversion of data by the modified version of the program CONTIN according to the relation,

$$\tilde{Y}_k = Y_k/(1 - \Delta B/\hat{B}), \qquad \text{(A1)}$$

where $\tilde{Y}_k$ represents the corrected data. If the values $y_k^2 = Y_k - 1$ are used as input data, this relation reads

$$\tilde{y}_k^2 = (y_k^2 + \Delta B/\hat{B})/(1 - \Delta B/\hat{B}). \qquad \text{(A2)}$$

Reduction of normalization errors from noisy data

With noisy data, the coefficients of the normalization error function (Eq. 14) should be calculated from the values of a best fit solution of the first order autocorrelation function. The program CONTIN calculates such solutions of minimum residual square when the value of the regularization parameter approaches zero. This can be done by fixing that parameter to a very small value, for instance by setting the appropriate control parameter to 10^{-9}. The analysis should be a weighted one and should take a constant background into account, because this results in better fits (the size distributions associated with those solutions are not of interest at this stage). The same data are then analyzed again, but now including the normalization error function and allowing the regularization parameter to be optimized by the program. There the coefficients of the normalization error function are entered as additional input data. The value of the relative baseline error obtained from that inversion is then used to correct the data according to one of the above given relations (A1 or A2). By repeating these operations, one minimizes the normalization errors.

The author wishes to thank Drs. E. Grell, Z. Kojro, and S. W. Provencher for stimulating discussions.

Received for publication 4 May 1988 and in final form 16 December 1988.

REFERENCES

Aragon, S. R., and R. Pecora. 1976. Theory of dynamic light scattering from polydisperse systems. *J. Chem. Phys.* 64:2395–2404.

Bott, S. 1983. Polydispersity analysis of QELS data by a smoothed inverse Laplace transform. *In* Measurements of Suspended Particles by Quasi-Elastic Light Scattering. B. E. Dahneke, editor. John Wiley & Sons, New York. 129–157.

Brunner, J., P. Skrabal, and H. Hauser. 1976. Single bilayer vesicles prepared without sonication. Physico-chemical properties. *Biochim. Biophys. Acta.* 455:322–331.

Chu, B. 1983. Correlation function profile analysis in laser light scattering. I. General review on methods of data analysis. *NATO ASI (Adv. Sci. Inst.) Ser. Ser. A Life Sci.* 59:121–160.

Degiorgio, V., and J. B. Lastkova. 1971. Intensity-correlation spectroscopy. *Phys. Rev. A.* 4:2033–2050.

Flamberg, A., and R. Pecora. 1984. Dynamic light scattering study of micelles in a high ionic strength solution. *J. Phys. Chem.* 88:3026–3033.

Forsythe, G. E., M. A. Malcolm, and C. B. Moler. 1977. Subroutine QUANC8. *In* Computer Methods for Mathematical Computations. Prentice Hall Inc., Englewood Cliffs, NJ. 97–105.

Goll, J. H., and G. B. Stock. 1977. Determination by photon correlation spectroscopy of particle size distributions in lipid vesicle suspensions. *Biophys. J.* 19:265–273.

Huang, C., and J. T. Mason. 1978. Geometric packing constraints in egg phosphatidylcholine vesicles. *Proc. Natl. Acad. Sci. USA.* 75:308–310.

Hughes, A. J., E. Jakeman, C. J. Oliver, and E. R. Pike. 1973. Photon-correlation spectroscopy: dependence of linewidth error on normalization, clip level, detector area, sample time and count rate. *J. Phys. A Math. Nucl. Gen. 6:1327–1336.*

Jakeman, E. 1974. Photon correlation. *NATO ASI (Adv. Sci. Inst.) Ser. Ser. B.* 3:75–149.

Jakeman, E., and E. R. Pike. 1969. Spectrum of clipped photon-counting fluctuations of Gaussian light. *J. Phys. A Gen. Phys.* 2:411–412.

