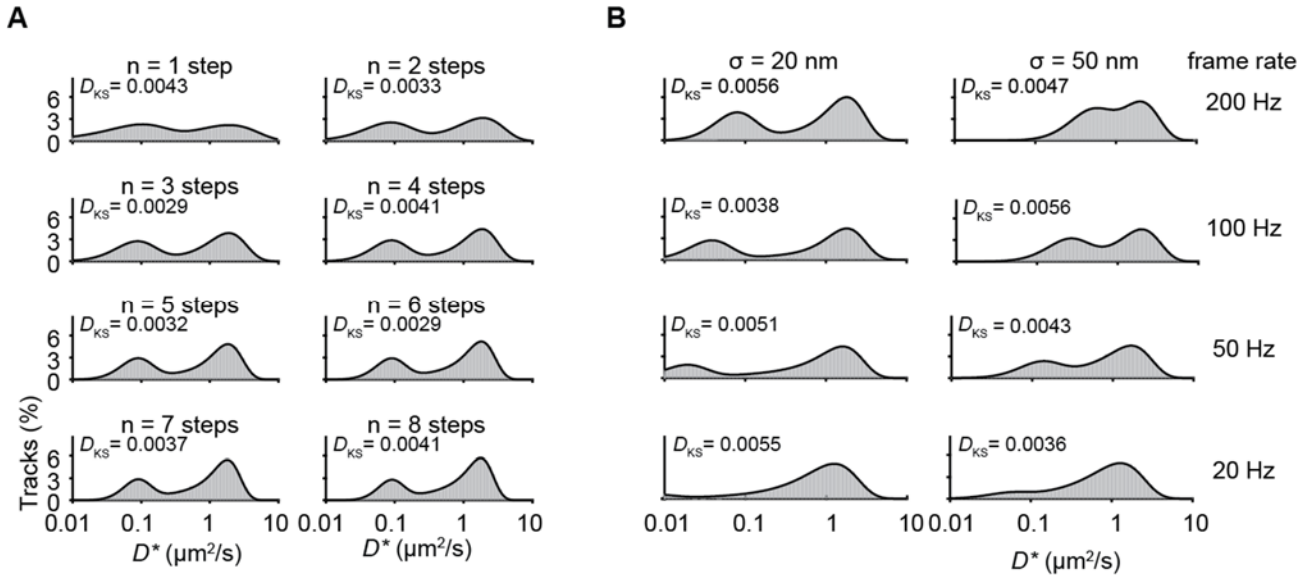


**Biophysical Journal, Volume 119**

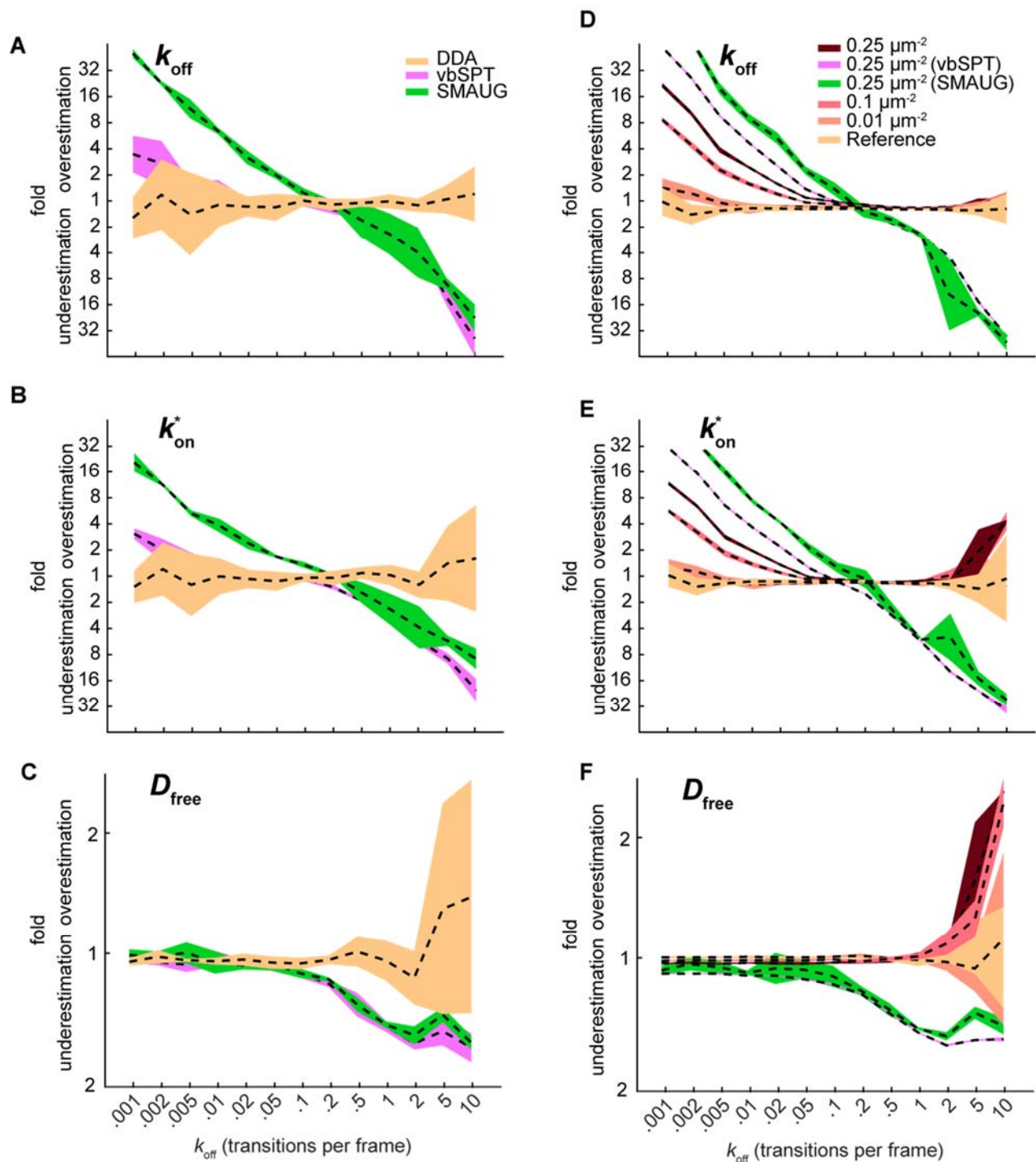
**Supplemental Information**

**Extracting Transition Rates in Particle Tracking Using Analytical Diffusion Distribution Analysis**

**Jochem N.A. Vink, Stan J.J. Brouns, and Johannes Hohlbein**

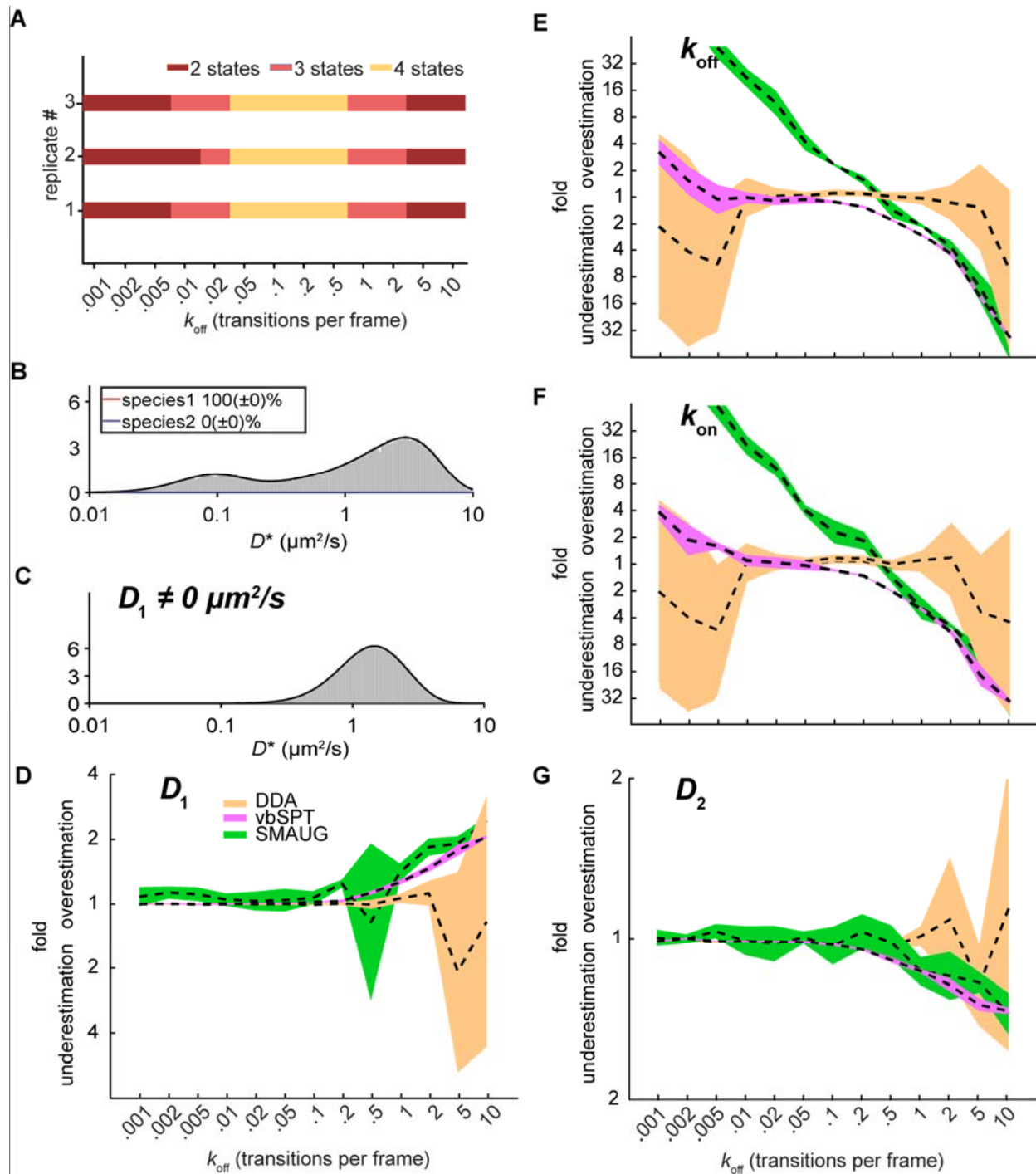


**Figure S1. The effect of track length, localization error and frame rate on the shape of  $D^*$  distributions. (A)**  $D^*$  distributions for different number of steps within a single trajectory, simulated (grey boxes) and DDA predicted distributions (black line) for 1-8 number of steps. Simulation parameters:  $k_{\text{on}}^* = 0.2 \text{ frame}^{-1}$ ,  $k_{\text{off}}^* = 0.2 \text{ frame}^{-1}$ ,  $D_{\text{free}} = 4 \mu\text{m}^2/\text{s}$ ,  $\sigma = 30 \text{ nm}$  (localisation precision) and  $n = 50.000$ . **(B)**  $D^*$  distributions for different localization errors and frame rates. Increasing the framerate shifts the peak of the bound population left to lower  $D^*$  values and increased localization errors shift this peak right to higher  $D^*$  values. At low frame rates, due to time averaging, the combined distribution of two states merge towards a single peak. Simulation parameters:  $k_{\text{on}}^* = 0.02 \text{ frame}^{-1}$ ,  $k_{\text{off}}^* = 0.02 \text{ frame}^{-1}$ ,  $D_{\text{free}} = 2 \mu\text{m}^2/\text{s}$  and  $n = 50.000$ . The Kolmogorov-Smirnov test statistic ( $D_{KS}$ ) is indicated at each histogram.



