

Cite this: *Phys. Chem. Chem. Phys.*, 2013, **15**, 845

Measuring a diffusion coefficient by single-particle tracking: statistical analysis of experimental mean squared displacement curves†

Dominique Ernst and Jürgen Köhler*

We provide experimental results on the accuracy of diffusion coefficients obtained by a mean squared displacement (MSD) analysis of single-particle trajectories. We have recorded very long trajectories comprising more than 1.5×10^5 data points and decomposed these long trajectories into shorter segments providing us with ensembles of trajectories of variable lengths. This enabled a statistical analysis of the resulting MSD curves as a function of the lengths of the segments. We find that the relative error of the diffusion coefficient can be minimized by taking an optimum number of points into account for fitting the MSD curves, and that this optimum does not depend on the segment length. Yet, the magnitude of the relative error for the diffusion coefficient does, and achieving an accuracy in the order of 10% requires the recording of trajectories with about 1000 data points. Finally, we compare our results with theoretical predictions and find very good qualitative and quantitative agreement between experiment and theory.

Received 28th September 2012,
Accepted 13th November 2012

DOI: 10.1039/c2cp43433d

www.rsc.org/pccp

1 Introduction

Diffusion is often exploited to examine interactions and movements of individual nanoscopic objects in soft matter and/or biological environments on a molecular length scale.^{1–10} Early experimental work was carried out using fluorescence recovery after photobleaching (FRAP)¹¹ which yields the ensemble average of the diffusing particles and which is, dictated by the diffraction of light, restricted in spatial resolution to length scales of about 200–300 nm. Alternatively, researchers employed fluorescence correlation spectroscopy (FCS),^{12,13} which provides the average over a number of individual objects that are registered sequentially and from which it is assumed that they behave uniformly. Since about two decades single-particle tracking (SPT) became a valuable tool to map out the movement of an individual particle with high spatial and temporal resolution.^{1–4,14–21} The developed methodology covers techniques where the movement of an individual particle can be followed by recording its diffraction-limited image on a sequence of CCD frames,¹ sophisticated approaches that compensate the Brownian motion using

electrokinetic forces,¹⁴ techniques that use structured illumination by actively designing the point-spread function of the microscope,^{21,22} as well as methods that rely on a spatial modulation of the light that travels to or comes from the particle.¹⁹ Fascinating results have been obtained, for example in biophysics the movement of molecules, viruses, or motor proteins could be made visible,^{4–6} and in the materials science transport processes through nanoporous structures^{7,8} or the manifestation of diffusion anomalies in liquid crystals and mesoporous structures could be followed.^{9,23,24}

Typically the fluorescence of a particle is monitored as a function of time and the position of the particle is extracted from the data with sub-diffraction limited accuracy. This provides the trajectory $\mathbf{r}(t)$ of the particle that is commonly analysed in terms of the mean squared displacement (MSD). For a 2-dimensional diffusion process the MSD generally scales with a power law according to $\text{MSD}(\tau) = 4\tilde{D}\tau^\alpha$, where \tilde{D} is the generalized diffusion coefficient, and α the anomaly parameter. For $\alpha = 1$ the underlying process corresponds to normal diffusion (Brownian motion)²⁵ and \tilde{D} reduces to the diffusion coefficient D known from Einstein.²⁶ Otherwise the process is called subdiffusive ($\alpha < 1$) or superdiffusive ($\alpha > 1$).

For obvious reasons an experimental trajectory can only be recorded with a distinct temporal resolution, it suffers from localisation errors due to the movement of the particle during data acquisition,²⁷ it is affected by unavoidable signal-to-noise

Experimental Physics IV and Bayreuth Institute of Macromolecular Research (BIME), University of Bayreuth, 95440 Bayreuth, Germany. E-mail: juergen.koehler@uni-bayreuth.de; Fax: +49 921 55 4002; Tel: +49 921 55 4000

