

# Revisiting the method of cumulants for the analysis of dynamic light-scattering data

Barbara J. Frisken

The method of cumulants is a standard technique used to analyze dynamic light-scattering data measured for polydisperse samples. These data, from an intensity–intensity autocorrelation function of the scattered light, can be described in terms of a distribution of decay rates. The method of cumulants provides information about the cumulants and the moments of this distribution. However, the method does not permit independent determination of the long-time baseline of the intensity correlation function and can lead to inconsistent results when different numbers of data points are included in the fit. The method is reformulated in terms of the moments about the mean to permit more robust and satisfactory fits. The different versions of the method are compared by analysis of the data for polydisperse-vesicle samples. © 2001 Optical Society of America

OCIS codes: 120.5820, 290.5870.

## 1. Introduction

The method of cumulants<sup>1</sup> for the analysis of dynamic light-scattering (DLS) data was first introduced by Koppel in 1972. He showed that the logarithm of the field–field correlation function is equivalent to the cumulant-generating function. Information about the cumulants of the distribution of decay rates can thus be obtained from the correlation function measured for polydisperse samples.

The method, as commonly used,<sup>2–5</sup> allows the logarithm of the field-correlation function to be written in terms of a polynomial in the delay time  $\tau$ , which is a function that can be fitted easily by use of linear least-squares techniques. This function has several disadvantages. Most remarkably, parameters obtained in the fits are not invariant as more data points are included. In addition, fitting this function requires that the long-time baseline of the intensity correlation function be an assumed rather than a floating parameter. Using the baseline as a floating parameter makes it possible to detect problems in the data and to fit data when a little bit of noise is present.

The ubiquitous use of nonlinear fitting routines makes formulation in terms of a polynomial unnecessary. Reformulating the method in terms of the moments of the distribution rather than of the cumulants results in more satisfactory and robust fits and permits independent fitting of the long-time baseline. Furthermore, it is not necessary to limit the fit to a restricted range of the data. In this paper data from measurements of polydisperse lipid vesicles are used to highlight the differences between the traditional and the reformulated versions of the method. Like the original cumulant method, this reformulated moment method is most reliable for monomodal decay-rate distributions of finite width.

## 2. Theory

### A. Dynamic Light Scattering

DLS measurements involve the analysis of the time autocorrelation function of scattered light as performed by a digital correlator. The normalized time autocorrelation function of the intensity of the scattered light  $g^{(2)}(\tau)$  for a given delay time  $\tau$  is given by<sup>2</sup>

$$g^{(2)}(\tau) = \frac{\langle I(t)I(t + \tau) \rangle}{\langle I(t) \rangle^2}, \quad (1)$$

where  $I(t)$  and  $I(t + \tau)$  are the intensities of the scattered light at times  $t$  and  $t + \tau$ , respectively, and the braces indicate averaging over  $t$ .

In most cases of practical interest the intensity–intensity time autocorrelation function may also be

---

The author (frisken@sfu.ca) is with the Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.

Received 11 December 2000; revised manuscript received 3 April 2001.

0003-6935/01/244087-05\$15.00/0

© 2001 Optical Society of America

expressed in terms of the field-field time autocorrelation function  $g^{(1)}(\tau)$  as

$$g^{(2)}(\tau) = B + \beta[g^{(1)}(\tau)]^2, \quad (2)$$

with  $g^{(1)}(\tau)$  given by

$$g^{(1)}(\tau) = \frac{\langle E(t)E^*(t+\tau) \rangle}{\langle E(t)E^*(t) \rangle}, \quad (3)$$

where  $E(t)$  and  $E(t + \tau)$  are the scattered electric fields at times  $t$  and  $t + \tau$ , respectively, and  $\beta$  is a factor that depends on the experimental geometry. Equation (2) is known as the Siegert relation.<sup>6</sup> The factor  $B$ , commonly referred to as the baseline, is the long-time value of  $g^{(2)}(\tau)$ . Although the factor  $B$  should be equal to one, in practice, a small amount of noise in the measurement can result in values that differ from unity by small ( $\sim 10^{-4}$ ) amounts. In this case assuming that the baseline is one changes the parameter estimates and increases the deviation of the fit from the data. Larger deviations of the baseline from one can indicate that there is a problem with the data.