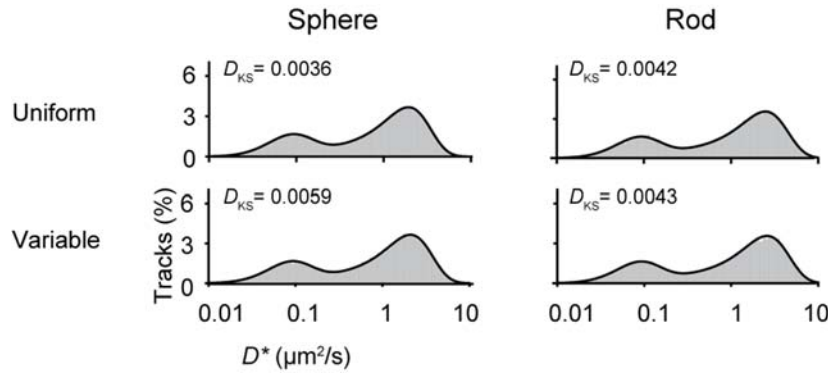
**Figure S2. The effect of small data size and tracking errors on the extraction of parameters**  
 (A-C) Same as Figures 2D-F but with 1.000 tracks instead of 50.000 tracks. (D-F) Same as Figures 2D-F but with increasing particle densities. The simulations were modified in order that

multiple particles were present in the same area at the same time point and therefore had a probability of being erroneously linked. The extraction accuracy was also compared to SMAUG and vbSPT for the highest tested density (0.25 localizations/ $\mu\text{m}^2$ ). The reference (yellow) indicates the standard simulation (without tracking errors) run elsewhere in the manuscript. Tracking was done with algorithms described previously (1) and a tracking window of 0.8  $\mu\text{m}$ . Other parameters included in all figures were:  $k_{\text{on}}^* = k_{\text{off}}$ ,  $D_{\text{free}} = 4 \mu\text{m}^2/\text{s}$  and  $\sigma = 30 \text{ nm}$ .

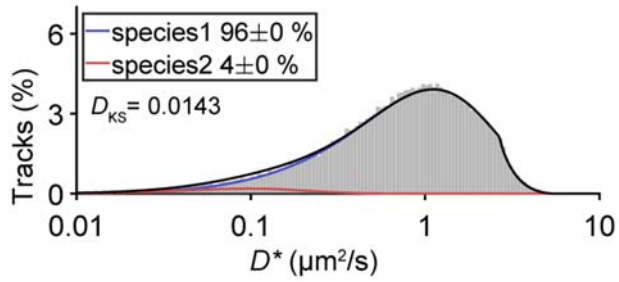


**Figure S3. Model selection for multiple states and parameter extraction of systems with more than one mobile state. (A)** The number of states extracted from vbSPT in a simulated two-state system. After removing restrictions on the maximum amount of states in vbSPT, the number of states fitted under some conditions differed from the amount of states modelled (2

states). Indicated are which simulated replicates contained which number of states (dark red: 2 states, light red: 3 states and yellow 4 states. **(B)** Result of fit with two species on a simulation containing only a single species  $k_{\text{on}}^* = k_{\text{off}} = 0.2 \text{ frame}^{-1}$ ,  $D_{\text{Free}} = 4 \mu\text{m}^2/\text{s}$ ,  $\sigma = 30 \text{ nm}$  and step number 4. **(C)**  $D^*$  distribution for a system with two mobile states, simulated (grey boxes) and DDA predicted distributions (black line). Parameters used were:  $k_{\text{on}}^* = k_{\text{off}} = 0.2 \text{ frame}^{-1}$ ,  $D_1 = 1 \mu\text{m}^2/\text{s}$ ,  $D_2 = 4 \mu\text{m}^2/\text{s}$ ,  $\sigma = 30 \text{ nm}$  and step number 4. **(D-G)** Same as Figures 2D-F but with two mobile states (which both are estimated from the data). Parameters used were  $k_{\text{on}}^* = k_{\text{off}}$ ,  $D_1 = 1 \mu\text{m}^2/\text{s}$ ,  $D_2 = 4 \mu\text{m}^2/\text{s}$  and  $\sigma = 30 \text{ nm}$ .



