



Supplementary Materials for

Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes

Francisco Balzarotti,* Yvan Eilers, Klaus C. Gwosch, Arvid H. Gynnå, Volker Westphal,
Fernando D. Stefani, Johan Elf, Stefan W. Hell*

*Corresponding author. Email:

Published 22 December 2016 on *Science* First Release
DOI: 10.1126/science.aak9913

This PDF file includes

Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 and S2
References

Contents of Supplementary Materials

1	FISHER INFORMATION AND CRAMÉR-RAO BOUNDS FOR MINFLUX	4
2	CRAMÉR-RAO BOUNDS FOR GEOMETRIES OF INTEREST	7
2.1	One-dimensional localization with two exposures	7
2.2	Two-dimensional localization with four doughnut exposures.....	10
3	POSITION ESTIMATORS.....	14
3.1	Maximum likelihood estimators	14
3.1.1	<i>MLE for 1D position with two parabolas</i>	<i>14</i>
3.1.2	<i>MLE for 2D position with four doughnut-shaped beams</i>	<i>15</i>
3.1.3	<i>MLE for 1D position with two Gaussian-shaped beams</i>	<i>16</i>
3.2	Other estimators	17
3.2.1	<i>Linearized least mean squared (LMS) estimator</i>	<i>18</i>
3.2.2	<i>Modified least mean squared (mLMS) estimator.....</i>	<i>19</i>
3.2.3	<i>Numerically unbiased mLMS (numLMS) estimator.....</i>	<i>19</i>
4	CAMERA LOCALIZATION	23
4.1	Static localization precision	23
5	DIFFUSION COEFFICIENT ESTIMATION	26
5.1	Algorithm overview	26
5.2	Applicability and performance of the OLSF estimator.....	27
6	MATERIALS AND METHODS	30
6.1	PSF and parameter vector \mathbf{p}	30
6.2	Localization performance map	31
6.3	MINFLUX nanoscopy	33
6.3.1	<i>Data acquisition.....</i>	<i>33</i>
6.3.2	<i>Trace segmentation.....</i>	<i>34</i>
6.3.3	<i>Localization of emission events</i>	<i>35</i>
6.3.4	<i>Clustering into nano-domains.....</i>	<i>35</i>
6.3.5	<i>Splitting events into equal number of photon events</i>	<i>35</i>
6.3.6	<i>Calculation of experimental covariance for fixed number of photons</i>	<i>35</i>
6.3.7	<i>Drift correction in post processing.....</i>	<i>36</i>
6.4	MINFLUX tracking	36
6.4.1	<i>Tracking of predefined trajectory.....</i>	<i>36</i>
6.4.2	<i>Tracking initiation: intensity thresholding and Gaussian localizations.....</i>	<i>36</i>
6.4.3	<i>The MINFLUX tracking routine</i>	<i>37</i>
6.4.4	<i>Trace segmentation and blinking statistics.....</i>	<i>37</i>
6.4.5	<i>Trajectory reconstruction</i>	<i>39</i>

6.4.6	<i>Error definition</i>	39
6.4.7	<i>Diffusion coefficient estimation</i>	40
6.4.8	<i>Localization precision</i>	40
6.5	Experimental setup.....	40
6.5.1	<i>Optical setup</i>	40
6.5.2	<i>System lock</i>	41
6.5.3	<i>Measurement control software</i>	42
6.6	Sample preparation.....	42
6.6.1	<i>Buffers</i>	42
6.6.2	<i>Cleaning of coverslips and objective slides</i>	42
6.6.3	<i>Flow channel</i>	43
6.6.4	<i>Fluorescent microspheres</i>	43
6.6.5	<i>Immobilized single ATTO 647N molecule</i>	43
6.6.6	<i>Immobilized labeled DNA origami</i>	44
6.6.7	<i>E. coli sample preparation</i>	44
SUPPLEMENTARY FIGURES		46
	Fig. S1.....	46
	Fig. S2.....	47
	Fig. S3.....	48
	Fig. S4.....	50
	Fig. S5.....	51
	Fig. S6.....	52
	Fig. S7.....	53
	Fig. S8.....	54
	Fig. S9.....	56
	Fig. S10.....	58
	Fig. S11.....	60
	Fig. S12.....	62
	Fig. S13.....	64
	Fig. S14.....	66
SUPPLEMENTARY TABLES		67
	Table S1.....	67
	Table S2.....	68

