

# A new robust estimator of polydispersity from dynamic light scattering data

## *Supporting Information*

Valentin Roger<sup>1</sup>, Hervé Cottet<sup>2</sup>, Luca Cipelletti<sup>1,\*</sup>

<sup>1</sup> Laboratoire Charles Coulomb (L2C), UMR 5221 CNRS, Université de Montpellier, 34095 Montpellier, France

<sup>2</sup> Institut des Biomolécules Max Mousseron (IBMM), UMR 5247 CNRS, Université de Montpellier, Ecole Nationale Supérieure de Chimie de Montpellier, 34095 Montpellier, France

\*email: luca.cipelletti@umontpellier.fr

We provide here details on:

1. The method used to integrate the area under  $g_1$ , Eq. 12 of the main text;
2. The PDFs of the diffusion coefficient used for the simulations and the corresponding  $g_2 - 1$  functions (Fig. SI1);
3. The accuracy and precision on the first cumulant and on the PI indexes as a function of the fitting range, for data simulated for a  $P(D)$  with  $\sigma_D = 1$  (Figs. SI2 and SI3);
4. The accuracy and precision on the first cumulant, as a function of sample polydispersity  $\sigma_D$ , of the CC and MA methods, for simulated  $g_2 - 1$  functions with a noise level  $\bar{\varepsilon} = 0.001$  (Fig. SI4);
5. The accuracy and precision of CC and MA and of the new polydispersity index as a function of sample polydispersity  $\sigma_D$ , for simulated  $g_2 - 1$  functions with a noise level  $\bar{\varepsilon} = 0.01$  (Fig. SI5);
6. The method used to convert the TEM and SEC size distribution data to the intensity-weighted  $P(D)$  curves shown in Fig. 5a and 5c of the main text.
7. The fits to the experimental  $g_2(t)$  for a polydisperse mixture of PSS polymers, and a comparison of the polydispersity index as obtained by the various methods.

## 1 Integration of the area under $g_1$

In order to minimize the impact of data noise on the measurement of the area under  $g_1$ , we implement Eq. 12 of the main text, proceeding as follows. First, we fit  $g_2$  to a (usually

2-nd order) Frisken cumulant function, Eq. 10 of the main text. A normalized intensity correlation function is obtained from

$$g_2^{(n)}(t) - 1 = [g_2(t) - B] / A, \quad (1)$$

where the superscript  $n$  stands for normalized and where  $A$  and  $B$  are the amplitude and baseline of the raw data obtained from the FC fit. If the quality of the fit is good, we also take  $\Gamma_1$  from the FC fit, otherwise we perform a CC fit on  $g_2^{(n)}(t) - 1$  to obtain a better estimate of the first cumulant. Next, we fit the tail of  $g_2^{(n)}(t)$  to a stretched exponential decay:

$$g_2^{(se)}(t) = A' \exp [-(t/\tau_c)^p] + 1, \quad (2)$$

with fitting range  $t \geq \Gamma_1^{-1}$ . Note that although  $g_2^{(n)}$  is normalized by construction,  $A'$  may be different from 1, because typically  $A'$ ,  $\tau_c$  and  $p$  are not independent fitting parameters. This is irrelevant, since the purpose of determining  $g_2^{(se)}(t)$  is just to obtain a noiseless function that reproduces well the large- $t$  behavior of the data, with no concern on the physical interpretation of the fitting parameters. Once  $A'$ ,  $\tau_c$  and  $p$  have been determined, the area under  $g_1$  is calculated from

$$\tau_{g_1} = \int_0^{\tau_{cut}} \sqrt{g_2^{(n)}(t) - 1} dt + \int_{\tau_{cut}}^{\infty} \sqrt{g_2^{(se)}(t) - 1} dt, \quad (3)$$

which is the practical implementation of Eq. 12 of the main text. The first integral in the r.h.s. of Eq. 3 is evaluated by integrating numerically the square root of the normalized, baseline-corrected  $g_2^{(n)}$ . The second integral may be calculated by numerical integration of the fit. Alternatively, one may use the equality