**Figure S4. Distributions of a population of cells with uniform and variable cell size.** The cell shapes were either spherical (left) or rod-shaped (right; radius to length ratio is 1:8). The average radius for uniform (upper row) and variable (lower row) cell sizes was the same:  $r_{\text{confined}} = \sqrt{5D_{\text{free}}t}$ . For the variable cell size, 60% of the cells were simulated with the same size as the average, whereas for 20% of the cells were simulated 25% smaller and for 20% of the cells were simulated with 25% larger cells. Further parameters used were:  $k_{\text{on}}^* = 0.02 \text{ frame}^{-1}$ ,  $k_{\text{off}}^* = 0.02 \text{ frame}^{-1}$ ,  $D_{\text{free}} = 4 \mu\text{m}^2/\text{s}$ ,  $\sigma = 30 \text{ nm}$  and  $dt = 0.01 \text{ s}$ . The Kolmogorov-Smirnov test statistic ( $D_{KS}$ ) is indicated at each histogram.



**Figure S5. DNA polymerase histogram for tracks with a step number of 2 steps.** The same condition and fitting parameters as for Figure 5A were used except that a two-step tracks are shown here. The maximum step size 5 pixels ( $0.6 \mu\text{m}$ ) that was initially applied to this dataset results in a discontinuous distribution, which is correctly captured by the ana-DDA fit. Data from previous study (2).



## Derivation of $D^*$ distributions of localization error

As mentioned in the methods section, the  $D^*$  distribution of localization error can be described by a summation of correlated gamma random variables. The extend by which the localization error affects the correlation of sequential steps can be quantified by calculating the correlation coefficient  $\rho_{ij} = \langle x, y \rangle / \sigma_x \sigma_y$  and the covariance of sequential steps as derived by Berglund(3)

$$\langle \Delta x_i, \Delta x_j \rangle = 2DR\Delta t - \sigma^2 \quad (1)$$

$$\text{for } |i - j| = 1,$$

where  $R$  is the motion blur coefficient caused by movement of the particle during the illumination time and  $D$  is the diffusion coefficient. We assume further that measurements were taken with very short illumination pulses leading to  $R \rightarrow 0$ . We further convert equation 20 to MSD notation

$$\langle \Delta x_i^2, \Delta x_j^2 \rangle = \langle \Delta x_i \Delta x_j \rangle^2 = \sigma^4. \quad (2)$$

After converting to two dimensions and assuming that  $\Delta x$  and  $\Delta y$  are independent, we get

$$\langle \Delta x_i^2 + \Delta y_i^2, \Delta x_j^2 + \Delta y_j^2 \rangle = 4 \langle \Delta x_i \Delta x_j \rangle^2 = 4\sigma^4. \quad (3)$$

To calculate the correlation coefficient  $\rho_{ij}$ , we use the following expression for the standard deviation of the MSD in two dimensions(4)

$$\sigma_{\text{MSD}} = 4D\Delta t + 4\sigma^2, \quad (4)$$

leading to

$$\rho_{ij} = \frac{\langle \Delta x_i^2 + \Delta y_i^2, \Delta x_j^2 + \Delta y_j^2 \rangle}{\sigma_{\text{MSD},i} \sigma_{\text{MSD},j}} = \frac{4\sigma^4}{(4D\Delta t + 4\sigma^2)^2}. \quad (5)$$

For most applications  $D\Delta t > \sigma^2$  and  $\rho$  can be neglected. However, for immobile particles  $D\Delta t = 0$ , and  $\rho = \frac{4\sigma^4}{(4\sigma^2)^2} = 1/4$ . For a number of  $n$  measured steps with localization error of an immobile particle, the correlation matrix  $\rho$  is therefore given by

$$\rho = \begin{bmatrix} 1 & 1/4 & 0 & 0 & \dots \\ 1/4 & 1 & 1/4 & 0 & \dots \\ 0 & 1/4 & 1 & 1/4 & \dots \\ 0 & 0 & 1/4 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}. \quad (6)$$