For monodisperse particles in solution the field-correlation function decays exponentially,  $g^{(1)}(\tau) = \exp(-\Gamma\tau)$ , with a decay rate of  $\Gamma = Dq^2$ , where  $D$  is the diffusion coefficient of the particles and  $q$  is the magnitude of the scattering wave vector. The scattering wave vector  $\mathbf{q}$  is defined as the difference between the incident and the scattered wave vectors, and its magnitude  $q$  is given by

$$q = \frac{4\pi n}{\lambda_0} \sin\left(\frac{\theta}{2}\right), \quad (4)$$

where  $n$  is the refractive index of the solvent,  $\lambda_0$  is the wavelength of the laser in vacuum, and  $\theta$  is the scattering angle. The Stokes-Einstein relation,  $D = k_B T / 6\pi\eta R_h$ , where  $k_B$  is Boltzmann's constant,  $T$  is the temperature, and  $\eta$  is the dynamic viscosity, relates the diffusion coefficient to the hydrodynamic radius  $R_h$  of the particles. For a polydisperse sample,  $g^{(1)}(\tau)$  can no longer be represented as a single exponential and must be represented as a sum or an integral over a distribution of decay rates  $G(\Gamma)$  by

$$g^{(1)}(\tau) = \int_0^\infty G(\Gamma)\exp(-\Gamma\tau)d\Gamma, \quad (5)$$

where  $G(\Gamma)$  is normalized so that

$$\int_0^\infty G(\Gamma)d\Gamma = 1. \quad (6)$$

## B. Method of Cumulants

Finding the precise functional form for the distribution of decay rates  $G(\Gamma)$  is problematic because the correlation function is measured discretely only over an incomplete range of  $\tau$  and there is always noise associated with the data.<sup>3</sup> There are several ways of using DLS data to characterize  $G(\Gamma)$ ,<sup>3</sup> but one of the

simplest is the method of cumulants first proposed by Koppel.<sup>1</sup> This method is based on two relations: one between  $g^{(1)}(\tau)$  and the moment-generating function of the distribution, and one between the logarithm of  $g^{(1)}(\tau)$  and the cumulant-generating function of the distribution. It is appropriate only for use in cases in which  $G(\Gamma)$  is monomodal.

In fact, as was discussed by Koppel,<sup>1</sup> the form of  $g^{(1)}(\tau)$  as given in Eq. (5) is equivalent to the definition of the moment-generating function  $M(-\tau, \Gamma)$  of the distribution  $G(\Gamma)$  (Ref. 7):

$$M(-\tau, \Gamma) = \int_0^\infty G(\Gamma)\exp(-\Gamma\tau)d\Gamma \equiv g^{(1)}(\tau). \quad (7)$$

The  $m$ th moment of the distribution  $m_m(\Gamma)$  is given by the  $m$ th derivative of  $M(-\tau, \Gamma)$  with respect to  $\tau$ :

$$\begin{aligned} m_m(\Gamma) &= \left. \frac{d^m M(-\tau, \Gamma)}{d(-\tau)^m} \right|_{-\tau=0} \\ &= \int_0^\infty G(\Gamma)\Gamma^m \exp(-\Gamma\tau)d\Gamma|_{-\tau=0}. \end{aligned} \quad (8)$$

Similarly, the logarithm of the field-correlation function is equivalent to the definition of the cumulant-generating function<sup>7</sup>  $K(-\tau, \Gamma)$

$$K(-\tau, \Gamma) = \ln[M(-\tau, \Gamma)] \equiv \ln[g^{(1)}(\tau)], \quad (9)$$

where the  $m$ th cumulant of the distribution  $\kappa_m(\Gamma)$  is given by the  $m$ th derivative of  $K(-\tau, \Gamma)$ :

$$\kappa_m(\Gamma) = \left. \frac{d^m K(-\tau, \Gamma)}{d(-\tau)^m} \right|_{-\tau=0}. \quad (10)$$

By making use of the fact that the cumulants, except for the first, are invariant under a change of origin, one can write the cumulants in terms of the moments about the mean as

$$\kappa_1(\Gamma) = \int_0^\infty G(\Gamma)\Gamma d\Gamma \equiv \bar{\Gamma}, \quad (11)$$

$$\kappa_2(\Gamma) = \mu_2, \quad (12)$$

$$\kappa_3(\Gamma) = \mu_3, \quad (13)$$

$$\kappa_4(\Gamma) = \mu_4 - 3(\mu_2)^2 \dots, \quad (14)$$

where  $\mu_m$  are the moments about the mean, as defined by

$$\mu_m = \int_0^\infty G(\Gamma)(\Gamma - \bar{\Gamma})^m d\Gamma. \quad (15)$$

The first cumulant describes the average decay rate of the distribution. The second and the third cumulants correspond directly to the appropriate moments about the mean: The second moment corresponds to the variance, and the third moment provides a measure of the skewness or asymmetry of the distribu-

tion. The first two cumulants must be positive, but the third cumulant can be positive or negative.

