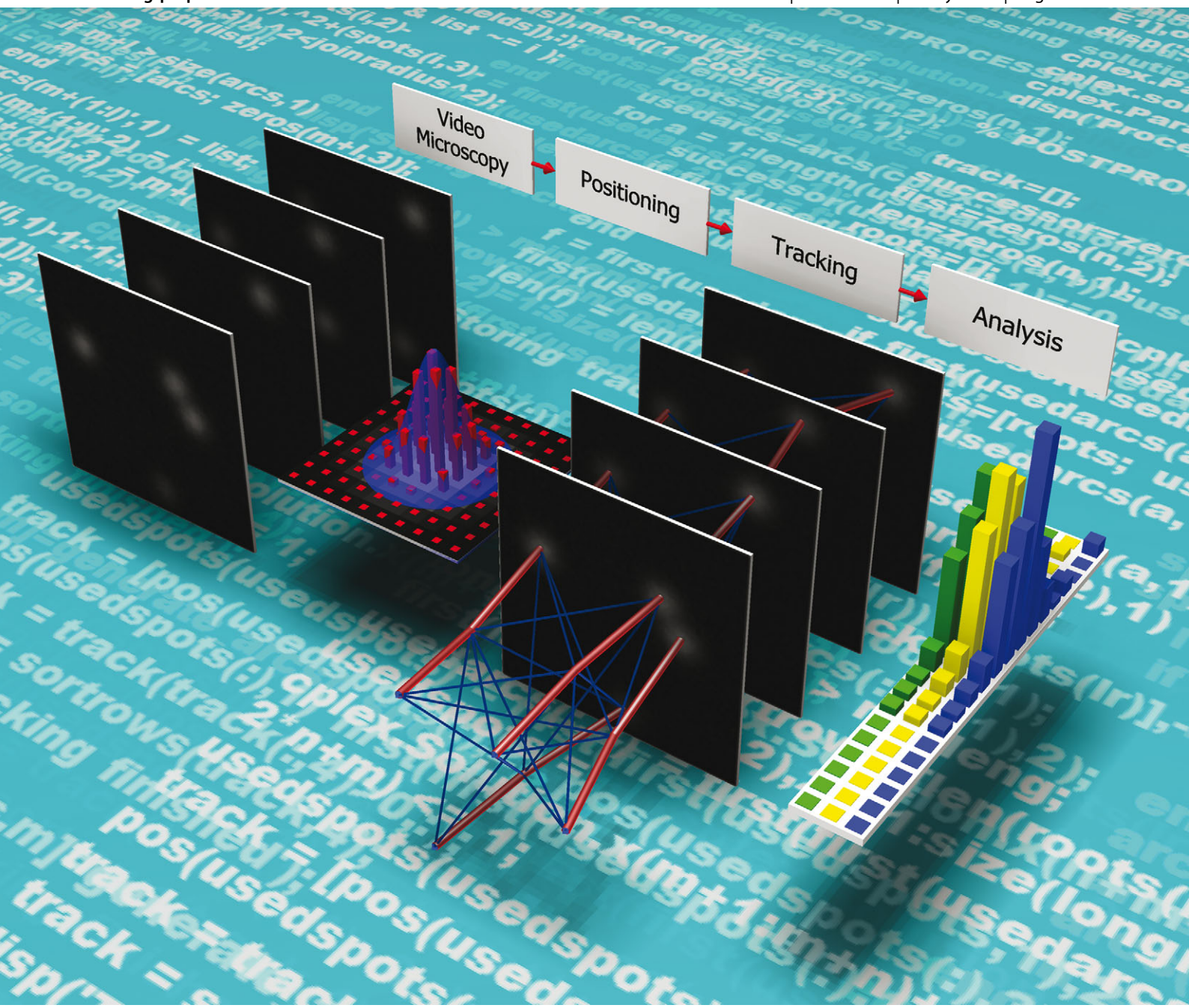
PCCP

Physical Chemistry Chemical Physics

www.rsc.org/pccp

Volume 15 | Number 17 | 7 May 2013 | Pages 6151–6550



ISSN 1463-9076

PAPER

Wöll, Karrenbauer *et al.*

A novel method for automatic single molecule tracking of blinking molecules at low intensities

PCCP

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2013, **15**, 6196

A novel method for automatic single molecule tracking of blinking molecules at low intensities†

Dominik Wöll,*[ab] Christoph Kölbl,[b] Beate Stempfle[b] and Andreas Karrenbauer*[acd]

Received 28th December 2012,
Accepted 6th February 2013

DOI: 10.1039/c3cp44693j

www.rsc.org/pccp

Single molecule tracking provides unprecedented insights into diffusional processes of systems in life and material sciences. Determination of molecule positions with high accuracy and correct connection of the determined positions to tracks is a challenging task with, so far, no universal solution for single fluorescing molecules tackling the challenge of low signal-to-noise ratios, frequent blinking and photo bleaching. Thus, the development of novel algorithms for automatic single molecule fluorescence tracking is essential to analyse the huge amount of diffusional data obtained with single molecule widefield fluorescence microscopy. Here, we present a novel tracking model using a top-down polyhedral approach which can be implemented effectively using standard linear programming solvers. The results of our tracking approach are compared to the ground truth of simulated data with different diffusion coefficients, signal-to-noise ratios and particle densities. We also determine the dependency of blinking on the analysed distribution of diffusion coefficients. To confirm the functionality of our tracking method, the results of automatic tracking and manual tracking by a human expert are compared and discussed.

Introduction

The possibility of observing single fluorescent molecules in real time and *in situ* has provided a lot of new insights into the dynamics of systems in biology and material sciences. Single molecule microscopy (SMM) allows for the parallel observation of translational and rotational motion of many single fluorescent molecules if their concentration is reasonable low, so that the signals of these emitters are separated further than the diffraction limit. In contrast to tracing single colloids with video microscopy or following the motion of single highly labelled particles, viruses or antibodies using widefield fluorescence microscopy,[1–3] tracking of single fluorescent molecules bears the challenge that the fluorescent spots are rather weak with a low signal-to-noise (S/N)-ratio and show significant changes in signal intensity.[4–6] These changes are caused by reorientation,[7,8] changes in fluorescence quantum yield due to alterations in the molecule environment[9,10] enabling reversible electron transfer reactions,[11] intersystem crossing to dark triplet states[12] and/or out-of-focus motion and a concomitant change in the single molecule point spread function.[13] In the extreme case, a fluorescent molecule is dark for several recorded frames, a phenomenon which is termed blinking.[14–16] Furthermore, single fluorophores photo bleach,[17] *i.e.* they are irreversibly destroyed under the illumination conditions used. Thus, only a limited number of photons per molecule is emitted. One challenge of single molecule fluorescence measurements consists of finding a suitable way to distribute these photons in time to obtain the best possible information on the system under investigation. For very good fluorophores several millions of photons can be emitted before photo bleaching occurs[18] which means that molecules can be observed for several seconds up to minutes at the typical irradiation conditions (1–10 kW m^{-2}). The fact that the fluorescence signals of single molecules cannot be classified according to their intensity or shape in different frames and even disappear in some frames causes severe problems in single molecule tracking. Thus, many tracking algorithms which have been developed for single particle tracking in video microscopy fail in tracking single molecules.