The summation of gamma random variables given a certain correlation matrix has been previously derived in terms of confluent Lauricella series(5). Using the definitions above, this equation can be written as

$$f_D(x|0, n) = \Phi_2(1, \dots, 1; n; -\frac{y}{\lambda_1}, \dots, -\frac{y}{\lambda_n}) x^{-1+n} / \det(A) \Gamma(n), \quad (7)$$

where  $\Phi_2$  is the confluent Lauricella function,  $\lambda_1 - \lambda_n$  are the eigenvalues of the matrix  $A = B \cdot B$ , where  $B$  is an  $n \times n$  matrix with diagonal values  $\sigma^2$ , and  $C$  is an  $n \times n$  matrix with values  $C_{ij} = \sqrt{\rho_{ij}}$ .

This summation, for each number of measured steps  $n$  is the modified distribution for immobile particles taking into account the correlation between sequential measured displacements. To implement this distribution in the calculation of our total  $D$  distributions, we subtract the fraction of immobile particles after  $n$  time steps ( $W_{contS1}(t_{S1} = 4t_f)$ , Eq.5 ) multiplied with the distribution of expected  $D^*$  for  $n$  time steps  $f_D(x|0, n)$  (Eq. 1) and replace it with the same fraction of immobilized particles multiplied with the distribution calculated based on the Lauricella series. The calculation of confluent Lauricella series was implemented from MATLAB code described in Martos-Naya et al. (2016)(6).

The equation above can be further refined to experimental data, if there is a large difference in localization error between particles. In that case, there is another correlation factor due to the difference in brightness/focus of particles. As some particles might show a dynamic brightness, e.g. by diffusing in and out the excitation/detection focus, localizations of this track will have a higher precision the brighter the emission of the particle is, altering the correlation matrix to

$$\begin{aligned}\rho_{ij} &= 1 \text{ for } i = j, \\ \rho_{ij} &= \frac{1}{4} + \frac{3}{4}r \text{ for } |i - j| = 1, \\ \rho_{ij} &= r \text{ for } |i - j| > 1,\end{aligned}$$

where  $r$  is the correlation coefficient between two steps within the same track not sharing any localizations ( $|i - j| > 1$ ). We found that this correlation coefficient can be experimentally determined by measuring correlation of displacement of immobilized particles, or by measuring the correlation of estimated localization errors within tracks. This can be done by making a matrix in MATLAB where the rows are the different tracks and the columns are either the different step size of immobilized particles or the estimated localization errors. The built-in function *coerrcoef* then automatically calculates the correlation coefficient of this dataset.

### **Derivation of confinement corrections**

The effective measured MSD given a diffusion coefficient  $D$  and a timestep  $\Delta t$  have been previously derived for a spherical geometry in multiple dimensions(7), from which we derived the effective diffusion coefficient given the geometry and the real diffusion coefficient for spherical or rod-shaped geometries. First, the authors defined the zeros  $\alpha_m$  at which  $j'_1(\alpha_m) =$

0, with  $j'_1$  being the derivative of the spherical Bessel function of the first kind. This can be rewritten as

$$(\alpha_m^2 - 2)\sin(\alpha_m) + 2\alpha_m \cos(\alpha_m) = 0. \quad (8)$$

Subsequently the effective measured MSD within a spherical confined space of radius  $r$  is equal to

$$MSD = r^2 \left( \frac{6}{5} - 12 \sum_{m=1}^{\infty} e^{-\frac{\alpha_m^2 t D}{r^2}} \frac{1}{\alpha_m^2 (\alpha_m^2 - 2)} \right). \quad (9)$$

This infinite series converges to zero. We therefore used the first 10.000 terms for calculation as a reasonable approximation. Because the previous equation refers to the three-dimensional MSD we use the following relation to calculate the observed diffusion coefficient we divide by  $6t$ ,

$$D_{\text{obs}} = \frac{MSD}{6t} = \frac{r^2}{6t} \left( \frac{6}{5} - 12 \sum_{m=1}^{\infty} e^{-\frac{\alpha_m^2 t D}{r^2}} \frac{1}{\alpha_m^2 (\alpha_m^2 - 2)} \right). \quad (10)$$