The basis of the cumulant expansion that is usually used in the analysis of DLS data lies in expanding the logarithm of  $g^{(1)}$  in terms of the cumulants of the distribution. This relation follows from the fact that the  $m$ th cumulant is the coefficient of  $(-\tau)^m/m!$  in the Taylor expansion of  $K(-\tau, \Gamma)$  about  $\tau = 0$ , as given by

$$\begin{aligned} \ln[g^{(1)}(\tau)] &\equiv K(-\tau, \Gamma) \\ &= -\bar{\Gamma}\tau + \frac{\kappa_2}{2!}\tau^2 - \frac{\kappa_3}{3!}\tau^3 + \frac{\kappa_4}{4!}\tau^4 \dots \end{aligned} \quad (16)$$

To take advantage of this form and use linear least-squares methods to fit this function to the data requires that a key assumption be made about the data: The baseline must be assumed to be exactly one. Then a fit can be made to

$$\ln[g^{(2)}(\tau) - 1] = \ln \frac{\beta}{2} - \bar{\Gamma}\tau + \frac{\kappa_2\tau^2}{2!} - \frac{\kappa_3\tau^3}{3!} + \dots \quad (17)$$

Equation (17) is the traditional fitting function that is described in many DLS texts.<sup>3-5</sup>

Although most modern correlators do an excellent job of measuring the baseline, small amounts of noise can lead to small deviations from unity. Nonlinear fitting routines permit the possibility of fitting the data to  $g^{(2)}$  directly. From Eq. (17), we obtain

$$g^{(2)} = B + \beta \exp\left(-2\bar{\Gamma}\tau + \kappa_2\tau^2 - \frac{\kappa_3}{3}\tau^3 \dots\right). \quad (18)$$

Using the form of Eq. (18) makes it possible to determine  $B$  from the data.

One can also express the field-correlation function in terms of the moments about the mean by first rewriting  $\exp(-\Gamma\tau)$  in terms of its mean value:

$$\exp(-\Gamma\tau) = \exp(-\bar{\Gamma}\tau)\exp[-(\Gamma - \bar{\Gamma})\tau]. \quad (19)$$

By substituting Eq. (19) into Eq. (5), we obtain

$$g^{(1)}(\tau) = \exp(-\bar{\Gamma}\tau) \int_0^\infty G(\Gamma)\exp[-(\Gamma - \bar{\Gamma})\tau]d\Gamma. \quad (20)$$

The exponential function inside the integral in Eq. (20) can be expressed as a series expansion:

$$\begin{aligned} g^{(1)}(\tau) &= \exp(-\bar{\Gamma}\tau) \int_0^\infty G(\Gamma) \left[ 1 - (\Gamma - \bar{\Gamma})\tau \right. \\ &\quad \left. + \frac{(\Gamma - \bar{\Gamma})^2}{2!}\tau^2 - \frac{(\Gamma - \bar{\Gamma})^3}{3!}\tau^3 + \dots \right] d\Gamma. \end{aligned} \quad (21)$$

This expansion is exact as long as all terms in the expansion are included. After integration the corre-

lation function can be expressed in terms of the moments about the mean, as defined in Eq. (15):

$$g^{(1)}(\tau) = \exp(-\bar{\Gamma}\tau) \left( 1 + \frac{\mu_2}{2!}\tau^2 - \frac{\mu_3}{3!}\tau^3 + \dots \right). \quad (22)$$

Expression (22) was derived by Pusey *et al.*,<sup>8</sup> but they went on to expand the logarithm of Eq. (22) to obtain a function for  $\ln[g^{(1)}(\tau)]$ . This expansion adds an extra approximation to the derivation that is unnecessary. Instead, the moment-based expression for  $g^{(1)}$  [Eq. (22)] and the Seigert relation [Eq. (2)] can be used directly to derive a third expression for  $g^{(2)}$ :

$$g^{(2)} = B + \beta \exp(-2\bar{\Gamma}\tau) \left( 1 + \frac{\mu_2}{2!}\tau^2 - \frac{\mu_3}{3!}\tau^3 \dots \right)^2. \quad (23)$$

The form of Eq. (23) also permits the direct fitting of  $B$  and has the advantage that it eliminates stability problems that are inherent to Eq. (18) at large  $\tau$ .