1 Fisher information and Cramér-Rao bounds for MINFLUX

In order to quantify how the proposed measurement scheme indeed increases the photon efficiency regarding the estimation of an emitter position, we calculate its Fisher information and its Cramér-Rao Bound (CRB).

We begin with a description of the statistics of photon collection, followed by the construction of a likelihood function. This is used for calculating the Fisher information that a given photon count measurement holds on the parameters of its statistics. Upon a multivariate space transformation, we obtain the Fisher information that said photons hold of the position of the emitter.

An emitter at a position $\bar{r}_m \in \mathbb{R}^d$ is exposed to a number of K different light intensities $\{I_0(\bar{r}), \dots, I_{K-1}(\bar{r})\}$ yielding a collection $\bar{n} = \{n_0, n_1, \dots, n_{K-1}\}$ of acquired photons per exposure (see fig. 1). Each number of photons n_i follows Poissonian statistics with a mean λ_i that depends on the corresponding light intensity $I_i(\bar{r}_m)$, where background and detector dark counts contributions are assumed to be negligible (background contributions are taken into account in eq. (S28)). When emitter saturation is avoided, the Poissonian mean λ_i can be approximated by

$$\lambda_i = c_e q_e \sigma_a I_i(\bar{r}_m) \quad (\text{S1})$$

where c_e is the collection efficiency of the system, q_e is the quantum yield and σ_a is the absorption cross-section of the emitter at the wavelength of the illumination I_i .

To facilitate the study in terms of the total acquired photons

$$N = n_0 + \dots + n_{K-1} \quad (\text{S2})$$

and to make the localization scheme independent of the intrinsic brightness of the molecules, the probabilities for measuring the collection of photons $P(\bar{n})$ is conditioned to N , yielding multinomial statistics:

$$P(n_i) \sim \text{Poisson}(\lambda_i) \text{ with } i \in [0, K-1] \Rightarrow P(\bar{n}|N) \sim \text{Multinomial}(\bar{p}, N)$$

$$P(\bar{n}|N) = \frac{N!}{n_0! \dots n_{K-1}!} \prod_{i=0}^{K-1} p_i^{n_i} \quad (\text{S3})$$

with the components of the parameter vector \bar{p} being

$$p_i^{(0)}(\bar{r}_m) = \frac{\lambda_i}{\sum_{j=0}^{K-1} \lambda_j} \approx \frac{I_i(\bar{r}_m)}{\sum_{j=0}^{K-1} I_j(\bar{r}_m)} \text{ with } i \in [0, \dots, K-1] \quad (\text{S4})$$

where, for the case of negligible dark counts (denoted by the (0) superscript), the brightness of the molecule is canceled out.

It should be stressed that, for the multinomial distribution, the number of *independent* photon acquisitions n_i (and parameters p_i) is $K-1$, as stated in eq. (S2) and

eq. (S4). Therefore, the number of independent elements p_i span a $(K - 1)$ -dimensional space, termed *reduced* \bar{p} -space henceforth. $P(\bar{n}|N)$ can thus be written as:

$$P(\bar{n}|N) = \frac{N!}{n_0! \dots n_{K-1}!} \left(\prod_{i=0}^{K-2} p_i^{n_i}(\bar{r}_m) \right) \left(1 - \sum_{j=0}^{K-2} p_j(\bar{r}_m) \right)^{n_{K-1}} \quad (\text{S5})$$