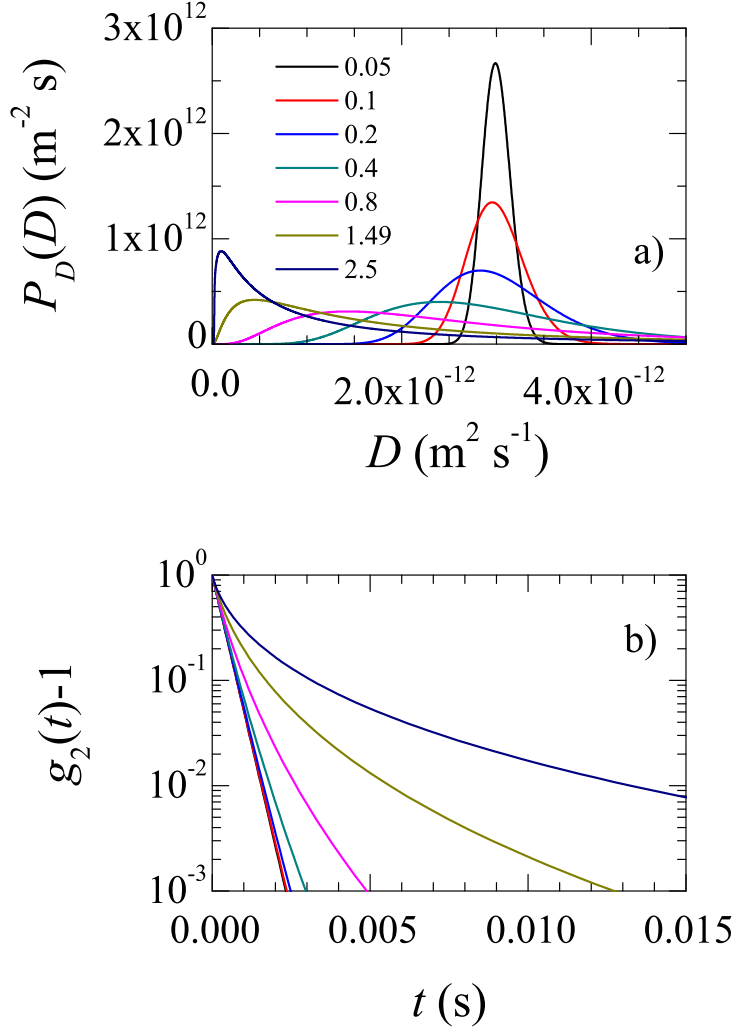
$$\int_{\tau_{cut}}^{\infty} \sqrt{g_2^{(se)}(t) - 1} dt = \sqrt{A'} \left[ \gamma \left( \frac{1}{p}, \infty \right) - \gamma \left( \frac{1}{p}, \frac{\tau_{cut}^p}{2} \right) \right], \quad (4)$$

where  $\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$  is the lower incomplete gamma function [1], available in many scientific calculation packages.

## 2 Probability distribution functions used for the simulations

Figure SI1a) shows some of the PDF of  $D$  used to generate the simulated correlation functions. The corresponding  $g_2 - 1$  are displayed in a semi-log scale in Fig. SI1b. Note that for a relative polydispersity  $\sigma_D \leq 0.2$ , the intensity correlation functions are very close to straight lines, the behavior of monodisperse samples. All the PDF have the same first moment,  $\overline{D} = 3 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$ , which corresponds to a size of about 70 nm for Brownian objects in water at room temperature. The PDFs have a log-normal shape, but the tails are truncated at small and large  $D$  to avoid unphysical values of the diffusion coefficient. Specifically,  $P(D)$  is obtained by normalizing the following tail-truncated log-normal distribution:

$$\frac{1}{D\sqrt{2\pi\beta}} \exp \left[ -\frac{(\ln D - \mu)^2}{2\beta^2} \right] f_+(D) f_-(D), \quad (5)$$



**Figure SI1.** a):  $P(D)$  used to generate the simulated  $g_2 - 1$ . Curves are labelled by  $\sigma_D$ , the relative standard deviation of  $P(D)$ . For the sake of clarity, not all the distributions are shown. b): Intensity correlation functions generated from the PDFs shown in a) (same line colors). For clarity, in this graph no noise has been added to  $g_2 - 1$ .