### 3. Testing the Model Functions

The three model functions of Eqs. (17), (18), and (23) with terms up to the second moment about the mean  $\mu_2$  were fitted to data measured from palmitoyl-oleoyl phosphatidylcholine (POPC) vesicles formed by extrusion through polycarbonate membranes with 200-nm pores at an extrusion pressure of 35 psi ( $2.4 \times 10^5$  Pa).<sup>9</sup> The apparatus used for the light-scattering experiments was a Model ALV DLS/SLS-5000 (ALV-Laser GmbH, Langen, Germany) that used a He-Ne laser as the light source. Figure 1 shows results of a measurement in which light scattered by the sample was collected at  $90^\circ$  from the transmitted beam. The model functions were fitted to the data by use of nonlinear fitting routines; weights were calculated from standard deviations provided by the ALV-Laser software. Fits were

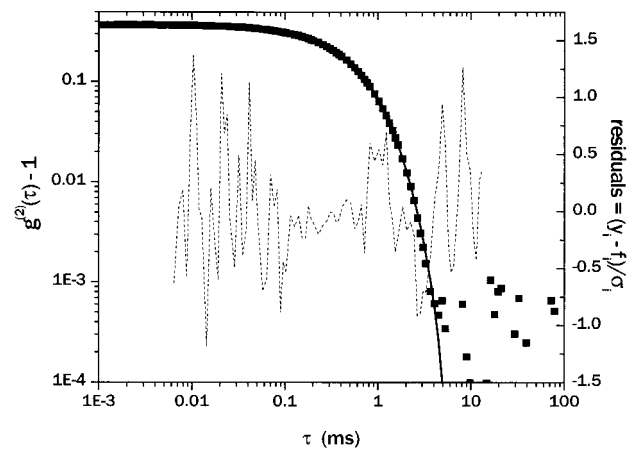


Fig. 1. Sample data taken for POPC vesicles formed by extrusion through polycarbonate membranes. The curve through the data is a fit of Eq. (23) to the data. The dashed curve shows the weighted residuals: the difference of the fit from the data divided by the uncertainty in each point.

**Table 1. Fit of Eq. (17) to DLS Data for POPC Vesicles Extruded through 200-nm Pores<sup>a</sup>**

$\tau_{\max}$ (ms)	$\beta$	$\chi^2$	$\Gamma$ (1/ms)	$\mu_2$ (1/ms <sup>2</sup> )
0.20	2.64	0.3753 <sub>3</sub>	0.92 <sub>1</sub>	0.2 <sub>1</sub>
0.41	2.22	0.3750 <sub>2</sub>	0.913 <sub>4</sub>	0.11 <sub>2</sub>
0.82	2.29	0.3750 <sub>1</sub>	0.908 <sub>2</sub>	0.085 <sub>5</sub>
1.64	12.6	0.3740 <sub>3</sub>	0.900 <sub>2</sub>	0.070 <sub>2</sub>
3.27	620	0.370 <sub>2</sub>	0.860 <sub>5</sub>	0.023 <sub>3</sub>

<sup>a</sup>The subscripts refer to the error in the final digit of the parameter. The parameter  $B$  is assumed equal to 1. The fit was made to data ranging from 6.4  $\mu$ s to a maximum delay time  $\tau_{\max}$ .

made to data corresponding to delay times of 6.4  $\mu$ s to maximum times ranging from 0.31 to 200 ms. It is traditional to fit the cumulant function to data to the point at which the amplitude has fallen to 10% of the original amplitude, which would be 1.84 ms for these data. The fit shown in Fig. 1 is a fit of Eq. (23) to data from  $\tau = 6.4 \mu$ s–13.1 ms. The correlation time of  $\bar{\tau} = 1/\bar{\Gamma}$  for these data is approximately 1.23 ms.