To quantify how much information the measured photon collection \bar{n} holds on the d -dimensional position of the molecule $\bar{r}_m = [r_{m1} \dots r_{md}]^T$, the Fisher information matrix $F_{\bar{r}_m}$ is to be calculated. It is evident from (S5) that the photon collection \bar{n} holds information on the success probability \bar{p} , which, in turn, holds information on the emitter position \bar{r}_m . Therefore, in a first step, the Fisher information $F_{\bar{p}}$ that the photon collection \bar{n} carries on \bar{p} is calculated. Thereafter, a coordinate transformation will result in the Fisher information $F_{\bar{r}_m}$ that the photon collection holds on the emitter position \bar{r}_m . The Fisher information is generally calculated as follows (38):

$$\{F_{\bar{\theta}}\}_{ij} = -E \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \mathcal{L}(\bar{\theta}|\bar{m}) \Big| \bar{\theta} \right) \quad (\text{S6})$$

where $\bar{\theta}$ is the parameter vector to be estimated and $\mathcal{L}(\bar{\theta}|\bar{m})$ the likelihood of the parameter $\bar{\theta}$ given the measurement \bar{m} . Thus, the Fisher Information $F_{\bar{p}}$ is given by:

$$\{F_{\bar{p}}\}_{ij} = E \left(-\frac{\partial^2}{\partial p_i \partial p_j} \ln \mathcal{L}(\bar{p}|\bar{n}) \Big| \bar{p} \right) \text{ with } i, j \in [0, \dots, K - 2] \quad (\text{S7})$$

Here $\mathcal{L}(\bar{p}|\bar{n}) = P(\bar{n}|\bar{p})$ is the likelihood of having the parameter \bar{p} when measuring a collection of photons \bar{n} . Operating with the definitions of equations (S5) and (S7) yields an expression for the Fisher information matrix for \bar{p} :

$$\{F_{\bar{p}}\}_{ij} = N \left(\frac{1}{p_{K-1}} + \delta_{ij} \frac{1}{p_i} \right) \text{ with } i, j \in [0, \dots, K - 2] \quad (\text{S8})$$

where δ_{ij} is the Kronecker delta function. Note that failing to take into account the interdependence of the parameters p_i in (S5) (i.e., that $p_{K-1}(\bar{r}_m) = 1 - \sum_{j=0}^{K-2} p_j(\bar{r}_m)$), results in an incorrect Fisher information matrix $F_{\bar{p}}$.

Finally, the Fisher information matrix $F_{\bar{r}_m}$ can be obtained from $F_{\bar{p}}$ upon reparametrization:

$$F_{\bar{r}_m} = \mathcal{J}^{*T} F_{\bar{p}} \mathcal{J}^* \quad (\text{S9})$$

where $\mathcal{J}^* \in \mathbb{R}^{(K-1) \times d}$ is the Jacobian matrix of the transformation from the \bar{r} -space to the *reduced* \bar{p} -space.

Combining equations (S8) and (S9) finally yields an expression for the Fisher information on the molecule position \bar{r}_m :

$$F_{\bar{r}_m} = \mathcal{J}^{*T} N \begin{bmatrix} \frac{1}{p_{K-1}} + \frac{1}{p_0} & \dots & \frac{1}{p_{K-1}} \\ \vdots & \ddots & \vdots \\ \frac{1}{p_{K-1}} & \dots & \frac{1}{p_{K-1}} + \frac{1}{p_{K-2}} \end{bmatrix} \mathcal{J}^*, \text{ with } \mathcal{J}^* = \begin{bmatrix} \frac{\partial p_0}{\partial r_{m1}} & \dots & \frac{\partial p_0}{\partial r_{md}} \\ \vdots & \ddots & \vdots \\ \frac{\partial p_{K-2}}{\partial r_{m1}} & \dots & \frac{\partial p_{K-2}}{\partial r_{md}} \end{bmatrix} \quad (\text{S10})$$