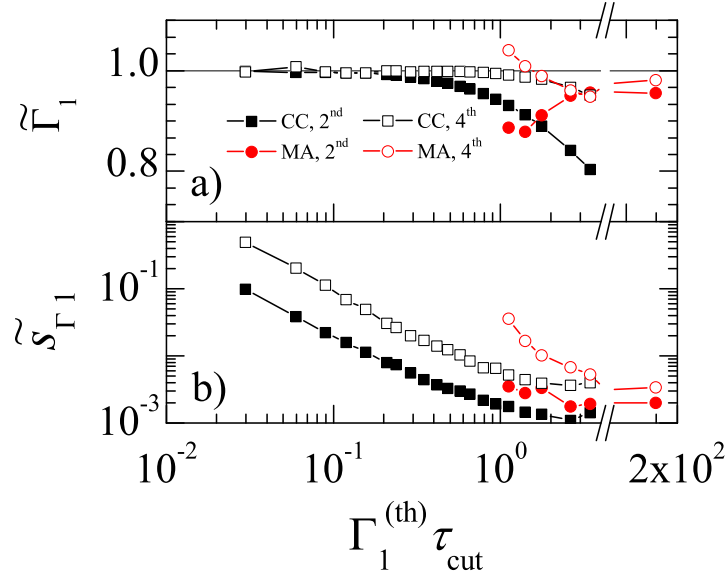
where the truncating function is

$$f_{\mp}(D) = \frac{1}{2} \left[ 1 \pm \tanh \left( 5 \frac{D - D_{\mp}}{D_{\mp}} \right) \right], \quad (6)$$

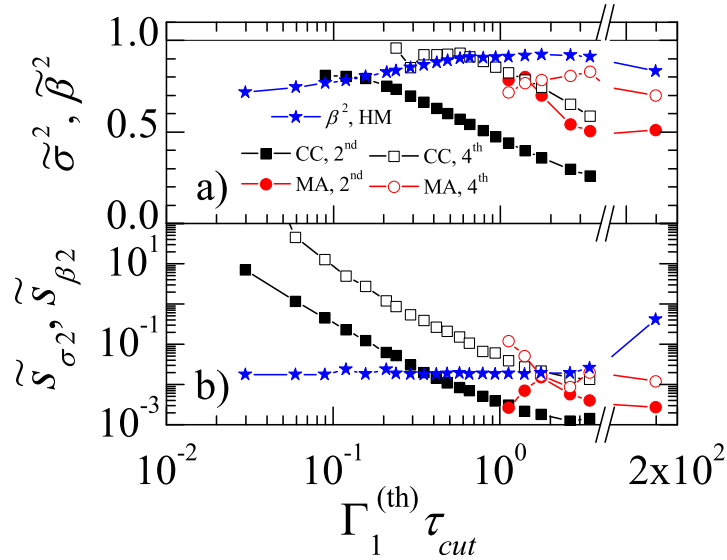
and where the lower and upper cutoff  $D_{\mp}$  correspond to a particle size of  $10 \mu\text{m}$  and  $1 \text{ nm}$ , respectively. The values of  $\sigma_D$  quoted in the main text are calculated using the truncated PDF, Eq. 5.

### 3 Optimum fitting range for data generated from a PDF with $\sigma_D = 1$

Figure SI2 shows the impact of the choice of the fitting range on the accuracy and precision of the first cumulant, as in Fig. 2 of the main text, but for a more polydisperse sample ( $\sigma_D = 1$  instead of  $\sigma_D = 0.2$  as in Fig. 2 of the main text).



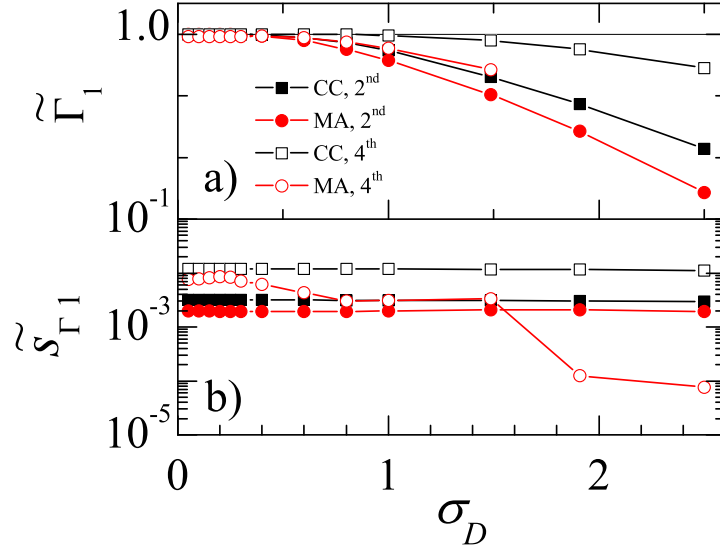
**Figure SI2.** Choice of the fitting range for obtaining  $\Gamma_1$ . a): normalized first cumulant as a function of the maximum delay time included in the fit,  $\tau_{cut}$ . The curves are labelled by the order and the kind of the fit. b) normalized uncertainty on the first cumulant, same symbols as in a). Data are obtained by analyzing 103 simulated  $g_2 - 1$ , generated from a  $P_D(D)$  with  $\sigma_D = 1$  and using a noise level  $\bar{\varepsilon} = 0.001$ .