The results for the traditional fitting function, as given by Eq. (17), are shown in Table 1. Results for the three parameters  $\beta$ ,  $\Gamma$ , and  $\kappa_2 = \mu_2$  are not stable in the sense that the parameters vary as the number of data points included in the fit is changed. The standard error in the parameters is shown as a subscript in the appropriate decimal place. The table also includes results for the usual goodness-of-fit parameter  $\chi^2$ , defined by

$$\chi^2 = \frac{1}{N - m} \sum_{i=1}^N \frac{(y_i - f_i)^2}{\sigma_i^2}, \quad (24)$$

where  $N$  is the number of data points,  $m$  is the number of parameters, and  $y_i$ ,  $f_i$ , and  $\sigma_i$  are the data, the fit, and the uncertainty in the data, respectively, at a given delay time  $\tau_i$ .

The results for the second model function, as given in Eq. (18), are shown in Table 2. This function is difficult to fit at small delay times because  $B$  is not specified, and it is difficult to fit at large delay times because the positive term in the exponential that is increasing as  $\tau^2$  makes the function unstable for large  $\tau$ . In the region in which a fit is obtained the parameters vary as more data points are fitted so that

**Table 2. Fit of Eq. (18) to DLS Data for POPC Vesicles Extruded through 200-nm Pores<sup>a</sup>**

$\tau_{\max}$ (ms)	$\chi^2$	$B$	$\beta$	$\Gamma$ (1/ms)	$\mu_2$ (1/ms <sup>2</sup> )
1.64	0.227	0.996 <sub>2</sub>	0.379 <sub>2</sub>	0.903 <sub>3</sub>	0.12 <sub>2</sub>
3.28	0.213	0.9984 <sub>3</sub>	0.3770 <sub>4</sub>	0.906 <sub>2</sub>	0.101 <sub>5</sub>
6.55	0.284	0.9996 <sub>1</sub>	0.3754 <sub>2</sub>	0.905 <sub>1</sub>	0.082 <sub>4</sub>
13.11	0.308	0.9998 <sub>1</sub>	0.3751 <sub>1</sub>	0.900 <sub>1</sub>	0.076 <sub>3</sub>
26.2	0.486	1.0001 <sub>1</sub>	0.3745 <sub>2</sub>	0.900 <sub>1</sub>	0.06 <sub>3</sub>
52.8	Unstable	Unstable	Unstable	Unstable	Unstable

<sup>a</sup>The subscripts refer to the error in the final digit of the parameter. The fit was made to data ranging from 6.4  $\mu$ s to a maximum delay time  $\tau_{\max}$ .

**Table 3. Fit of Eq. (23) to DLS Data for POPC Vesicles Extruded through 200-nm Pores<sup>a</sup>**

$\tau_{\max}$ (ms)	$\chi^2$	$B$	$\beta$	$\Gamma$ (1/ms)	$\mu_2$ (1/ms <sup>2</sup> )
1.64	0.227	0.998 <sub>1</sub>	0.378 <sub>1</sub>	0.907 <sub>2</sub>	0.12 <sub>1</sub>
3.28	0.211	0.9991 <sub>2</sub>	0.3760 <sub>3</sub>	0.908 <sub>1</sub>	0.099 <sub>5</sub>
6.55	0.253	0.9998 <sub>1</sub>	0.3753 <sub>2</sub>	0.907 <sub>1</sub>	0.087 <sub>4</sub>
13.11	0.265	0.9999 <sub>1</sub>	0.3751 <sub>2</sub>	0.907 <sub>1</sub>	0.083 <sub>4</sub>
26.2	0.379	1.00007 <sub>7</sub>	0.3750 <sub>2</sub>	0.907 <sub>2</sub>	0.080 <sub>4</sub>
52.8	0.421	1.00004 <sub>6</sub>	0.3750 <sub>2</sub>	0.907 <sub>2</sub>	0.081 <sub>4</sub>
105.3	0.525	1.00001 <sub>5</sub>	0.3750 <sub>2</sub>	0.907 <sub>2</sub>	0.082 <sub>5</sub>

<sup>a</sup>The subscripts refer to the error in the final digit of the parameter. The fit was made to data ranging from 6.4  $\mu$ s to a maximum delay time  $\tau_{\max}$ .

no satisfactory determination of the polydispersity, in particular, can be made.