[a] *Zukunftskolleg, University of Konstanz, Universitätsstr. 10, 78464 Konstanz, Germany*
[b] *Department of Chemistry, University of Konstanz, Universitätsstr. 10, 78464 Konstanz, Germany. E-mail: dominik.woell@uni-konstanz.de; Tel: +49 7531 88 2024*
[c] *Department of Computer and Information Science, University of Konstanz, Universitätsstr. 10, 78464 Konstanz, Germany*
[d] *Max Planck Institute for Informatics, Campus E1 4, 66123 Saarbrücken, Germany. E-mail: andreas.karrenbauer@mpi-inf.mpg.de; Tel: +49 681 9325 1058*

† Electronic supplementary information (ESI) available: Open source tracking code written in Matlab, examples for input and output of Matlab routine, distribution of analyzed diffusion coefficients, correlation between ground truth and determined diffusion coefficients, time traces of simulated blinking molecules, jumps observed for tracking with a high maximum step length, comparison between Gaussian fit and center of mass fit, rationalization of the chosen cost function. See DOI: 10.1039/c3cp44693j

The way from the recorded single molecule microscopy movies to the results about their motion includes typically the following steps: (i) exact determination of the positions of each fluorescent molecule, (ii) reasonably connecting the positions to single molecule tracks and (iii) statistical analysis of these tracks.

Already before the first optical observations of single chromophores,[19,20] procedures have been described to optimize accuracy of positioning polymer beads in differential interference contrast (DIC) light-microscopy images with nm accuracy.[21] Nowadays, single molecule positions are most often determined using centre of mass or Gaussian fits, with the latter being the best option for low signal-to-noise levels of around 4.[22] The dependence of localization accuracy on different parameters has been discussed by Thompson *et al.*[23] and further modified by Mortensen *et al.*[24] Further optimization of the localization procedure can yield molecule positions down to the nm-range and beyond.[25] However, for single chromophores with fixed transition dipole moment, it has to be taken into account that polarization effects due to chromophore orientation limit the resolution to *ca.* 10 nm for high numerical apertures.[26] High localization accuracies are also essential for superresolution techniques based on photo bleaching[27,28] or stochastic on- and off-switching of fluorescence such as PALM,[29] STORM,[30] dSTORM,[31,32] *etc.*

One approach to analyse the positioning data is Image Correlation Analysis[33] which can give valuable information about the mobility of molecules. In particular for high chromophore densities which are beyond the capabilities of single molecule tracking these techniques give valuable information.[34] One of the main limitations, namely the recording speed of the CCD cameras used, has been recently overcome using two colour-labelling. This allowed the authors for a parallel, albeit stepwise detection of fast single molecule motion.[35]

However, single molecule tracks give more information than correlations especially in systems with spatial and/or dynamic heterogeneities.[36,37] Thus, after localization, the positions of subsequent frames have to be connected to tracks.[38] Different approaches have been developed for this purpose, but up to now it remains challenging to improve and develop algorithms which can be used not only for special tasks but also for a universal set of problems.[39]

Single particle tracking procedures started with the connection of one point with its closed neighbour in consecutive frames.[40] For each trajectory constructed in this way, the mean squared displacement MSD(t) can be computed.[41,42] In 1999, Chetverikov and Verestoy published a new algorithm called *IPAN Tracker*.[43] Using a competitive linking process that develops as the trajectories grow, this algorithm deals better with incomplete trajectories, high spot densities, faster moving particles and appearing and disappearing spots. Sbalzarini and Koumoutsakos used the same approach as the *IPAN tracker*, but did not make any assumptions about the smoothness of trajectories.[44] Their algorithm was implemented as *ParticleTracker* in *ImageJ*. The *SpotTracker*[45] is a very powerful tool to follow single spots throughout one movie, but it can only proceed spot by spot. The algorithm proposed by Bonneau *et al.*[46] falls into the same category of greedy algorithms that iteratively compute shortest paths in space-time, which are not revised subsequently. One of the most accurate solutions to SPT is provided by multiple-hypothesis tracking (MHT). This method chooses the largest no conflicting ensemble of single molecule paths simultaneously accounting for all positions in each frame. Jaqaman *et al.* used such an approach where they first linked positions in consecutive frames and combined these links into entire trajectories.[47] Both steps were globally optimized which yielded a very likely solution to the tracking problem. Dynamic multiple-target tracing was used by Sergé *et al.* to generate dynamic maps of tracked molecules at high density.[48] Subtracting detected peaks from the images allows them for a detection of low intensity peaks which would be otherwise hidden in movies of high particle density. Peak positions were connected using statistical information from past trajectories. Yoon *et al.* describe a Bayesian-based inference approach to track low intensity and blinking molecules.[49] Recent reviews of methods for cell, particle and molecule tracking were published by Meijering *et al.*[50] and Yao *et al.*[51]

For the analysis of the tracks, different approaches have been developed.[39,52] The most common approach is analysis of the mean square displacement at different time intervals[3,53–55] which can readily distinguish between different modes of motion such as normal diffusion, anomalous diffusion, confined diffusion, drift and active transport.[42] Alternatively, probability distribution of squared displacements[56] and radii of gyration[52,57,58] can be used to analyse single molecule tracks. The latter method is especially powerful to distinguish between mobile and immobile molecules.[59]

For all single molecule tracking algorithms, it is important to be aware of the dependence of single molecule diffusion coefficients on the track length.[60] The longer the track length, the more reliable is the obtained diffusion coefficient. In our and others experience, a trajectory of at least 20 points is required to obtain reasonable results.[61]

In this paper, we present a method for tracking individual molecules which globally optimizes the likelihood of the connections between molecule positions fast and with high reliability even for high spot densities and blinking molecules. Our method uses cost functions which can be freely chosen to combine costs for distances between spots in space and time and which can account for the reliability of positioning a molecule. To this end, we describe a top-down polyhedral approach to the problem of tracking many individual molecules. This immediately yields an effective implementation using standard linear programming solvers. Furthermore, we derive an unbiased estimator for diffusion coefficients of blinking molecules, which proves well in the experimental evaluation of our approach.

Methods

Generation of synthetic trajectories

Synthetic trajectories were generated to test our localization and tracking algorithms. Random walk simulations with 500 steps were performed using four different diffusion coefficients

(10^{-15}, 10^{-14}, 10^{-13} and 10^{-12} m^2 s^{-1}), five different signal-to-noise ratios (1, 2, 3, 4 and 5) and five different particle densities (100, 200, 300, 400 and 500 particles per frame). These values can be found in many tracking challenges of practical importance, but were also chosen to determine the sensitivity of our algorithms. The initial x- and y-position of each spot was chosen uniformly at random in the range [0; 500]. The single molecule trajectories were used to construct movies with 500 frames. Two dimensional Gaussian functions were constructed with their centre located at the positions obtained from the random walk simulations and their widths being diffraction limited (*ca.* 300 nm). The time between consecutive frames was chosen as 0.1 s and the resolution as 100 nm per pixel. Gaussian white noise was added to the movies corresponding to the signal-to-noise ratio defined as $\mathrm{SNR} = \frac{A_{\mathrm{signal}}}{\sigma_{\mathrm{noise}}}$ with the amplitude A_{signal} of the centre of the 2D Gaussian and the standard deviation σ_{noise} of the Gaussian white noise.