**Figure SI3.** Choice of the fitting range for obtaining the polydispersity index. a): normalized polydispersity index  $\tilde{\sigma}^2$  and new polydispersity index  $\tilde{\beta}^2$  as a function of the maximum delay time included in the fit,  $\tau_{cut}$ . Black and red symbols: same as in Fig. SI2; blue stars:  $\beta^2$  normalized by its theoretical value. b) normalized uncertainty on the polydispersity indexes, same symbols as in a). Data are obtained by analyzing the same simulated functions as in Fig. SI2.

Figure SI3 shows the impact of the choice of the fitting range on the accuracy and precision of the polydispersity indexes, as in Fig. 3 of the main text, but for a more polydisperse sample ( $\sigma_D = 1$  instead of  $\sigma_D = 0.2$  as in Fig. 3 of the main text).

#### 4 Accuracy and precision on $\Gamma_1$ as a function of sample polydispersity, for data with noise $\bar{\varepsilon} = 0.001$

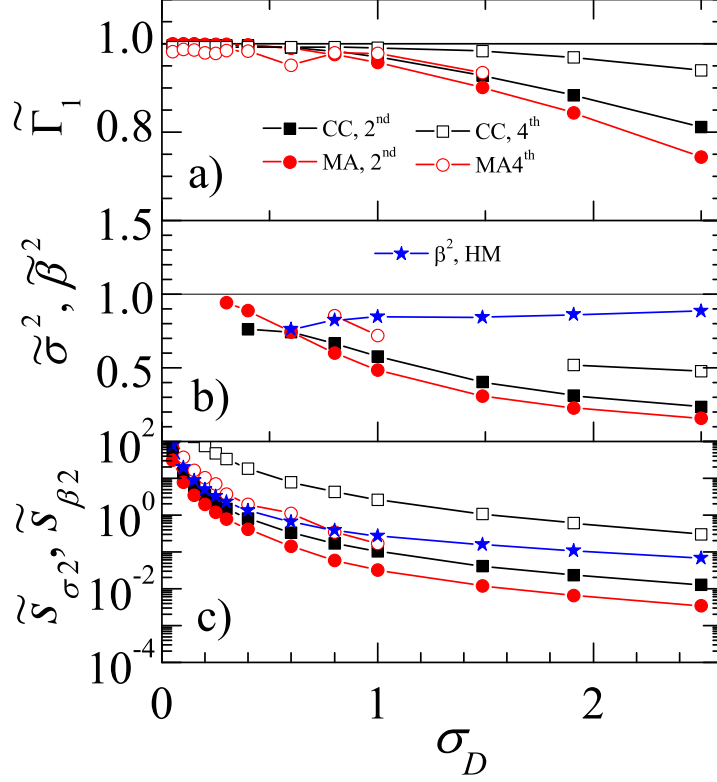


**Figure SI4.** Normalized  $\Gamma_1$  (panel a) and its normalized uncertainty (b) as a function of sample polydispersity  $\sigma_D$ , obtained by analyzing simulated  $g_2 - 1$  functions with a noise level  $\bar{\varepsilon} = 0.001$ .

Figure SI4 shows the accuracy and precision on the first cumulant, as a function of sample polydispersity  $\sigma_D$ , of the CC and MA methods, for simulated  $g_2 - 1$  functions with a noise level  $\bar{\varepsilon} = 0.001$ . The accuracy and precision on the polydispersity indexes for the same functions are shown in Fig. 4 of the main text.