† Electronic supplementary information (ESI) available: Analysis of all measured trajectories. See DOI: 10.1039/c2cp43433d

limitations²⁸ and last but not least it is inherently of finite length. Hence, it is of crucial importance to understand how accurate the diffusion coefficient can be extracted from a real experimental MSD curve.^{25,29,30} Although the mathematical framework for the MSD analysis is known for many years, the implications of the experimental limitations on the accuracy of the measured parameters have been considered in detail only recently.^{29,31} These studies address the achievable precision for the diffusion coefficient that can be obtained from a given MSD curve as a function of the experimental parameters mentioned above. In order to test their theoretical results the authors had to rely on fictitious tracking experiments based on simulations rather than on experimental data. The reason is that a sophisticated statistical analysis of the tracking data requires a very large data set which is difficult to obtain, because the particle might get lost for tracking due to diffusion out of the focal volume or due to photobleaching. Often it is already a great challenge to register trajectories consisting of some hundred data points. Naively speaking, a trajectory of arbitrary length could be acquired by repeating a tracking experiment under exactly the same experimental conditions on several nanoparticles. However, since even nominally identical nanoparticles are slightly different in shape and size, the statistics of the (unknown) size distribution of the nanoparticles will be superimposed on the statistics of the diffusion coefficient. Moreover, the precision with which the diffusion coefficient can be determined from a MSD curve depends on the accuracy of the MSD data points and on the number of fitting points that are taken into account.^{29,30} Therefore, the numerous theoretical and numerical developments that allow assessment of the experimental shortcomings still await experimental verification.

In this work we use single-particle orbit tracking, which allows us to obtain single-particle trajectories that consist of more than 1.5×10^5 data points with a temporal resolution of 4 ms and a spatial accuracy of better than 10 nm.³² Such a long trajectory can be divided into a sequence of segments, where each segment can be considered as an individual trajectory that, by definition, has been recorded on exactly the same particle under identical experimental conditions. This enables us to evaluate the statistics of the diffusion coefficient extracted from the segments as a function of the length of the segments and to compare the results with the theoretical predictions made in the literature.^{25,29}

2 Experimental section

2.1 Sample preparation

For the single-particle tracking experiments we use fluorescent beads with a diameter of 20 nm that are loaded with Nile red (Molecular Probes, 20 mg ml⁻¹ dissolved in water). This solution is further diluted in water to a concentration of 0.1 nM and subsequently mixed with pure glycerol (Sigma) resulting in a concentration of 2 pM for the tracers. From that solution a drop of about 25 μ l is sandwiched between two microscope coverslips that are cleaned with acetone. In order to prevent evaporation of the solvent (and the resulting flow field in the sample) the edges of the coverslips are sealed with grease (High-Vacuum Grease, Wacker).

This construct is mounted on top of a 3-axis piezo stage (Tritor 102, piezosystem Jena) providing a scan range of 100 μ m for each axis. All experiments are performed at room temperature, *i.e.* (21 ± 0.5) °C.

2.2 Experimental setup

The home-built setup for single-particle orbit tracking has been described in great detail in a separate paper.³² Briefly, the output from an Ar/Kr-ion laser (Innova 70C Spectrum, Coherent) operated at 514 nm is guided through a deflection unit consisting of two mutually perpendicular arranged acousto optical deflectors (AOD, DTSX-400-532, Pegasus) that generate a rotating light orbit. This orbit is projected via a dichroic beam splitter (z532RDC, AHF) towards an infinity-corrected water-immersion objective (UPLSAPO, 60 \times , NA = 1.2, Olympus). This results in a focussed laser beam with a waist of $w = 270$ nm that rotates on an orbit with a radius of $R = 190$ nm in the focal plane of the objective. The frequency of the rotation can be adjusted by the AODs and is set to 1 kHz.