This expression can be simplified to the form:

$$F_{\bar{r}_m} = N \sum_{i=0}^{K-1} \frac{1}{p_i} \begin{bmatrix} \left(\frac{\partial p_i}{\partial r_{m1}}\right)^2 & \dots & \frac{\partial p_i}{\partial r_{m1}} \frac{\partial p_i}{\partial r_{md}} \\ \vdots & \ddots & \vdots \\ \frac{\partial p_i}{\partial r_{md}} \frac{\partial p_i}{\partial r_{m1}} & \dots & \left(\frac{\partial p_i}{\partial r_{md}}\right)^2 \end{bmatrix} \quad (\text{S11})$$

Once the Fisher information matrix is obtained, a lower bound for the covariance matrix of the molecule position $\Sigma(\bar{r}_m)$ can be derived from the Cramér-Rao inequality:

$$\Sigma(\bar{r}_m) \geq \Sigma_{CRB}(\bar{r}_m) = F_{\bar{r}_m}^{-1} \quad (\text{S12})$$

Thus, a CRB $\Sigma_{CRB}(\bar{r}_m)$ is obtained for each possible molecule position \bar{r}_m . We choose to analyze the arithmetic mean $\tilde{\sigma}_{CRB}$ of the eigenvalues σ_i^2 of this matrix and their isotropy \mathbb{I} as performance metrics.

$$\tilde{\sigma}_{CRB} = \sqrt{\frac{1}{d} \text{tr}(\Sigma_{CRB})} \quad (\text{S13})$$

$$\mathbb{I} = \frac{\min_{i \in [1,d]} \sigma_i}{\max_{i \in [1,d]} \sigma_i} \quad (\text{S14})$$

The use of $\tilde{\sigma}_{CRB}$ is equivalent to the commonly used norm (38) for the Fisher information matrix $\phi_q[F] = (\text{tr}(F^q)/d)^{1/q}$ with $q = -1$.

Table S1 displays explicit expressions for eq. (S4), (S8), (S13) and the *reduced* Jacobian for a set of dimensionalities d and number of exposures K .

2 Cramér-Rao bounds for geometries of interest

The general results of the previous section will now be evaluated for a number of specific conditions of interest, namely, dimensionality, number of exposures, beam shapes and arrangement.

For the sake of simplicity in the experimental design, in this work we focus on localization schemes where the respective beam intensities $I_i(\vec{r})$ are all equal except for a displacement by a distance \vec{r}_{b_i} , such that

$$I_i(\vec{r}) = I(\vec{r} - \vec{r}_{b_i}) \quad (\text{S15})$$

The beams of interest are the quadratic beam (meaning a beam with a central zero whose intensity increases quadratically with distance r from the center), standing wave, doughnut beam and Gaussian beam. The first one, though physically impossible, is used as a first relevant (i.e. quadratic) order approximation of the doughnut beam and the standing wave around its minima. The following definitions will be used:

$$\text{Quadratic} \quad I_{quad}(\vec{r}) = A_{quad} r^2 \quad (\text{S16})$$

$$\text{Doughnut} \quad I_{doughnut}(\vec{r}) = A_0 4e \ln 2 \frac{r^2}{fwhm^2} e^{-4 \ln 2 \frac{r^2}{fwhm^2}} \quad (\text{S17})$$

$$\text{Standing wave} \quad I_{sw}(\vec{r}) = A_i \sin^2(\vec{k} \cdot \vec{r}) \quad (\text{S18})$$

$$\text{Gaussian} \quad I_{gauss}(\vec{r}) = A_0 e^{-4 \ln 2 \frac{r^2}{fwhm^2}} \quad (\text{S19})$$