Positioning of single molecules

The positioning procedure was empirically optimized for spots with low intensity as they are commonly obtained in single molecule microscopy. Fig. 1 visualizes the steps of this procedure for a simulated image with spots of signal-to-noise ratio 5. The original image 1 is filtered with two different average filters $n \times n$ and $m \times m$, respectively, where $m > n$ resulting in the filtered images 2 and 3, respectively. Subtraction of 2 from 3 resulted in image 4, to which another 3×3 average filter was applied to obtain 5 which was subsequently binarized, *i.e.* pixels with a value above a fixed, but arbitrary threshold were set to 1, others to 0. If at least 6 pixels (number arbitrary) can be found directly connected to each other, their centre is taken as the initial position of the spot. The exact spot position is calculated within the original image 1, starting from the above mentioned initial position and using either a centre of mass or a Gaussian fitting algorithm, respectively.[22] The strength of our filtering procedures lies in the recognition of closely lying spots. The two spots in close vicinity in Fig. 1 are better distinguished in image 5 where filtering results in a groove around spots compared to image 2 where significant intensity can be found also between spots. Using our positioning procedure, the positions assigned to the spots in image 6 are very close to the simulated positions.

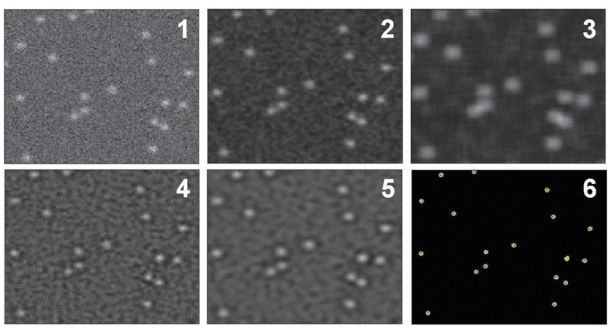


Fig. 1 Visualization of the procedure used for single molecule positioning.

Tracking algorithm

Our tracking procedure is based on a linear programming formulation and described in detail elsewhere.[62] A compact description of all tracking possibilities by linear equations and linear inequalities is derived, which allows for computing the one with the maximum likelihood. To this end, each pair of spots that may be consecutive in a trajectory is equipped with a cost-value, which describes the likelihood of the corresponding displacement.

Moreover, in order to avoid significant fragmentation of tracks, costs are assigned to each spot for being the first or the last in a trajectory. Each of the decisions, *i.e.* to start a trajectory with a spot, to continue a trajectory with a spot, and to end a trajectory at a spot, is modelled by a binary variable. Using the Markov property that a transition only depends on the state itself and not on the history that led to it, the cost-function is linear in these binary variables. A feasible solution to the considered optimization problem is subject to the constraints that

(i) a trajectory starts at a unique spot,
(ii) a trajectory ends at a unique spot, and
(iii) each spot belongs to at most one trajectory.

Since these constraints can be modelled by linear equations, a so-called linear programming problem (LP) is obtained, which may be solved efficiently with standard software, *e.g.* CPLEX. It is worth to emphasize that it is sufficient to constrain the variables to the interval [0;1] to obtain a binary optimum solution. A MATLAB script is published as Open Source, which was used to compute the results of this paper using the MATLAB connector of CPLEX as a black box.

Parameters for tracking

As cost-function, we used '$\Delta x^2 + \Delta y^2 + \Delta t^2$' with the time Δt between consecutive positionings and Δx and Δy being the changes of the spatial coordinates x and y within the time interval Δt. The rationale for using this function which includes a superlinear Δt^2 term is based on the following observation: if time is not penalized, then track fragmentation becomes more likely as shown in Fig. S19 and in the two movies in the ESI.†

It is important to note that our implementation allows for arbitrary functions on spatial (Δx, Δy) and temporal (Δt) distance of positionings and also enables us to include the localization accuracy of a spot in its costs for being tracked or not. Additionally, we provide a further parameter called the join-radius, which allows for merging two spots in the same frame within this radius to remedy the influence of duplicated positions belonging to the same spot due to noise. This feature may be simply turned off by setting the radius to 0.

Results and discussion

Positioning of single molecules

The positioning accuracy of our procedure was tested for different sets of simulated images with varying signal-to-noise (S/N) ratios (see also Methods section). The number of molecules in the movies was varied from 100 to 500. Fig. 2 gives an impression of the spot densities used for our simulations.

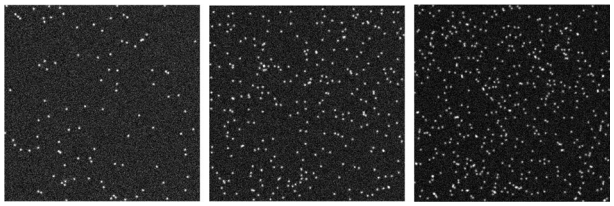


Fig. 2 Examples for simulated images with S/N-ratio 5 and (from left to right) 100, 300 and 500 spots.

Typical images with different S/N-ratios are shown in Fig. 3. The localization accuracy of spots depends on the S/N-ratio. The distribution of the distance between simulated and localized positions of approximately 104 spots for different S/N-ratios and number of molecules is shown in Fig. 4. Low S/N-ratios result in a significant broadening of distributions and thus low localization accuracy. The number of molecules has only a small influence on the distributions which become slightly broader with increasing molecule density due to the poorer localization of positions with nearby spots.

Determination of diffusion coefficients from track positions

The diffusion coefficients were determined from single molecule tracks using track radius analysis (also known as radius of gyration analysis).[52,57–59] The advantage of this method lies in the high accuracy of the obtained diffusion coefficients which are robust even for small trajectories of approx. 10 frames.

In the following, we show how to obtain an unbiased estimator for the diffusion coefficient in a normal diffusion model. Therefore, we consider position updates of the form

$$x_i = x_{i-1} + s_i n_i \quad \text{for } i > 1$$

where n_i is a standard normal random variable (with mean value $\mu = 0$ and standard deviation $\sigma = 1$) and the step-lengths s_i are dependent on the diffusion coefficient and the time between the observations. Hence, the coordinates x_i are Gaussian random variables with an expectation value of

$$E[x_i] = x_1 + \sum_{k=2}^{i} s_k E[n_k] = x_1$$

$$\begin{aligned} E[x_i x_j] &= {x_1}^2 + x_1 \sum_{k=2}^{i} s_k E[n_k] + x_1 \sum_{k=2}^{j} s_k E[n_k] \\ &\quad + \sum_{k=2}^{i} \sum_{l=2}^{j} s_k s_l E[n_k n_l] \\ &= {x_1}^2 + \sum_{k=2}^{\min\{i,j\}} {s_k}^2 \end{aligned}$$

In the case of normal diffusion, we have ${s_k}^2 = 2D(t_k - t_{k-1})$, where t_k denotes the time of the kth observation. Thus, we may write

$$E[x_i x_j] = {x_1}^2 + 2D(t_{\min\{i,j\}} - t_1).$$

We use the unbiased estimator for the variance of the x-coordinates of a track $x_1,\ldots,x_N$, which is determined as

$$\begin{aligned} {\hat{\sigma}_x}^2 &:= \frac{1}{N-1} \sum_{i=1}^{N} \left(x_i - \frac{1}{N} \sum_{k=1}^{N} x_k \right)^2 = \frac{1}{N-1} \sum_{i=1}^{N} {x_i}^2 \\ &\quad - \frac{2}{N(N-1)} \sum_{i=1}^{N} x_i \sum_{k=1}^{N} x_k + \frac{1}{N(N-1)} \left(\sum_{k=1}^{N} x_k \right)^2 \\ &= \frac{1}{N-1} \sum_{i=1}^{N} {x_i}^2 - \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \end{aligned}$$