The emission of the fluorescent nanoparticles is collected with the same objective, passes the dichroic and is focussed either onto the chip of a CCD (sencam qe, PCO) or an avalanche photo diode (SPCM-AQR-14, Perkin Elmer). Residual laser light that passes the dichroic is suppressed by a dielectric optical filter (HQ545LP, OD = 6 at 514 nm, AHF). To spot the location of the tracers we operate the setup in widefield mode. Therefore the deflection unit is switched off and an additional lens in the excitation path defocusses the laser light to an area of $80 \times 80 \mu\text{m}^2$. With the aid of the piezo stage an appropriate particle is moved close to the position where the light orbit will appear (centre of the field of view). Subsequently, the optics are switched to confocal mode, the light orbit is generated and the algorithm for automated tracking is started.

We record the emission intensity of the fluorescent particle which is modulated by the frequency of the rotating laser focus. By demodulating this emission signal we are able to calculate the x -, y -position of the particle with respect to the centre of the orbit. The position provides a feedback signal for the piezo and the particle is moved (together with the sample) back to the centre of the orbit. These steps (collect emission – calculate position – move piezo) are repeated continuously, which allows us to reconstruct the movement of a fluorescent tracer particle for more than 10 minutes with a spatial resolution of better than 10 nm. The temporal resolution of the experiments is $\Delta t = 4$ ms which results in trajectories of $N = 1.5 \times 10^5$ x -, y -position pairs.

3 Results and discussion

An example for a typical trajectory measured with our setup is displayed in Fig. 1. It represents 1.52×10^5 data points and corresponds to an elapsed time of 608 s which is indicated by the colour code, where blue corresponds to the start of the trajectory and red to its end. In order to mimic to have only shorter trajectories with less data points we cut the long trajectory into segments that were treated as independent

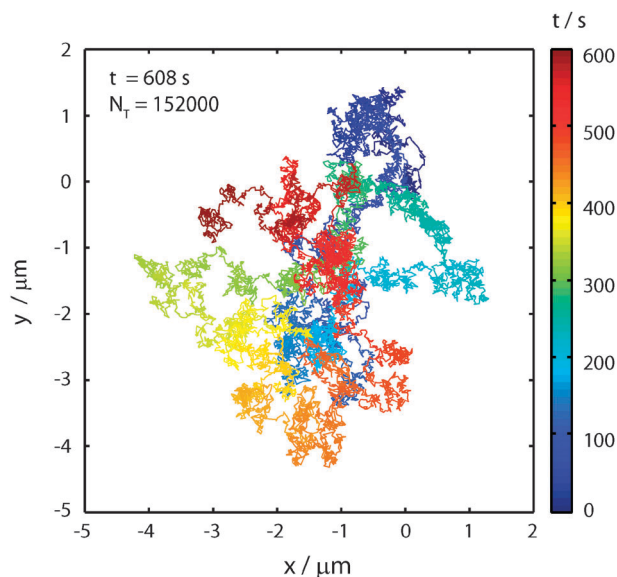


Fig. 1 Example of a trajectory of a 20 nm sized bead in pure glycerol. The colour code refers to the elapsed time of 608 s ($W = 1.52 \times 10^5$ data points; blue corresponds to the start and red corresponds to the end).

trajectories of shorter length. In the following we denote the length of the full trajectory as N_T (here $N_T = 1.52 \times 10^5$) and the length of a segment as N_{seg} . For our study we choose $N_{\text{seg}} = 10, 20, 40, 60, 80, 100, 200, 400, 600, 800,$ and 1000 , which yields ensembles of N_T/N_{seg} short trajectories of equal length. The idea is now to determine the diffusion coefficient D from the slope of the time-averaged MSDs of each segment and to examine the statistical variation of D within each ensemble of trajectories. Yet, according to ref. 29 there exists an optimum number of data points of the MSD that should be considered to obtain the best result for the diffusion coefficient. This can be understood as follows. For increasing lag times the accuracy of the data points in the MSD decreases due to the progressively decreasing averaging of the available data. For example, the first data point of the MSD represents an average over $(N_{\text{seg}} - 1)$ positions of the particle whereas the last data point has not been averaged at all. Hence, fitting the slope of the MSD curve by taking too many data points into account leads to a deterioration rather than an improvement of the result. On the other hand, the very first points of the MSD are stronger subjected to localisation errors, either due to noise (static error) or due to blurring of the position of the particle during data acquisition (dynamic error). Both effects average out for MSD points at longer lag times.