where A_{quad} is the concavity of the parabolic beam, A_0 is the peak intensity for the doughnut and Gaussian cases, $fwhm$ is a size-related parameter for the doughnut case (peak diameter occurs at $\ln(2)^{-1/2} fwhm \approx 1.2 fwhm$) and the full-width-half-maximum for the Gaussian beam, A_i are the respective peak intensities of the standing waves and \vec{k} determines the spacing between successive intensity minima and its direction. The quadratic approximation of the doughnut beam ($A_{quad} = A_0 4e \ln 2 / fwhm^2$) holds as long as

$$4 \ln 2 r^2 \ll fwhm^2 \quad (\text{S20})$$

2.1 One-dimensional localization with two exposures

In this case, the CRB of eq. (S13) is reduced to

$$\tilde{\sigma}_{CRB} \Big|_{k=2}^{d=1} = \frac{1}{\sqrt{N}} \frac{\sqrt{p_0(1-p_0)}}{\left| \frac{dp_0}{dr_0} \right|} \quad (\text{S21})$$

where p_0 is the parameter of a binomial distribution. By having the beams at positions $r_{b_0} = -L/2$ and $r_{b_1} = L/2$ (separated by a distance L) and using the beam definitions of eq. (S16)-(S18), with $\bar{r} = x$, the following expressions are obtained for the corresponding binomial parameters and CRB:

$$p_0^{quad}(x) = \frac{1 \left(1 + \frac{x}{L/2}\right)^2}{1 + \left(\frac{x}{L/2}\right)^2} \quad (\text{S22a})$$

$$\tilde{\sigma}_{CRB}^{quad}(x) = \frac{1}{\sqrt{N}} \frac{L}{4} \left[1 + \left(\frac{x}{L/2}\right)^2\right] \quad (\text{S22b})$$

$$\tilde{\sigma}_{CRB}^{quad}(0) = \frac{1}{\sqrt{N}} \frac{L}{4} \quad (\text{S22c})$$

$$p_0^{doughnut}(x) = \frac{\frac{1}{2} \left(1 + \frac{x}{L/2}\right)^2 e^{-\frac{4 \ln 2 xL}{fwhm^2}}}{\left[1 + \left(\frac{x}{L/2}\right)^2\right] \cosh\left(\frac{4 \ln 2 xL}{fwhm^2}\right) - 2 \frac{x}{L/2} \sinh\left(\frac{4 \ln 2 xL}{fwhm^2}\right)} \quad (\text{S22d})$$

$$\sigma_{CRB}^{doughnut}(x) = \frac{\frac{1}{\sqrt{N}} \frac{L}{4} \left| \left[1 + \left(\frac{x}{L/2}\right)^2\right] \cosh\left(\frac{4 \ln 2 xL}{fwhm^2}\right) - 2 \frac{x}{L/2} \sinh\left(\frac{4 \ln 2 xL}{fwhm^2}\right) \right|}{\left| 1 + \ln 2 \frac{L^2}{fwhm^2} \left[\left(\frac{x}{L/2}\right)^2 - 1\right] \right|} \quad (\text{S22e})$$

$$\tilde{\sigma}_{CRB}^{doughnut}(0) = \frac{1}{\sqrt{N}} \frac{L}{4} \frac{1}{1 - \ln 2 \frac{L^2}{fwhm^2}} \quad (\text{S22f})$$

$$p_0^{sw}(x) = \frac{\sin^2 \left[k \left(x + \frac{L}{2} \right) \right]}{\sin^2 \left[k \left(x + \frac{L}{2} \right) \right] + \sin^2 \left[k \left(x - \frac{L}{2} \right) \right]} \quad (\text{S22g})$$

$$\tilde{\sigma}_{CRB}^{sw}(x) = \frac{\left| \cos \left[2k \left(x + \frac{L}{2} \right) \right] + \cos \left[2k \left(x - \frac{L}{2} \right) \right] - 2 \right|}{4k\sqrt{N} \sin kL} \quad (\text{S22h})$$

$$\tilde{\sigma}_{CRB}^{sw}(0) = \frac{1}{2k\sqrt{N}} \tan \frac{kL}{2} \quad (\text{S22i})$$

Figure S1 shows, for each beam shape, the intensities $I_i(x)$ (A-C), the binomial success probability parameter $p(x)$ (D-F) and the CRB (G-I) for beam separations L of 25 nm, 50 nm and 150 nm, a beam size parameter ($fwhm$) of 300 nm, a k -value of $k = 2\pi/\lambda$, with $\lambda = 600$ nm and a total number of photons $N = 100$.