In expectation, we get

$$\begin{aligned} (N-1) \cdot E\left[{\hat{\sigma}_x}^2\right] &= \sum_{i=1}^{N} E\left[{x_i}^2\right] - \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} E[x_i x_j] \\ &= N{x_1}^2 + 2D \sum_{i=1}^{N} (t_i - t_1) - N{x_1}^2 \\ &\quad - \frac{2D}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(t_{\min\{i,j\}} - t_1 \right) \\ &= 2D \sum_{i=2}^{N} (t_i - t_1) - \frac{2D}{N} \sum_{i=1}^{N} (2N - 2i + 1)(t_i - t_1) \\ &= \frac{2D}{N} \sum_{i=2}^{N} (2i - 1 - N)(t_i - t_1). \end{aligned}$$

Because the random displacements of the x- and y-coordinates are independent, we obtain the same for $E[\hat{\sigma}_y]$. This motivates the definition of

$$\hat{D} := \frac{N(N-1)}{4} \frac{{\hat{\sigma}_x}^2 + {\hat{\sigma}_y}^2}{\sum_{i=2}^{N} (2i - 1 - N)(t_i - t_1)},$$

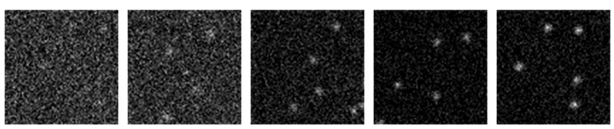


Fig. 3 Examples for simulated spots with S/N-ratio 1, 2, 3, 4 and 5 (from left to right).

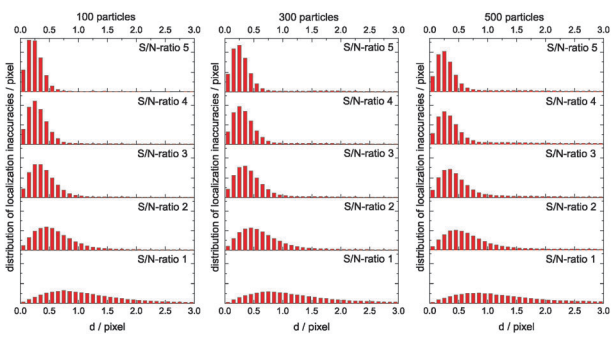


Fig. 4 Localization accuracy for different signal-to-noise ratios and particle densities.

which is an unbiased estimator for the diffusion coefficient because $E[\hat{D}] = D$. For the special case of uniform time steps, we substitute $t_i = i\tau$ and obtain

$$\hat{D} := \frac{3}{2\tau}\frac{\hat{\sigma}_x{}^2 + \hat{\sigma}_y{}^2}{N+1} = \frac{3}{2\tau}\frac{N}{N^2-1}R_g{}^2$$

with the track radius R_g as described in previous papers:[52,59]

$$R_g{}^2 = tr\begin{pmatrix} \frac{1}{N}\sum_{j=1}^{N}(x_j - \langle x\rangle)^2 & \frac{1}{N}\sum_{j=1}^{N}(x_j - \langle x\rangle)(y_j - \langle y\rangle) \\ \frac{1}{N}\sum_{j=1}^{N}(x_j - \langle x\rangle)(y_j - \langle y\rangle) & \frac{1}{N}\sum_{j=1}^{N}(y_j - \langle y\rangle)^2 \end{pmatrix}$$

$$= \frac{1}{N}\left(\sum_{j=1}^{N}(x_j - \langle x\rangle)^2 + \sum_{j=1}^{N}(y_j - \langle y\rangle)^2\right)$$

Effect of localization accuracy on determination of diffusion coefficients

Since the localization of the centres of the spots is subject to some uncertainty, we extend our model as follows. We define $\tilde{x}_i := x_i + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0,\varepsilon^2)$ are independent normally distributed random variables with variance ε^2. Hence, we obtain

$$E[\tilde{x}_i] = E[x_1]$$

$$E[\tilde{x}_i\tilde{x}_j] = E[x_ix_j] + E[\varepsilon_i] \times E[x_j] + E[x_i] \times E[\varepsilon_j] + E[\varepsilon_i\varepsilon_j]$$

$$= E[x_ix_j] + \varepsilon^2\delta_{ij}$$

with Kronecker's $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$. Since we can estimate the track variances only based on the measured data x_i, we get in expectation

$$(N-1)\cdot E\left[\hat{\sigma}_x{}^2\right] = \sum_{i=1}^{N}E\left[\tilde{x}_i{}^2\right] - \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\tilde{x}_i\tilde{x}_j\right]$$

$$= \sum_{i=1}^{N}E\left[x_i{}^2\right] + N\varepsilon^2 - \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[x_ix_j\right] - \varepsilon^2$$

$$= (N-1)\varepsilon^2 + \frac{2D}{N}\sum_{i=2}^{N}(2i-1-N)(t_i - t_1)$$

Hence, we now have a bias in the estimator $\hat{D}$ as defined above depending on ε:

$$E[\hat{D}] = D + \frac{N(N-1)\varepsilon^2}{2\sum_{i=2}^{N}(2i-1-N)(t_i - t_1)}$$

Thus, uniform time steps yield

$$E[\hat{D}] = D + \frac{3\varepsilon^2}{\tau(N+1)}$$

Determination of tracking accuracy for simulated data

Tracking procedures have to meet several conditions to be suitable. Apart from practical aspects such as tracking speed and memory consumption, the number of false positives is the key factor which has to be minimized in order to obtain reliable results.

False positives are connections between positions in different frames which have been set even though the positions do not belong to the same molecule/particle. They can result in severe errors in single molecule tracking and cause wrong interpretations of collected data. Thus, the number of false positives should be kept as low as possible.

False negatives are connections between positions in different frames which are not recognized by the tracking algorithm. One challenge is the simultaneous analysis of fast and slow particles since tracking of slow particles is more reliable than tracking of fast particles which can move in and out of focus and the trajectories of which cross each other more frequently.

The fraction of false positives and false negatives for movies of different diffusion coefficients and S/N-ratios was calculated comparing ground truth and analysed connections between points. The results are shown in Fig. 5. The fraction of false positive connections depends on the S/N-ratio. For tracking with a maximum step length of 5 pixels (500 nm), the number of false positives is approx. 13% at a very low S/N-ratio of 1, whereas it decreases to approx. 3% for S/N-ratios between 3 and 5. A longer maximum step length of 15 pixels increases the number of false positives due to an increased amount of possible connections. Otherwise, the fraction of false positive connections does not significantly depend on the diffusion coefficient of the particles. A stronger variation is found in the fraction of false negatives. Due to the large displacements of molecules with diffusion coefficients of 10^{-12} m^2 s^{-1} many connections were not detected with a maximal displacement of 5 pixels (see open symbols in Fig. 5 (right)). For lower diffusion coefficients the fraction of false negatives decreases from approx. 65% for a S/N-ratio of 1 to approx. 15% for a S/N-ratio of 5 and does not significantly depend on the number of molecules.