Jakeman, E., E. R. Pike, and S. Swain. 1971. Statistical accuracy in the digital autocorrelation of photon counting fluctuations. *J. Phys. A Gen. Phys.* 4:517–534.

Koppel, D. E. 1972. Analysis of macromolecular polydispersity in intensity correlation spectroscopy: the method of cumulants. *J. Chem. Phys.* 57:4814–4820.

Koppel, D. E. 1974. Statistical accuracy in fluorescence correlation spectroscopy. *Phys. Rev. A.* 10:1938–1945.

Mandel, L. 1959. Fluctuations of photon beams: the distribution of the photo-electrons. *Proc. Phys. Soc.* 74:233.

McWhirter, J. G., and E. R. Pike. 1978. On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind. *J. Phys. A Math. Gen.* 11:1729–1745.

Oliver, C. J. 1978. The extraction of spectral parameters in photo-correlation spectroscopy. *Adv. Phys.* 3:387–435.

Oliver, C. J. 1981. Recent developments in photon correlation and spectrum analysis techniques. II. Information from photodetection spectroscopy. *NATO ASI (Adv. Sci. Inst.) Ser. Ser. B.* 73:121–160.

Ostrowsky, N., D. Sornette, P. Parker, and E. R. Pike. 1981. Exponential sampling method for light scattering polydispersity analysis. *Optica Acta.* 28:1059–1070.

Pecora, R., and S. R. Aragon. 1974. Theory of light scattering from hollow spheres *Chem. Phys. Lipids.* 13:1–10.

Phillips, D. L. 1962. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.* 9:84–97.

Provencher, S. W. 1979. Inverse problems in polymer characterization: direct analysis of polydispersity with photon correlation spectroscopy. *Makromol. Chem.* 180:201–209.

Provencher, S. W. 1982*a*. A constrained regularization method for inverting data represented by linear algebraic or intergral equations. *Comp. Phys. Commun.* 27:213–227.

Provencher, S. W. 1982*b*. CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comp. Phys. Comm.* 27:229–242.

Provencher, S. W. 1982*c*. CONTIN Users Manual. EMBL technical report DA05. European Molecular Biology Laboratory, Heidelberg.

Provencher, S. W., J. Hendrix, L. De Maeyer, and N. Paulussen. 1978. Direct determination of molecular weight distributions of polystyrene in cyclohexane with photon correlation spectroscopy. *J. Chem. Phys.* 69:4273–4276.

Pusey, P. N., and J. M. Vaughan. 1975. Light scattering and intensity fluctuation spectroscopy. *In* Dielectric and Related Phenomena. M. Davies, editor. The Chemical Society, London. 48–105.

Saleh, B. E. A., and M. F. Cardoso. 1973. The effect of channel correlation on the accuracy of photon counting digital autocorrelators. *J. Phys. A Math. Nucl. Gen.* 6:1897–1909.

Schulz, G. V. 1939. Über die Kinetik der Kettenpolymerisation. V. *Z. Phys. Chemie.* 43:25.

Stelzer, E. 1982. Untersuchungen zur Präparation und Charakterisierung von Phospholipid-Vesikeln definierter Grösse. Diplomarbeit, Universität Frankfurt, Frankfurt a.M.

Stelzer, E., H. Ruf, and E. Grell. 1983. Analysis and resolution of polydisperse systems. *In* Photon Correlation Techniques. E. O. Schulz-Dubois, editor. Springer-Verlag, Berlin. 329–334.

Twomey, S. 1963. On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear systems produced by quadrature. *J. Assoc. Comput. Mach.* 9:84–97.

Weiner, B. B., and W. W. Tscharnuter. 1987. Uses and abuses of photon correlation spectroscopy in particle sizing. *In* Particle Size Distribution. T. Provder, editor. American Chemical Society, Washington, DC. 48–61.

Wiff, D. R. 1973. Techniques used in applying regularization to the ill-posed problem of determining a molecular weight distribution from sedimentation equilibrium. *J. Poly. Sci. Polym. Symp.* 43:219–234.