As a consequence of this, we first have to find out the optimum number of data points that should be considered for fitting the slope of the MSD. In the following, the protocol for doing so will be explained on the example of $N_{\text{seg}} = 1000$ which yields an ensemble of 152 trajectories of equal length and the same number of MSD curves. For this ensemble we fitted the slope, D^* , of each MSD curve by an unweighted linear fit to the first n data points. More precisely, we have skipped the very first data point of the MSDs, because it turned out that it is strongly affected by residual oscillations of the piezo. These

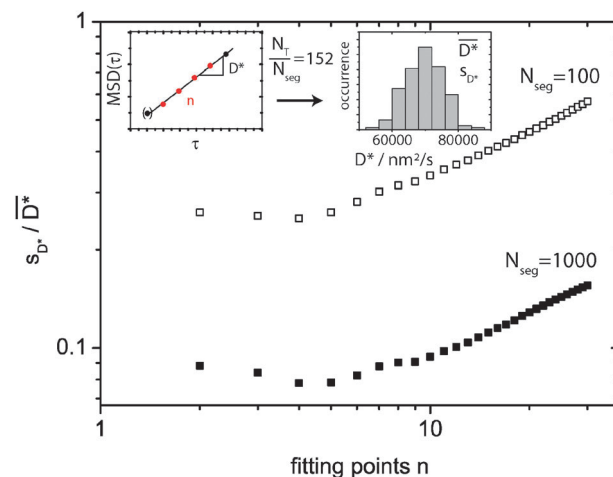


Fig. 2 Relative error of the slope D^* obtained from unweighted linear fits to the MSD curve as a function of the number of fitting points n for the segment length $N_{\text{seg}} = 100$ (open symbols) and $N_{\text{seg}} = 1000$ (full symbols). The inset top left displays schematically a MSD curve as a function of the lag time τ and the data points that are considered for the linear fit (red) to obtain D^* . For all fits the first data point of the MSD (brackets) is ignored (for details see text). The inset top right shows as an example for the distribution of the slopes within the ensemble of N_T/N_{seg} trajectories for $n = 4$ and $N_{\text{seg}} = 100$, from which \bar{D}^* (first moment; empirical mean value) and s_{D^*} (second moment; empirical standard deviation) can be calculated.

oscillations affect the position determination and reduce the accuracy of the first point of the MSD curve, whereas the influence of these oscillations on the accuracy of the succeeding MSD points level off (for details see Experimental section and ref. 32). Therefore the fit was applied to the data points from 2 to $(n + 1)$ and the slope D^* of the MSD curves was determined as a function of n . In order to be compatible with the existing literature we prefer the slope D^* of the MSD curves rather than the diffusion coefficient $D = D^*/4$.²⁹ An example for the distribution of D^* is shown in the top right inset of Fig. 2 for $n = 4$, *i.e.* taking only the data points 2–5 for fitting the MSDs into account as indicated schematically in the top left inset of Fig. 2. Subsequently, we determined from each histogram the first and the second moment providing the empirical mean \bar{D}^* and the empirical standard deviation s_{D^*} for this parameter, and plotted the ratio s_{D^*}/\bar{D}^* as a function of the number of fitting points n . The result of this procedure is shown in Fig. 2 for the examples of $N_{\text{seg}} = 100$ and $N_{\text{seg}} = 1000$. For both samples, the relative error s_{D^*}/\bar{D}^* first decreases for growing n and then rapidly increases if more fitting points are taken into account. Here we find an optimum for the accuracy of the slope of the MSDs for $n = 4$. While the relative accuracy that can be achieved for D^* (about 8% for $N_{\text{seg}} = 1000$, and about 25% for $N_{\text{seg}} = 100$) clearly depends on the lengths of the segments, it is interesting to note that the number of fitting points n that yield the optimum result does not.