The maximum step length allowed for tracking is a critical parameter in particular for particles with high diffusion coefficients. Setting the maximum step length too low will cause an increase in the number of false negatives. On the other hand, many false positive connections will be tracked for a large maximum step length. Thus, it is very important to choose a reasonable value which can be estimated from the following

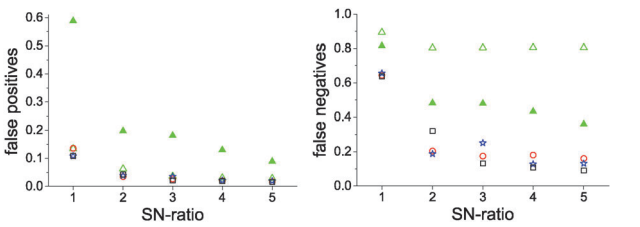


Fig. 5 Fraction of false positives (left) and false negatives (right) after automatic single molecule positioning and tracking for different S/N-ratios. The simulated diffusion coefficients are 10^{-12} m^2 s^{-1} (green triangles), 10^{-13} m^2 s^{-1} (red circles), 10^{-14} m^2 s^{-1} (blue stars) and 10^{-15} m^2 s^{-1} (black squares). For the diffusion coefficient of 10^{-12} m^2 s^{-1}, a tracking radius of 5 pixels (open triangles) was compared to a tracking radius of 15 pixels (full triangle). All data points represent the averages of movies with 100, 200, 300, 400 and 500 simulated particles.

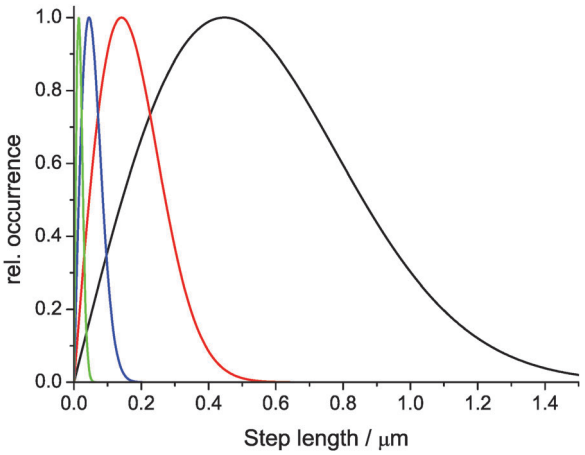


Fig. 6 Step length distributions for diffusion coefficients of 10^{-12} m^2 s^{-1} (black), 10^{-13} m^2 s^{-1} (red), 10^{-14} m^2 s^{-1} (blue) and 10^{-15} m^2 s^{-1} (green) and a time step of 0.1 s according to eqn (1).

considerations. Fig. 6 shows the distribution of step lengths for the four different diffusion coefficients used in this study. The distribution can be expressed as[63]

$$P(r,t) = \frac{1}{\pi 4Dt}\exp\left(-\frac{r^2}{4Dt}\right)2\pi r \qquad (1)$$

It follows that the probability of finding a step length within a radius of R is

$$W(R,t) = \int_0^R P(r,t)\mathrm{d}r = 1 - \exp\left(-\frac{R^2}{4Dt}\right) \qquad (2)$$

For our simulated integration time of 0.1 s, eqn (2) can be used to show that 99% of the step lengths are within 1.357 μm for a diffusion coefficient of 10^{-12} m^2 s^{-1}, within 0.429 μm for 10^{-13} m^2 s^{-1}, within 0.136 μm for 10^{-14} m^2 s^{-1} and within 0.043 μm for 10^{-15} m^2 s^{-1}. Thus setting the step length for tracking to 5 pixels (0.5 μm) is suitable for tracking molecules with a diffusion coefficient below 10^{-13} m^2 s^{-1}. The tracks of faster molecules, however, cannot reliably be found anymore because the number of false negatives is too high (see Fig. 5 (right)). Increasing the maximum allowed step length to 15 pixels = 1.5 μm (full symbols in Fig. 5), a distance which includes more than 99% of all step lengths for $D = 10^{-12}$ m^2 s^{-1}, decreases the number of false negatives, but simultaneously the number of false positives increases due to the possibility to connect positions which are further apart, in particular for movies with low S/N-ratio.

Analysis of all steps essential to determine diffusion coefficients

Positioning and tracking are the key steps in the determination of diffusion coefficients. To distinguish the errors appearing in these steps, we analysed four different cases shown in Fig. 7. First, in case 1, the distribution of diffusion coefficient as directly calculated from the ground truth tracks was determined. The upper left graph in Fig. 8 shows the distribution of the logarithm of diffusion coefficients due to the finite number of sample points in the corresponding random walks.

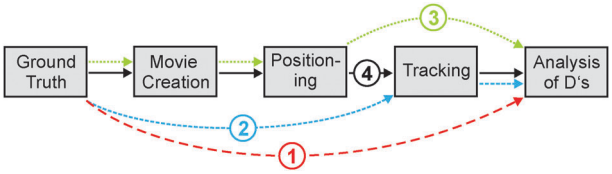


Fig. 7 Steps taken from the simulated ground truth data to the analysis of the trajectories. For analysis of the accuracy of each single step, different ways were used for analysis, which are labeled 1 to 4.

All distributions resemble a Gaussian shape with a width of almost one decade and their maximum at the diffusion coefficient used for the simulations.

In a second set of analyses (case 2, Fig. 8) the same ground truth positions were tracked using our polyhedral model solved with CPLEX. The distribution of diffusion coefficients is very similar to case 1, with the exception of the molecules with $D \geq 10^{-13}$ m^2 s^{-1}. For these velocities, the maximum distance typically set to 5 pixels resulted in a distribution shifted to lower diffusion coefficients due to the fact that at $D \geq 10^{-13}$ m^2 s^{-1} average displacements between frames are getting close to the set value of 5 pixels and thus the simulated tracks are chopped into many short tracks by the tracking method. Changing the maximum distance to 15 pixels resulted in a high correlation between tracked and ground truth data for $D = 10^{-12}$ m^2 s^{-1}. For $D > 10^{-13}$ m^2 s^{-1}, reasonable correlations are only possible for sufficiently low spot densities. For high spot densities the probability of positions of other tracks moving into the tracking range of another molecule is too high, and thus the tracking algorithm in general returns diffusion coefficients lower than the real value.

The third set of analyses, case 3, allows for an investigation of the influence of positioning inaccuracies on the distribution of diffusion coefficients. Movies were constructed from the ground truth positions with different S/N-ratios. Our positioning algorithm was applied to these movies and, where possible, the positions matched to the positions of the ground truth tracks. With the determined positions of each track, a diffusion coefficient was obtained. The distribution of diffusion coefficients of these tracks resembles the distributions of the ground truth tracks with the exception of low diffusion coefficients with low S/N-ratios where the poor localization accuracy results in a seemingly higher diffusion coefficient than simulated.