The results for the third model function, as given in Eq. (23), are shown in Table 3. Again, the function is hard to fit at small delay times, but as soon as the maximum delay time is greater than a time corresponding to several correlation times the parameters are well determined with minimal variation as the number of data points fitted increases. As well as doing a better job of determining the parameters, the third model function is also much more robust; bad guesses of the initial parameters still lead to quick convergence to the solution. Table 4 compares the number of iterations required for convergence from different starting parameters by use of Eqs. (18) and (23).

The difference in effectiveness among the fitting functions is due to the different expansions used. One derives the expansion in terms of cumulants, Eq. (16), by making an expansion about  $\tau = 0$ . Thus Eq. (16) is accurate only near  $\tau = 0$ , not very useful if one wishes to fit the whole data set to obtain as much information from it as possible. In contrast, one derives the expansion in terms of moments about the mean, Eq. (22), by making an expansion about  $\bar{\Gamma}$ . Equation (22) should be most accurate near  $\bar{\Gamma}$ ; this seems a more appropriate point of expansion when trying to determine the distribution function.

**Table 4. Comparison of the Robustness of the Fits of Model Function 2 [Eq. (18)] and Model Function 3 [Eq. (23)] to DLS Data for POPC Vesicles Extruded through 200-nm Pores<sup>a</sup>**

Model Function	Number of Iterations		Initial $\Gamma$ (1/ms)	Initial $\mu_2$ (1/ms <sup>2</sup> )
	Required for Convergence			
2	3		1	0.1
2	20		2	0.1
2	16		1	0.2
2	8		1	0.01
3	3		1	0.1
3	6		2	0.1
3	3		1	0.2
3	3		1	0.01

<sup>a</sup>Initial values of  $B = 1.0$  and  $\beta = 0.4$  were used in all cases. The fits were made to data ranging from 6.4  $\mu$ s to 6.55 ms.

In principle, a fit that uses Eq. (23) could include  $\mu_3$ . For the vesicle data, including  $\mu_3$  sometimes decreases  $\chi^2$  but sometimes increases it; this parameter does not seem to be significant for these data. Furthermore, use of  $\mu_3$  does not change the values of the other parameters significantly in these fits.

#### 4. Conclusions

Data from polydisperse samples as measured in DLS experiments can be analyzed in terms of the moments about the mean of the distribution function that describes the polydispersity of the sample. The traditional fitting function, as derived from the cumulants of the distribution, has several problems associated with it: It results in parameter values that depend on the number of data points fitted, and it does not permit an independent fit of the baseline  $B$ . This paper has demonstrated how a more robust fitting function can be obtained by the avoidance of the logarithm of  $g^{(1)}(\tau)$  and the direct expansion of  $g^{(1)}(\tau)$  in terms of the moments about the mean. The function can be fitted to the entire data set, gives consistent results for fitting parameters when different numbers of points are fitted, is more robust to bad guesses of the initial parameters, and permits an independent fit of the baseline  $B$ .

The author gratefully acknowledges helpful discussions with Art Bailey.

#### References and Note

1. D. E. Koppel, "Analysis of macromolecular polydispersity in intensity correlation spectroscopy: the method of cumulants," *J. Chem. Phys.* **57**, 4814–4820 (1972).
2. B. J. Berne and R. Pecora, *Dynamic Light Scattering* (Krieger, Malabar, Fla., 1990).
3. P. Stepanek, "Data analysis in dynamic light scattering," in *Dynamic Light Scattering*, W. Brown, ed. (Oxford University, Oxford, UK, 1993), pp. 177–240.
4. R. Pecora, *Dynamic Light Scattering* (Plenum, New York, 1985).
5. B. Chu, *Laser Light Scattering: Basic Principles and Practice* (Academic, Boston, 1991).
6. K. Schätzel, "Single-photon correlation techniques," in *Dynamic Light Scattering*, W. Brown, ed. (Oxford University, Oxford, UK, 1993), pp. 76–148.
7. A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics* (Wiley, New York, 1994), Chap. 4.
8. P. N. Pusey, D. E. Koppel, D. W. Schaefer, R. D. Camerini-Otero, and S. H. Koenig, "Intensity fluctuation spectroscopy of laser light scattered by solutions of spherical viruses," *Biochemistry* **13**, 952–960 (1974).
9. For details of extrusion methods for vesicle production as used in these experiments, please see B. J. Frisken, C. Asman, and P. J. Patty, "Studies of vesicle extrusion," *Langmuir* **16**, 928–933 (2000), and references therein.