In order to facilitate a quantitative comparison of the data shown in Fig. 2 with the theoretical predictions in the literature²⁹ we have to resort to the reduced localisation error $x = \sigma^2/\Delta t$, where σ is the localisation error, D the diffusion

coefficient, and Δt the temporal resolution. A good estimate for the diffusion coefficient is obtained from a linear fit to the data points 2–5 of the time-averaged MSD from the total trajectory of 1.52×10^5 data points, which yields $D_T = 17.45 \times 10^{-3} \mu\text{m}^2 \text{s}^{-1}$. Using $\sigma = 7.5 \text{ nm}^{32}$ and $\Delta t = 4 \text{ ms}$ we obtain $x \approx 0.8$ and find a good agreement between experiment and theory (see Fig. 6 for $x = 1$ in ref. 29).

Next we investigated the influence of the length of the segments on the outcome of the experimentally determined diffusion coefficient. Therefore we analysed the MSDs from 15 200 segments with a length of $N_{\text{seg}} = 10$, from 7600 segments with a length of $N_{\text{seg}} = 20$ and so on until 152 segments with a length of $N_{\text{seg}} = 1000$, by applying a linear fit to the data points 2–5 of the respective MSD curves. The resulting distributions for the diffusion coefficients $D = D^*/4$ are shown in Fig. 3 as a function of N_{seg} . For better comparison the histograms were scaled with the number of sub-trajectories in each ensemble, N_T/N_{seg} . For the short segments the distributions were asymmetric and extremely broad with a long tail toward larger values of the diffusion coefficient testifying that diffusion coefficients obtained from trajectories consisting only of a few data points are not very meaningful. As N_{seg} increases the distributions narrow down and become more symmetric. The shape and profile of the experimental distributions presented here are consistent with theory.³⁰

The distributions presented in Fig. 3 can be interpreted as the empirical probability density functions (PDFs) to measure a distinct range of values for the diffusion coefficient for a given length N_{seg} of the trajectory. This tells us that for trajectories with a length in the order of 100 data points (black bold line in Fig. 3), which would be very reasonable for experiments in biological environments, the actual outcome of an experiment for the diffusion coefficient can vary by more than a factor of 2. For each of these empirical probability densities, the best

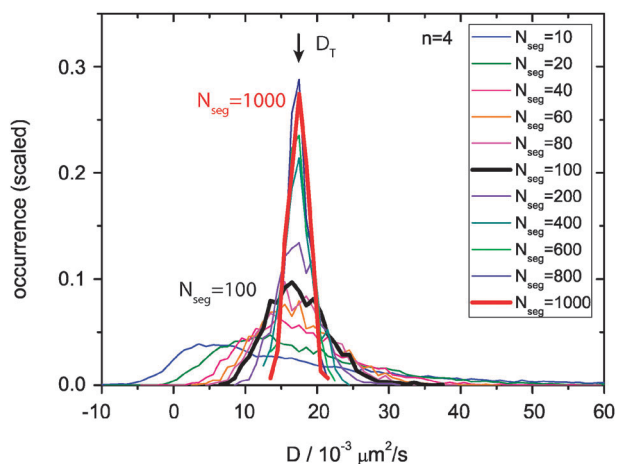


Fig. 3 Distributions of diffusion coefficients as a function of the segment length N_{seg} . For comparison, the distributions were scaled with the number of trajectories N_T/N_{seg} within the corresponding ensemble. The diffusion coefficients were obtained from linear fits to the MSD curves using $n = 4$ fitting points. The arrow on top of the distributions indicates the diffusion coefficient D_T determined from the time-averaged MSD of the full trajectory with a length of $N_T = 1.52 \times 10^5$ data points. The distributions for $N_{\text{seg}} = 100, 1000$ are highlighted by the bold lines.

estimate for the actual diffusion coefficient corresponds to the mean value \bar{D} (first moment) which is displayed in Fig. 4 (open squares) together with the corresponding empirical standard deviations (second moment, black error bars) as a function of N_{seg} . Interestingly the mean values show only little variation as a function of N_{seg} which is shown in the inset of Fig. 4 on an expanded scale.