The localization accuracy determines the lowest diffusion coefficient which can be determined by the corresponding experimental settings. For an extended analysis of the localization accuracy on the distribution of diffusion coefficients, we added a Gaussian noise with standard deviation $\sigma = 0.5$ or $\sigma = 1$ pixel, respectively, to the ground truth data. The analysis of the obtained tracks (directly from the perturbed ground truth without tracking) is shown in the ESI† (Fig. S4 and S5). A localization inaccuracy of 0.5 pixel (50 nm) causes the analysed diffusion coefficients to be too high for $D = 10^{-15}$ m^2 s^{-1}. Diffusion coefficients above 10^{-14} m^2 s^{-1} are perfectly obtained. If the localization inaccuracy is increased to 1.0 pixel (100 nm), the distribution obtained for molecules with

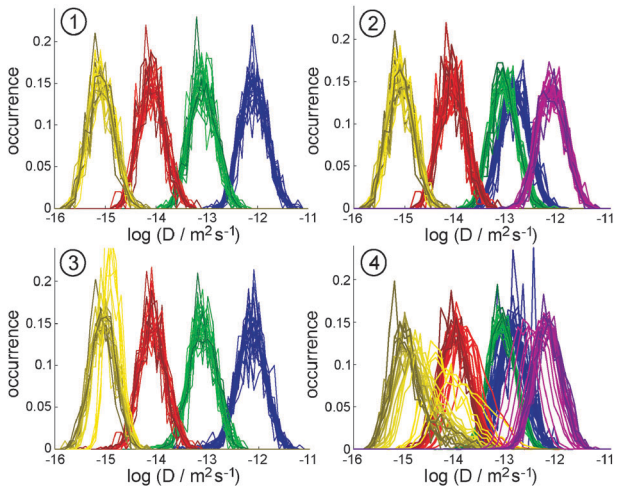


Fig. 8 Distributions of analysed diffusion coefficients for the simulated diffusion coefficients 10^{-15} (yellow), 10^{-14} (red), 10^{-13} (green), 10^{-12} m^2 s^{-1} (violet for a maximum step length of 15 or blue for a maximum step length of 5) and S/N-ratios encoded in the brightness of colors (bright = S/N-ratio 1; dark = S/N-ratio 5) for the four different cases shown in Fig. 7.

simulated D of 10^{-15} m^2 s^{-1} is significantly wrong and also analysis of $D = 10^{-14}$ m^2 s^{-1} becomes less reliable.

Case 4 describes the procedure which is applied for real movies to determine single molecule diffusion coefficients. Spots in movies are positioned, the positions tracked and a diffusion coefficient calculated from these tracks. For high S/N-ratios the obtained distributions are similar to the simulated distributions. Two trends can be observed in particular at low S/N-ratios: (i) molecules with low diffusion coefficients tend to be analysed as being faster as they were simulated and (ii) analysis of the motion of fast molecules in average results in a lower diffusion coefficient as the ground truth data. The former observation can be explained by the poorer localization accuracy at low S/N-ratios. This inaccuracy resembles diffusion and thus a low diffusion coefficient will be assigned even to immobile molecules, just as discussed in the previous paragraph and section 'effect of localization accuracy on determination of diffusion coefficients'. In the case of fast motion, the probability of trajectories crossing each other and moving in and out of focus is increased. This results in a certain likelihood of fragmentation of trajectories preferentially at large steps which are consequently excluded from the calculation of the diffusion coefficients.

Tracking of blinking data

The fluorescence intensity of single molecules is typically not constant, but shows blinking behaviour due to photochemical or photophysical quenching processes[4] such as intersystem crossing to the triplet state[9,64,65] or electron transfer.[10]

The lengths of on- and off-times typically show a power law distribution. In order to simulate blinking behaviour, we generated on- and off-times for our simulated tracks using the following procedure. At the beginning a molecule was uniformly at random set as on or off. The number of frames

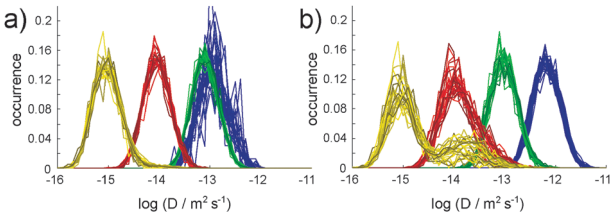


Fig. 9 Distributions of diffusion coefficients obtained after tracking of blinking data for a maximum step length of (a) 5 and (b) 15.

remaining in this state was determined randomly from a distribution so that the probability follows $P(t) = \left(\frac{t_{on/off}}{\tau}\right)^{-\alpha}$. The values for τ and α were set to reasonable values of $\tau = 1$ s and $\alpha = -2$, respectively. Four typical on- and off-time traces are presented in Fig. S16 of the ESI.†

The analysed distributions of diffusion coefficients for tracking of blinking ground truth data are shown in Fig. 9. For a maximum step length of 5 pixels, the distributions of diffusion coefficients resemble the ground truth data except for diffusion coefficients above 10^{-13} m^2 s^{-1} where, as also shown in Fig. 8, a step length of 5 pixels is not sufficient for tracking. Tracking with a maximum step length of 15 pixels gives good results for fast diffusing molecules, but for simulated diffusion coefficients of 10^{-14} m^2 s^{-1} and below results a long tail or even a second band at higher D values (see Fig. 9 (right)). The deviation of distribution from the ground truth distribution is caused by tracks which include at least one large jump from one molecule to another one which results in a significant increase of the track radius (see Fig. S17 in ESI†).

Comparison of the results from manual and automatic tracking

A comparison between the distributions obtained using manual and automatic tracking of single molecule trajectories, respectively, is presented in Fig. 10.

For our measurement parameters, good agreement between both ways of tracking is obtained for molecules with $\log(D/\text{m}^2\,\text{s}^{-1})$ between -14.5 and -13. For slower molecules, manual tracking results in a D-distribution shifted to lower diffusion coefficients. This is due to the fact that the human brain sorts out false positive positions which are in proximity of the track under observation and which in many cases of automatic tracking cause a (even though small) jump forward in one frame and backward in the next. Tracks including such jumps are assigned to too high diffusion coefficients. One suitable possibility to avoid such errors would be the identification and subsequent deletion of such events in an analysis step after tracking.

Trajectories of molecules moving faster than *ca.* 10^{-13} m^2 s^{-1} cross each other rather often resulting in an inherent ambiguity in connecting their positions. They often exhibit some rather large jumps which can be often recognized by our brain. The costs of these large jumps in our automatic tracking procedure are rather high and thus the tracking may jump to other molecules which occasionally appear to be closer in the

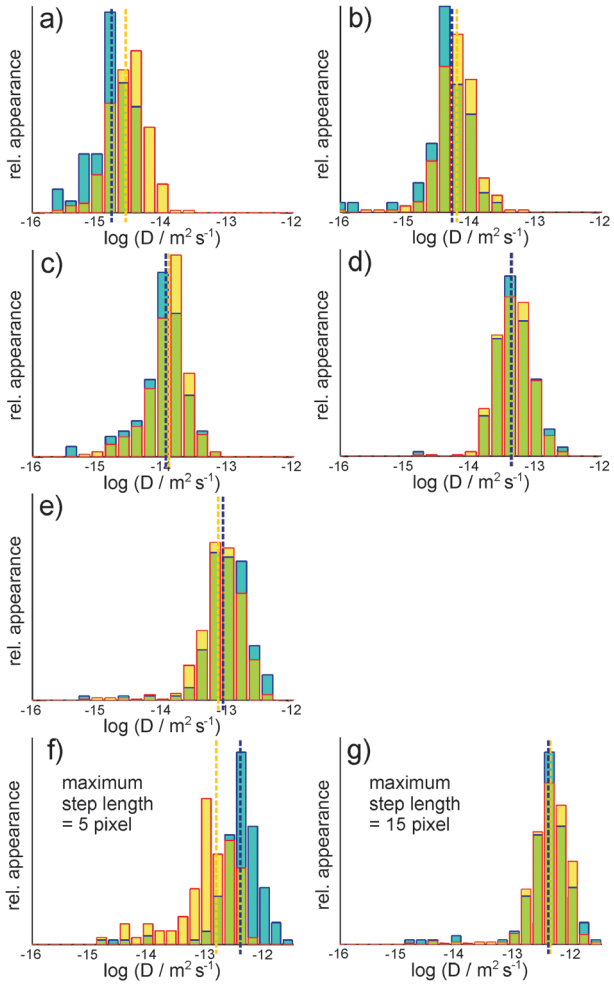


Fig. 10 Comparison of the distributions of diffusion coefficients for manual (cyan) and automatic (yellow) tracking. For automatic tracking a maximum step length of 5 pixels was used except for graph (g) where it was increased to 15 pixels to account for fast moving molecules. Data were obtained by tracking single, fluorescent perylene diimide molecules during the polymerization of styrene as described in detail by Stempfle *et al.*[59] Integration time was 0.029 s per frame, pixel size amounted to 53 nm and each frame contained approx. 100 spots.