We can define the above equation as a function to calculate the observed diffusion given a certain radius, frame time and diffusion coefficient in the presence of spherical confinement

$$D_{\text{obs}} = f_{\text{spherical}}(r, t, D). \quad (11)$$

We then substitute  $f_D(x|D, 1)$  for  $f_D(x|D_{\text{obs}}, 1)$  and use equation 12 to calculate the distribution under any number of steps and given a  $k_{\text{off}}$  and  $k_{\text{on}}^*$ .

For rod geometries, there is no analytically derived solution available. However, we can combine the spherical derivation with a derivation in the same study for circular 2D geometries. In this geometry, the authors defined zeros  $\beta_m$  of the function  $J'_1(\beta_m) = 0$ , where  $J'_1$  is the derivative of the Bessel function of the first kind. Subsequently, the effective measured MSD within a circular confined space of radius  $r$  is equal to

$$MSD = r^2 \left( 1 - 8 \sum_{m=1}^{\infty} e^{-\frac{\alpha_m^2 t D}{r^2}} \frac{1}{\alpha_m^2 (\alpha_m^2 - 1)} \right), \quad (12)$$

which we can again convert to a function to calculate the observed diffusion coefficient, but now as the MSD is two-dimensional, we divide by  $4t$

$$D_{\text{obs}} = \frac{MSD}{4t} = \frac{r^2}{4t} \left( 1 - 8 \sum_{m=1}^{\infty} e^{-\frac{\alpha_m^2 t D}{r^2}} \frac{1}{\alpha_m^2 (\alpha_m^2 - 1)} \right) = f_{\text{circular}}(r, t, D). \quad (13)$$

To calculate the effective measured MSD in a rod-shaped geometry, we split the cell in two parts: the hemispherical (consisting of two hemi-spheres) and the cylindrical part. If the cell is much longer than it is wide the cylindrical part dominates. For diffusion within a cylinder, movement along the cell length is not restricted, whereas movement along the width of the cell is constrained as given by equation 31. If the cell is as long as wide, we have a spherical cell for which the diffusion is described by equation 28. For cells featuring intermediate aspect ratios, we can calculate the ratio of these two domains via the ratio of their volumes

$$V_{\text{total}} = V_{\text{cylindrical}} + V_{\text{hemi-sphere}} = \pi l r^2 + \frac{4}{3} \pi r^3, \quad (14)$$

where  $r$  is the radius of the cell width and  $l$  the length of the cylindrical part of the rod-shaped cell. The observed diffusion coefficient along the cell length,  $D_{\text{obs},x}$  is not being restricted in the cylindrical part. The observed diffusion coefficient along the cell width  $D_{\text{obs},y}$  on the contrary is restricted within the cylindrical part. Therefore, we separately calculate these two observed diffusion coefficients

$$D_{\text{obs},x}(r, t, D) = \frac{V_{\text{spherical}}}{V_{\text{total}}} f_{\text{spherical}}(r, t, D) + \frac{V_{\text{cylindrical}}}{V_{\text{total}}} D, \quad (15)$$

$$D_{\text{obs},y}(r, t, D) = \frac{V_{\text{spherical}}}{V_{\text{total}}} f_{\text{spherical}}(r, t, D) + \frac{V_{\text{cylindrical}}}{V_{\text{total}}} f_{\text{circular}}(r, t, D). \quad (16)$$