The case of doughnut beams –displayed in fig. S1A,D,G– produced the intricate expressions eq. (S22d)-(S22f), all of which can be approximated by their quadratic beam counterparts eq. (S22a)-(S22c) as long as (i) the emitter resides in the central region

(described by eq. (S20)) and (ii) the beam separation is smaller than the overall beam size ($L \ll fwhm/\sqrt{\ln 2}$). These conditions encompass the regime of interest, therefore the quadratic beam case displayed in fig. S1B,E,H is the one to be analyzed further.

In the quadratic beam approximation, the CRB scales linearly with the beam separation L , which makes the localization more precise the closer the beams are. However, there is a tradeoff between precision at the origin and the field of view. Remarkably low CRB values below 5 nm are achieved with only 100 measured photons within a region of about 100 nm around the origin.

The standing wave case is displayed in fig. S1C,F,I. The depicted functions are based on the eq. (S22g)-(S22i), which can again be approximated by their quadratic beam counterparts (S22a)-(S22c) as long as (i) the emitter is close to an intensity zero ($\Delta r \ll \pi/k$) and (ii) the beam separation is smaller than half the sine period ($L \ll \pi/k$). In the case of $L = 150$ nm and $k = 2\pi/600$ nm⁻¹ these conditions are not fulfilled anymore. Note that the CRB is flat in this case.

The case of two Gaussian beams is treated in the same way as the other cases, despite the absence of local minima. Using the same definitions as above and eq. (S19) yields

$$p_0^{gauss}(x) = \frac{e^{\frac{4 \ln 2 x L}{fwhm^2}}}{2 \cosh \frac{4 \ln 2 x L}{fwhm^2}} \quad (S23a)$$

$$\tilde{\sigma}_{CRB}^{gauss}(x) = \frac{1}{\sqrt{N}} \frac{fwhm^2}{4 \ln 2 L} \cosh \frac{4 \ln 2 x L}{fwhm^2} \quad (S23b)$$

$$\tilde{\sigma}_{CRB}^{gauss}(0) = \frac{1}{\sqrt{N}} \frac{fwhm^2}{4 \ln 2 L} \quad (S23c)$$

Figure S4 shows the intensities $I_i(x)$ (A), the binomial parameter $p(x)$ (C) and the CRB (E) for beam separations L of 25 nm, 50 nm, 150 nm, 300 nm, 600 nm, and 1200 nm. The performance of the Gaussian beam case is poor for small beam separations. This is visible in the slope of the binomial parameter $p(x)$ in fig. S4C and leads to the high values of the CRBs in fig. S4E, in contrast with the situation found with beams that feature minima (fig. S1).

It should be noted from eq. (S23c) that the localization CRB for Gaussian beams scales as $\propto L^{-1}$ at the origin. This implies that the further apart the beams are, the more precise the emitter localization will be (as shown in fig. S4). This makes sense, as the series expansion of largely separated Gaussian beams around the $x = 0$ has significant contributions only from order 2 (as long as the field of view $\ll 3/[2 \ln(2) L]$). Therefore, they are not distinguishable from the parabolic beams discussed previously, mimicking intensity minima.

Increasing the beam separation also produces a faster growth of the CRB away from the origin (given the $\propto xL/fwhm^2$ dependence in eq. (S23b), reducing the method's effective *field of view*. Practical applications with these beams may be highly affected by background collection and aberrations.