respective frame. This is the main reason why the distribution of manually tracked fast moving molecules is shifted to higher *D*'s compared to our automatic procedure.

Possibilities to improve automatic tracking

For slow moving molecules, accurate tracking is restricted by localization accuracy. If higher precision is required, the S/N-ratio has to be improved to determine the molecule positions more accurately using 2D-Gaussian fitting. Since for widefield fluorescence microscopy measurements the noise is normally already kept at a minimum, the only possibility for higher accuracy is improving the signal. This is possible by either increasing the laser intensity or enhancing the integration time, thus collecting more photons per frame. This improve in signal intensity is, however, at the expense of earlier photo bleaching, and is only suitable when using very photostable dyes or particles. For very long integration times, mechanical stability of the sample has additionally to be assured or, if this is not possible, drift-correction has to be applied.

For tracking of molecules with rapid motion the integration time of the CCD camera should be set to a minimum (which at the current technical state-of-the-art is in the ms-range) in order to keep the distance molecules move from frame to frame within a few hundreds of nm. For these short integration times a rather high excitation light intensity is necessary to allow for a reasonable high S/N-ratio. However, as shown in the paper, localization accuracy is less important to determine diffusion coefficients of fast molecules. If necessary, improvements in localization accuracy are possible using stroboscopic illumination.[66] Furthermore, it is of considerable advantage to decrease the spot density, *i.e.* to reduce the concentration of fluorescent probes in order to minimize crossing of trajectories.

Further possible advances of the tracking algorithm are the implementation of elucidating tracking alternatives with only slightly higher costs. These alternatives could be presented to the human expert/user in order to decide which one is most likely. This approach could be even connected with learning algorithms to teach the computer how to process different situations.

We want to stress that our positioning and tracking method can be combined with the typical procedures of track analysis such as mean square displacement analysis and track radius analysis. Moreover, corrections to account for the inherent bias from averaging intensity over the entire integration time of the CCD camera may be incorporated seamlessly.[67] Even though we considered normal diffusion here, the method is not restricted to this type of motion. In our current implementation, we set one maximum step length for the whole movie. This might become problematic in heterogeneous media with a high concentration of particles and diffusion coefficients ranging over several orders of magnitude. However, such situations are detected easily. Due to the flexibility and efficiency of our linear programming approach, we could allow further or disallow incorrect connections between spots, and re-optimise without rendering our approach unpractical. This is subject to future work.

Conclusion

We present a suitable tracking method which, compared to other approaches, is especially powerful for tracking single fluorescent molecules, the signals of which often exhibit low S/N-ratios and frequent blinking. It performs a global optimization of connections between consecutive positions based on cost functions which can be adapted for different problems and which can also obtain information on the quality of positioning. Apart from the cost function, it only requires a minimum set of assumptions. The key parameter is the maximum allowed distance between consecutive positions which should be set appropriately in order to keep a good balance between false negative and false positive connections and to limit the amount of possibilities. Despite the still vast

amount of possibilities, our method has the advantage to be computationally feasible and fast, thus allowing typical single molecule tracking challenges to be solved within seconds using a typical personal computer or laptop. Tracking reliability is mainly depending on the quality of positioning, in particular the amount of false positive positions. We present our approach as open source code in Matlab using CPLEX as linear solver. Even though the method was shown for 2D tracking throughout this paper, it is easily adaptable to follow 3D motion.

Acknowledgements

D.W. and A.K. thank the Zukunftskolleg of the University of Konstanz for financial and administrative support and especially for its encouragement of interdisciplinary projects. A.K. is supported by the Max Planck Center for Visual Computing and Communication (www.mpc-vcc.org). Furthermore, we are grateful to Prof. Dr. Andreas Zumbusch and Franziska Rabold for fruitful discussions and testing of our algorithm.