In addition to the distribution of the diffusion coefficients as a function of the length of segments Qian *et al.*²⁵ derived an analytical expression for the standard deviation of these distributions which is given by $\sigma = \pm \bar{D} [2n/3(N_{\text{seg}} - n)]^{1/2}$ and which allows for comparison with the empirical standard deviation obtained experimentally. The calculated standard deviation is displayed in Fig. 4 by the red error bars. For better visualization both the experimental (black) and calculated (red) error bars are connected by the dashed and dotted lines, respectively. The systematically larger widths of the experimental error bars can be explained to result from errors in the position determination that were not considered in the theoretical calculation by Qian *et al.* Besides this minor discrepancy both characteristics are in good agreement. For $N_{\text{seg}} > 200$ the mean values of the distributions, \bar{D} , show only very little variation and are close to the value D_T obtained from the time-averaged MSD of the full trajectory (inset in Fig. 4, grey line).

However, it should be kept in mind that \bar{D} can only be determined if the distribution of D for the respective length of the trajectory is known. A single tracking experiment, for example with $N_{\text{seg}} = 100$, provides the diffusion coefficient only with an accuracy of about $\pm 25\%$. Achieving an accuracy of better than $\pm 10\%$ requires trajectories of $N_{\text{seg}} = 1000$ as testified in Fig. 2 and 3.

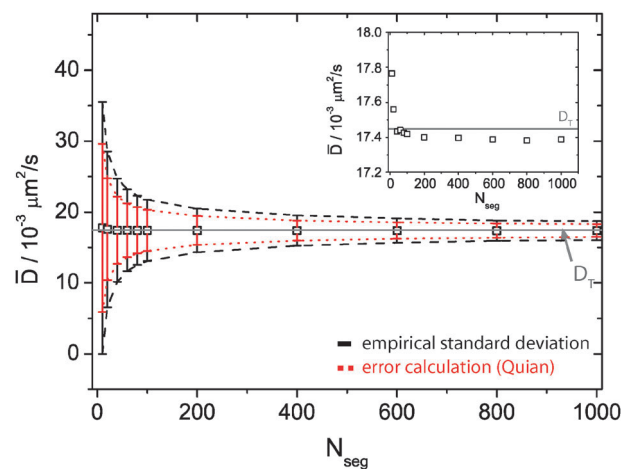


Fig. 4 Means of the diffusion coefficient, \bar{D} (open squares), and empirical standard deviations, s_D (black error bars), taken from the distributions shown in Fig. 3 as a function of the segment length N_{seg} . Theoretical values for the standard deviations (red error bars) have been calculated according to Qian *et al.*²⁵ using $\sigma = \pm \bar{D} [2n/3(N_{\text{seg}} - n)]^{1/2}$ and $n = 4$. As a guide for the eye the error bars are connected by the dashed (experimental data) and dotted (calculated data) lines. The inset displays the means of the diffusion coefficients on an enlarged scale. The grey line corresponds to the diffusion coefficient that is determined from the full time-averaged MSD curve of the full trajectory of 1.52×10^5 data points.

4 Conclusions

We analysed how the accuracy of the diffusion coefficient that is determined by fitting MSD curves depends on experimental parameters such as the length of the underlying trajectory and the number of fitting points of the MSD. This became possible by recording extremely long trajectories that could be cut into a large number of segments of variable lengths, thereby providing sufficiently large ensembles of (sub)trajectories that served for a statistical analysis. Our results were compared with theoretical predictions and we found good quantitative and qualitative agreement between experiment and theory.

The analysis presented was based on data from a single very long trajectory. In total we measured 7 very long trajectories from different tracers and found similar results. These are summarized in the ESI.†

Acknowledgements

We thank Florian Schwaiger for fruitful discussions and gratefully acknowledge financial support from the German Science Foundation (DFG) within the framework of the Research Unit “Nichtlineare Dynamik komplexer Kontinua” (FOR 608).