For the last step, we require a probability distribution function of the sum of the two distributions  $D_{\text{obs},x}$  and  $D_{\text{obs},y}$  and go back to the distribution of  $x$

$$x \sim N(0, \sqrt{2D_{\text{obs},x}t}), \quad (17)$$

representing a normal distribution with mean of zero and  $\sigma = \sqrt{2D_{\text{obs},x}t}$ . The distribution of the squared displacement is therefore a chi-square distribution

$$X^2/2D_{\text{obs},x}t = X^2/\sigma^2 \sim \chi_1^2. \quad (18)$$

The same holds for  $y$  with

$$Y^2/2D_{\text{obs},y}t \sim \chi_1^2. \quad (19)$$

To get to the distribution of  $D_{\text{obs}}$  we calculate

$$D_{\text{obs}} = D_{\text{obs},x}/2 \left( X^2/2D_{\text{obs},x}t \right) + D_{\text{obs},y}/2 \left( Y^2/2D_{\text{obs},y}t \right) = \frac{X^2 + Y^2}{4t}. \quad (20)$$

Consequently, the distribution of  $D_{\text{obs}}$  is a summation of two chi-square variables weighted by the different diffusion coefficients. The formula for this summation was given in a previous study(8) for the following case: Let  $X, Y \sim \chi_k^2$  two independent and identically distributed chi-square random variables with  $k$  degrees of freedom. Let  $Z := aX + bY$ , then the density function  $f_z$  is given by:

$$f_z = \theta(z) \frac{1}{(4ab)^{\frac{k}{2}}} \left( \frac{a-b}{8ab} \right)^{\frac{1}{2} - \frac{k}{2}} \frac{\Gamma\left(\frac{1}{2} + \frac{k}{2}\right)}{\Gamma(k)} e^{-\frac{a+b}{4ab}z} x^{\frac{k}{2} - \frac{1}{2}} I_{\frac{k}{2} - \frac{1}{2}}\left(\frac{b-a}{4ab}z\right) \quad (21)$$

In our case, where  $k = 1$ , this equation is simplified to

$$f_z = \frac{1}{(4ab)^{\frac{1}{2}}} \exp\left(-\frac{a+b}{4ab}z\right) I_0\left(\frac{b-a}{4ab}z\right), \quad (22)$$

where  $I_0$  is the zeroth order modified Bessel function of the first kind. When we substitute  $D_{\text{obs},x}/2$  and  $D_{\text{obs},y}/2$  for  $a$  and  $b$  respectively (combine equation 39 and 41), we obtain the following equation for the summation of two diffusion coefficients in two dimensions

$$f_D = \frac{1}{(D_{\text{obs},x}D_{\text{obs},y})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(D_{\text{obs},y} - D_{\text{obs},x})x\right) I_0\left(\frac{1}{2}(D_{\text{obs},y} - D_{\text{obs},x})x\right) \quad (23)$$

where  $x$  is the measured displacement as in equation 1. This distribution can then be used as substitution for  $f_D(x|D,1)$  and we can use equation 12 to solve the distribution under any number of steps and given a  $k_{\text{off}}$  and  $k_{\text{on}}^*$ .

## Supporting References

1. Crocker, J.C., and D.G. Grier. 1996. Methods of digital video microscopy for colloidal studies. *J. Colloid Interface Sci.* 179:298–310.
2. Uphoff, S., R. Reyes-Lamothe, F. Garza de Leon, D.J. Sherratt, and A.N. Kapanidis. 2013. Single-molecule DNA repair in live bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 110:8063–8068.
3. Berglund, A.J. 2010. Statistics of camera-based single-particle tracking. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 82.
4. Michalet, X. 2010. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 82:41914.
5. Paris, J.F. 2011. A Note on the Sum of Correlated Gamma Random Variables. *Available*

in *arxiv.org* at <http://arxiv.org/abs/1103.0505>.

6. Martos-Naya, E., J.M. Romero-Jerez, F.J. Lopez-Martinez, and J.F. Paris. A MATLAB TM program for the computation of the confluent hypergeometric function  $\Phi 2$ . .
7. Bickel, T. 2007. A note on confined diffusion. *Phys. A Stat. Mech. its Appl.* 377:24–32.
8. Bausch, J. 2013. On the efficient calculation of a linear combination of chi-square random variables with an application in counting string vacua. *J. Phys. A Math. Theor.*