Notes and references

1 R. Bausinger, K. von Gersdorff, K. Braeckmans, M. Ogris, E. Wagner, C. Bräuchle and A. Zumbusch, *Angew. Chem., Int. Ed.*, 2006, **45**, 1568–1572.
2 T. Kues, R. Peters and U. Kubitscheck, *Biophys. J.*, 2001, **80**, 2954–2967.
3 C. M. Anderson, G. N. Georgiou, I. E. G. Morrison, G. V. W. Stevenson and R. J. Cherry, *J. Cell Sci.*, 1992, **101**, 415–425.
4 F. Cichos, C. von Borczyskowski and M. Orrit, *Curr. Opin. Colloid Interface Sci.*, 2007, **12**, 272–284.
5 M. Lippitz, F. Kulzer and M. Orrit, *ChemPhysChem*, 2005, **6**, 770–789.
6 D. Bingemann, *Chem. Phys. Lett.*, 2006, **433**, 234–238.
7 S. Felekyan, R. Kühnemuth, V. Kudryavtsev, C. Sandhagen, W. Becker and C. A. M. Seidel, *Rev. Sci. Instrum.*, 2005, **76**, 083104.
8 M. Dorfschmid, K. Müllen, A. Zumbusch and D. Wöll, *Macromolecules*, 2010, **43**, 6174–6179.
9 J. Schuster, F. Cichos and C. von Borczyskowski, *Opt. Spectrosc.*, 2005, **98**, 712–717.
10 J. N. Clifford, T. D. M. Bell, P. Tinnefeld, M. Heilemann, S. M. Melnikov, J. Hotta, M. Sliwa, P. Dedecker, M. Sauer, J. Hofkens and E. K. L. Yeow, *J. Phys. Chem. B*, 2007, **111**, 6987–6991.
11 J. Hofkens, T. D. M. Bell, A. Stefan, E. Fron, K. Müllen and F. C. De Schryver, *Pure Appl. Chem.*, 2006, **78**, 2261–2266.
12 T. Cordes, J. Vogelsang and P. Tinnefeld, *J. Am. Chem. Soc.*, 2009, **131**, 5018–5019.
13 M. Böhmer and J. Enderlein, *J. Opt. Soc. Am. B*, 2003, **20**, 554–559.
14 W. Göhde, U. C. Fischer, H. Fuchs, J. Tittel, T. Basché, C. Bräuchle, A. Herrmann and K. Müllen, *J. Phys. Chem. A*, 1998, **102**, 9109–9116.
15 W. T. Yip, D. H. Hu, J. Yu, D. A. Vanden Bout and P. F. Barbara, *J. Phys. Chem. A*, 1998, **102**, 7564–7575.
16 M. Heidernätsch, M. Bauer, D. Täuber, G. Radons and C. von Borczyskowski, *Diffusion-fundamentals.org*, 2009, **11**, 111.
17 E. Füreder-Kitzmüller, J. Hesse, A. Ebner, H. J. Gruber and G. J. Schütz, *Chem. Phys. Lett.*, 2005, **404**, 13–18.
18 T. Weil, T. Vosch, J. Hofkens, K. Peneva and K. Müllen, *Angew. Chem., Int. Ed.*, 2010, **49**, 9068–9093.
19 W. E. Moerner and L. Kador, *Phys. Rev. Lett.*, 1989, **62**, 2535–2538.
20 M. Orrit and J. Bernard, *Phys. Rev. Lett.*, 1990, **65**, 2716–2719.
21 J. Gelles, B. J. Schnapp and M. P. Sheetz, *Nature*, 1988, **331**, 450–453.
22 M. K. Cheezum, W. F. Walker and W. H. Guilford, *Biophys. J.*, 2001, **81**, 2378–2388.
23 R. E. Thompson, D. R. Larson and W. W. Webb, *Biophys. J.*, 2002, **82**, 2775–2783.
24 K. I. Mortensen, L. S. Churchman, J. A. Spudich and H. Flyvbjerg, *Nat. Methods*, 2010, **7**, 377–383.
25 A. Pertsinidis, Y. X. Zhang and S. Chu, *Nature*, 2010, **466**, 647–653.
26 J. Enderlein, E. Toprak and P. R. Selvin, *Opt. Express*, 2006, **14**, 8111–8120.
27 X. H. Qu, D. Wu, L. Mets and N. F. Scherer, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11298–11303.
28 M. P. Gordon, T. Ha and P. R. Selvin, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 6462–6465.
29 E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz and H. F. Hess, *Science*, 2006, **313**, 1642–1645.
30 M. J. Rust, M. Bates and X. W. Zhuang, *Nat. Methods*, 2006, **3**, 793–795.
31 M. Heilemann, E. Margeat, R. Kasper, M. Sauer and P. Tinnefeld, *J. Am. Chem. Soc.*, 2005, **127**, 3801–3806.
32 M. Heilemann, S. van de Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld and M. Sauer, *Angew. Chem., Int. Ed.*, 2008, **47**, 6172–6176.
33 N. O. Petersen, P. L. Hoddelius, P. W. Wiseman, O. Seger and K. E. Magnusson, *Biophys. J.*, 1993, **65**, 1135–1146.
34 S. Semrau and T. Schmidt, *Biophys. J.*, 2007, **92**, 613–621.
35 S. Semrau, A. Pezzarossa and T. Schmidt, *Biophys. J.*, 2011, **100**, L19–L21.
36 D. Wöll, E. Braeken, A. Deres, F. De Schryver, H. Uji-i and J. Hofkens, *Chem. Soc. Rev.*, 2009, **38**, 313–328.
37 D. Wöll, H. Uji-i, T. Schnitzler, J. Hotta, P. Dedecker, A. Herrmann, F. C. De Schryver, K. Müllen and J. Hofkens, *Angew. Chem., Int. Ed.*, 2008, **47**, 783–787.
38 T. Schmidt, G. J. Schutz, W. Baumgartner, H. J. Gruber and H. Schindler, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 2926–2929.
39 M. J. Saxton, *Nat. Methods*, 2008, **5**, 671–672.
40 R. N. Ghosh and W. W. Webb, *Biophys. J.*, 1994, **66**, 1301–1318.
41 M. Goulian and S. M. Simon, *Biophys. J.*, 2000, **79**, 2188–2198.

42 M. J. Saxton and K. Jacobson, *Annu. Rev. Biophys. Biomol. Struct.*, 1997, **26**, 373–399.
43 D. Chetverikov and J. Verestoy, *Computing*, 1999, **62**, 321–338.
44 I. F. Sbalzarini and P. Koumoutsakos, *J. Struct. Biol.*, 2005, **151**, 182–195.
45 D. Sage, F. R. Neumann, F. Hediger, S. M. Gasser and M. Unser, *IEEE Trans. Image Process.*, 2005, **14**, 1372–1383.
46 S. Bonneau, M. Dahan and L. D. Cohen, *IEEE Trans. Image Process.*, 2005, **14**, 1384–1395.
47 K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid and G. Danuser, *Nat. Methods*, 2008, **5**, 695–702.
48 A. Sergé, N. Bertaux, H. Rigneault and D. Marguet, *Nat. Methods*, 2008, **5**, 687–694.
49 J. W. Yoon, A. Bruckbauer, W. J. Fitzgerald and D. Klenerman, *Biophys. J.*, 2008, **94**, 4932–4947.
50 E. Meijering, O. Dzyubachyk and I. Smal, in *Imaging and Spectroscopic Analysis of Living Cells: Optical and Spectroscopic Techniques*, ed. P. M. Conn, 2012, vol. 504, pp. 183–200.
51 C. Yao, J. Zhang, G. Wu and H. Zhang, *Comput. Math. Methods Med.*, 2012, 859398.
52 L. C. C. Elliott, M. Barhoum, J. M. Harris and P. W. Bohn, *Phys. Chem. Chem. Phys.*, 2011, **13**, 4326–4334.
53 C. Hellriegel, J. Kirstein and C. Bräuchle, *New J. Phys.*, 2005, **7**, 1–14.
54 T. Schmidt, G. J. Schutz, W. Baumgartner, H. J. Gruber and H. Schindler, *J. Phys. Chem.*, 1995, **99**, 17662–17668.
55 A. Kusumi, Y. Sako and M. Yamamoto, *Biophys. J.*, 1993, **65**, 2021–2040.
56 G. J. Schütz, H. Schindler and T. Schmidt, *Biophys. J.*, 1997, **73**, 1073–1080.
57 J. Rudnick and G. Gaspari, *Science*, 1987, **237**, 384–389.
58 C. A. Werley and W. E. Moerner, *J. Phys. Chem. B*, 2006, **110**, 18939–18944.
59 B. Stempfle, M. Dill, M. Winterhalder, K. Müllen and D. Wöll, *Polym. Chem.*, 2012, **3**, 2456–2463.
60 M. J. Saxton, *Biophys. J.*, 1997, **72**, 1744–1753.
61 M. C. Konopka and J. C. Weisshaar, *J. Phys. Chem. A*, 2004, **108**, 9814–9826.
62 A. Karrenbauer and D. Wöll, *arXiv:1212.5877*, 2012.
63 H. Qian, M. P. Sheetz and E. L. Elson, *Biophys. J.*, 1991, **60**, 910–921.
64 F. Köhn, J. Hofkens, R. Gronheid, M. Van der Auweraer and F. C. De Schryver, *J. Phys. Chem. A*, 2002, **106**, 4808–4814.
65 T. Ha, T. Enderle, D. S. Chemla, P. R. Selvin and S. Weiss, *Chem. Phys. Lett.*, 1997, **271**, 1–5.
66 C. Flors, J. Hotta, H. Uji-i, P. Dedecker, R. Ando, H. Mizuno, A. Miyawaki and J. Hofkens, *J. Am. Chem. Soc.*, 2007, **129**, 13970–13977.
67 D. Montiel, H. Cang and H. Yang, *J. Phys. Chem. B*, 2006, **110**, 19763–19770.