#### 5 Accuracy and precision on $\Gamma_1$ and on the PI indexes, for data with noise $\bar{\varepsilon} = 0.01$

Figure SI5 shows the impact of sample polydispersity on the accuracy and precision of  $\Gamma_1$  and of the polydispersity indexes, as in Fig. 4 of the main text, but for noisier correlation functions (the noise level is here  $\bar{\varepsilon} = 0.01$ , ten times larger than in Fig. 4 of the main text, corresponding indicatively to  $T_{exp} = 22$  s according to Fig. 1 of the main text). The results are similar to those for  $\bar{\varepsilon} = 0.001$ , except that the range over which acceptable fit parameters are obtained is further restricted. In particular, no method is able to measure polydispersity with less than 100% uncertainty for  $\sigma_D < 0.3$ ; the 2nd-order MA is the best method for  $0.3 \leq \sigma_D < 0.6$ ;  $\beta^2$  outperforms all other methods for  $\sigma \geq 0.6$ .



**Figure SI5.** Normalized  $\Gamma_1$  and polydispersity indexes as a function of sample polydispersity  $\sigma_D$ , obtained by analyzing simulated  $g_2 - 1$  functions with a noise level  $\bar{\varepsilon} = 0.01$ . a): first cumulant. b) polydispersity indexes. c): normalized uncertainty on the polydispersity indexes.

## 6 Conversion of the PDF obtained by TEM and SEC to intensity-weighted $P(D)$

In Figure 5a, 5c of the main text the equivalent log-normal PDF of  $D$  obtained by analyzing DLS data is compared to the distributions obtained by TEM or SEC. Since different techniques probe in general different moments of the size distribution, the TEM and SEC PDFs must be properly weighted prior to comparison with  $P(D)$ . For the PMMA data, one obtains from the TEM images  $N_R(R)$ , the number density of particles with radius between  $R$  and  $R + dR$ . Since  $P(D)$  is weighted by the scattering intensity (see Eq. 6 of the main text), we need to multiply  $N_R(R)$  by  $I_s(q, R) = 9(qR)^{-6} [\sin(qR) - qR \cos(qR)]^2$ , the form factor for a homogeneous sphere [2,3]. The intensity-weighted PDF of  $R$  is thus

$$P_R(R) = \frac{N_R(R)I_s(q, R)}{\int_0^\infty N_R(R)I(q, R)dR}. \quad (7)$$

As a final step,  $P_R(R)$  is transformed to a PDF of  $D$  (which we shall denote by  $P_{TEM}(D)$ ) using the Stokes-Einstein relation  $D(R) = \frac{k_B T}{6\pi\eta R}$  (Eq. 4 of the main text) and the standard probability transformation law:

$$P_{TEM}(D) = \left[ P_R(R) \left| \frac{dR}{dD} \right| \right]_{R=R(D)}. \quad (8)$$

We find a  $\sim 20\%$  discrepancy between  $\bar{D}$  obtained directly from DLS and from  $P_{TEM}(D)$ , due to the fact that the hydrodynamic radius probed by DLS is larger than the geometric

radius of the particles, seen in the TEM images. Since here we are interested only in comparing the shape and width of the distributions, in Fig. 5a of the main text we have rescaled  $D$  for the TEM data so that the average diffusion coefficient calculated from  $P_{TEM}(D)$  matches that obtained by DLS.

A similar series of transformations is applied to  $P_M^{(M)}(M)$ , the mass-weighted distribution of molecular masses obtained by SEC for the PSS polymers (the superscript denotes weighting by mass). The intensity-weighted PDF of  $M$  is given by

$$P_M(M) = \frac{P_M^{(M)}(M)M^{-1}I_c[q, R(M)]}{\int_0^\infty P_M^{(M)}(M)M^{-1}I_c[q, R(M)]dM}. \quad (9)$$

In the above equation, the term  $M^{-1}$  has been introduced to remove the mass weighting. For the intensity weighting, we use

$$I_c(q, R) = \frac{2}{(qR)^4} \left[ e^{-(qR)^2} + (qR)^2 - 1 \right], \quad (10)$$

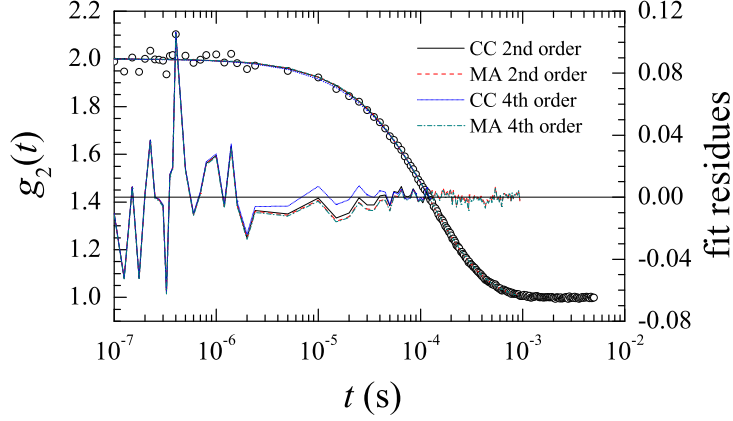
the form factor for a Gaussian coil [3]. The desired intensity-weighted PDF of  $D$  is finally calculated from