References

- J. Gelles, B. J. Schnapp and M. P. Sheetz, *Nature*, 1988, **331**, 450.
- B. J. Schnapp, J. Gelles and M. P. Sheetz, *Cell Motil. Cytoskeleton*, 1988, **10**, 47.
- B. W. Hicks and K. J. Angelides, *J. Membr. Biol.*, 1995, **144**, 231.
- T. Schmidt, G. J. Schütz, W. Baumgartner, H. J. Gruber and H. Schindler, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 2926.
- G. Seisenberger, M. U. Ried, T. Endreß, H. Buning, M. Hallek and C. Bräuchle, *Science*, 2001, **294**, 1929.
- E. J. G. Peterman, H. Sosa and W. E. Moerner, *Annu. Rev. Phys. Chem.*, 2004, **55**, 79.
- C. R. Haramagatti, F. H. Schacher, A. H. E. Müller and J. Köhler, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2278.
- J. Kirstein, B. Platschek, C. Jung, R. Brown, T. Bein and C. Bräuchle, *Nat. Mater.*, 2007, **6**, 303.
- B. Schulz, D. Täuber, F. Friedriszik, H. Graaf, J. Schuster and C. v. Borczyskowski, *Phys. Chem. Chem. Phys.*, 2010, **12**, 11555.
- D. Ernst, M. Hellmann, J. Köhler and M. Weiss, *Soft Matter*, 2012, **8**, 4886.
- E. L. Elson, *Annu. Rev. Phys. Chem.*, 1985, **36**, 379.
- E. L. Elson and D. Madge, *Biopolymers*, 1974, **13**, 1.
- E. L. Elson, *Biophys. J.*, 2011, **101**, 2855.
- Q. Wang and W. E. Moerner, *ACS Nano*, 2011, **5**, 5792.
- Q. Wang and W. Moerner, *Appl. Phys. B: Lasers Opt.*, 2010, **99**, 23.
- V. Levi, Q. Ruan, K. Kis-Petikova and E. Gratton, *Biochem. Soc. Trans.*, 2003, **31**, 997.
- M. Speidel, A. Jonáš and E.-L. Florin, *Opt. Lett.*, 2003, **28**, 69.
- A. J. Berglund and H. Mabuchi, *Appl. Phys. B: Lasers Opt.*, 2004, **78**, 653.
- H. Cang, C. S. Xu, D. Montiel and H. Yang, *Opt. Lett.*, 2007, **32**, 2729–2731.
- J. G. Ritter, R. Veith, J.-P. Siebrasse and U. Kubitschek, *Opt. Express*, 2008, **16**, 7142.
- S. R. P. Pavani, M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun and W. E. Moerner, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 2995.
- M. A. Thompson, M. D. Lew, M. Badieirostami and W. E. Moerner, *Nano Lett.*, 2010, **10**, 211.
- B. Schulz, D. Täubner, J. Schuster, T. Baumgärtel and C. v. Borczyskowski, *Soft Matter*, 2011, **7**, 7431.
- J. Schuster, F. Cichos and C. v. Borczyskowski, *Eur. Phys. J. E: Soft Matter Biol. Phys.*, 2003, **12**, 019.
- H. Qian, M. P. Sheetz and E. L. Elson, *Biophys. J.*, 1991, **60**, 910.
- A. Einstein, *Ann. Phys. (Berlin, Ger.)*, 1905, **322**, 549.
- M. Goulian and S. M. Simon, *Biophys. J.*, 2000, **79**, 2188.
- D. S. Martin, M. B. Forstner and J. A. Käs, *Biophys. J.*, 2002, **83**, 2109.
- X. Michalet, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2010, **82**, 041914.
- M. J. Saxton, *Biophys. J.*, 1997, **72**, 1744.
- X. Michalet and A. J. Berglund, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2012, **85**, 061916.
- D. Ernst, S. Hain and J. Köhler, *J. Opt. Soc. Am. A*, 2012, **29**, 1277.