$$P_{SEC}(D) = \left[ P_M(M) \left| \frac{dM}{dD} \right| \right]_{M=M(D)}. \quad (11)$$

Note that in Eqs. 10 and 11 one needs the functional relation between  $M$  and  $D$  and  $R$ . We assume a fractal-like shape of the polymer (as in the Mark-Houwink law relating the intrinsic viscosity to  $M$ ), implying  $D = KM^b$ . By fitting  $D_{hm}$  (obtained from separate DLS analysis on the PSS29k and PSS145k samples) *vs* the nominal mass to this power law, we find  $b = -0.64$  and  $K = 1.56 \times 10^{-8} \text{ m}^2\text{s}^{-1}(\text{g mol}^{-1})^b$ . The law  $R(M)$  follows from the Stokes-Einstein relation:  $R = k_B T / (6\pi\eta K M^b)$ . Equation 11 is applied separately to the PDFs of the PSS29k and PSS145k samples; the PDF for the mixture is simply obtained by normalizing the sum of those of the individual species, since each species contributes equally to the scattering signal.

## 7 Analysis of the experimental $g_2(t)$ for the PSS mixture

Figure SI6 shows the experimental intensity correlation function for the mixture of PSS29k and PSS145k polymers, together with fits using 2nd and 4th order CC and MA. Although the difference in the quality of the fits is rather small, the polydispersity values issued from the various cumulant and moment methods differ significantly, as seen in Table SI T1. The results of the analysis of one single correlation function, as it is the case here, should be taken with some caution, since in the main text we have shown that the uncertainty on the polydispersity indexes can be significant. However, some comments are in order. The 2nd order CC fit performs well, while the 4th order CC fit overestimates  $\sigma_D$  almost by a factor of 2. This may seem surprising since for simulated data (see Fig. 4a of the main text)  $\sigma^2$  for the 4th order CC was shown to be close to the expected value. However, Fig. 4b shows that for 4th order CC the uncertainty is quite large, of the order of 40% for  $\sigma^2$ . The anomalously large value of  $\sigma$  found here is therefore consistent, within uncertainty, with the results for simulated data. This emphasises the importance of the precision with which the polydispersity index can be obtained, especially given that a poor performance



**Figure SI6.** Left axis: CC and MA fits of order 2 and 4 of the intensity correlation function for the mixture of PSS29k and PSS145k polymers. Right axis: fit residues.

SEC	CC 2nd	MA 2nd	CC 4th	MA 4th	from $\beta^2$
0.80	0.68	0.60	1.58	0.61	0.71

**Table SI T1** Polydispersity of the mixture of PSS29k and PSS145k, as obtained by SEC, by CC and MA or order 2 and 4, and by the new polydispersity index  $\beta$ . For the latter, we quote the relative standard deviation of the equivalent log normal distribution, obtained through  $\sigma = \sqrt{\exp \beta^2 - 1}$  (see Eq. 14 and the following discussion in the main text).

often cannot not be inferred from the fit quality. This is indeed the case here, since the 4th order CC fit is excellent in spite of the aberrant value of  $\sigma$  (see Fig. SI6). The 2nd and the 4th order MA fits underestimate  $\sigma$  by about 25%; the polydispersity issued from the new index  $\beta$  is the closest to the expected one (as estimated by SEC), consistently with our findings for simulated data in the regime of large polydispersity,  $\sigma_D > 0.6$ .

## References and notes

- [1] Abramowitz, Milton; Stegun, Irene A., eds. (1965), Chapter 6, Section 5, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*.
- [2] For the dilute suspensions of our experiments, the scattered intensity is proportional to the form factor.
- [3] P. N. Pusey, *Introduction to Scattering Experiments*, in *Neutrons, X-Rays and Light Scattering Methods Applied to Soft Condensed Matter*, P. Linder and Th. Zemb eds, North Holland, Amsterdam (2002).