It is straightforward to state that a proper manipulation of the beams $I_i(\vec{r})$ enables great flexibility of the CRB which can be adapted for different applications. We have illustrated this point by using the doughnut and Gaussian beam shapes and the separation L as a parameter, but the possibilities are not at all limited to these cases. Arbitrary beam shapes generated, for example, by a spatial light modulator (SLM) could be used to optimize for specific properties, e.g. a spatially flatter CRB. Optimizations of the beams according to the application specifications (dimensionality, speed, field of view, precision, etc.) and to the implementation technology will follow on further studies. In this work we focus on the many benefits on speed, precision and observation time that a collection of displaced doughnut beams can bring for single molecule localization applied for tracking and imaging.

2.2 Two-dimensional localization with four doughnut exposures

Localizing an emitter in a plane requires at least three exposures due to the constraining eq. (S2). In this work, the pattern shown in fig. 2 is extensively utilized. The pattern is composed of four exposures with doughnut shaped excitation profiles eq. (S17): three equidistant beams with their zero on a circle of diameter L and a fourth one placed at the origin. The respective displacements are chosen to be:

$$\begin{aligned} \bar{r}_{b_0} &= [0,0]^T \\ \bar{r}_{b_i} &= \frac{L}{2} \cdot [\cos(\alpha_i), \sin(\alpha_i)]^T, \quad \text{for } \alpha_i = i \cdot \frac{2\pi}{3} \text{ and } i \in [1,3] \end{aligned} \quad (\text{S24})$$

This particular choice of an excitation beam pattern (EBP) has some advantages over other more straightforward possible choices, e.g., a square four-beam pattern or a triangular three-beam pattern. Doughnut beams conveniently increase the Fisher information at the central region of the EBP, as discussed previously for the one-dimensional case. However, successive maxima and minima can lead to indeterminations in the position. This means the same parameter vector \vec{p} is obtained at multiple positions of the emitter, and therefore similar measured photon collections \bar{n} . The central doughnut beam is included to counteract position ambiguities.

As an example, fig. S2A shows the likelihood function $\mathcal{L}(\vec{r}_m|\bar{n}) = P(\bar{n}|N, \vec{r}_m)$ as a function of the molecule position \vec{r}_m for an arbitrary combination of collected photons $\bar{n} = \{n_1, n_2, n_3\}$ from only three peripheral doughnut exposures. Though multiple positions exhibit local maxima for the likelihood function \mathcal{L} , a unique ML estimator is defined as long as one of them is absolute. However, fluctuation-driven hopping between two local maxima with similar likelihood values make such an estimator badly behaved.

The inclusion of a fourth doughnut exposure at the origin enhances the localization procedure by including radial information. This is readily visible from the $\bar{p}(\bar{r})$ function visualization in fig. S7D-G; note that the shown function results from a measured doughnut. The radial information resolves most indeterminations, especially in the region surrounding the beam pattern origin, leading to a better behaved estimator. This is illustrated in fig. S2B, where the likelihood function \mathcal{L} displays a single maximum. Interestingly, the photon distribution is *exactly the same* as in the previous example, except for the collection of zero photons during the central doughnut exposure. This example illustrates how the *measurement of zero photons* fundamentally improves the localization. (Emissions from the sample are obviously not required for this improvement).

The inclusion of additional exposures (or structured detection) is not the only way to avoid multiple local maxima in the likelihood function. The use of non-identical beams should allow neutralizing this inconvenience without further exposures.

Using eq. (S11) and eq. (S12), an explicit expression for the two-dimensional ($d = 2$) localization CRB Σ_{CRB} and the arithmetic mean of its eigenvalues $\tilde{\sigma}_{CRB}$ can be derived:

$$\Sigma_{CRB}(\bar{r}_m) = \frac{1}{N} \left(\left[\sum_{i=0}^{K-1} \frac{1}{p_i} \left(\frac{\partial p_i}{\partial x} \right)^2 \right] \left[\sum_{i=0}^{K-1} \frac{1}{p_i} \left(\frac{\partial p_i}{\partial y} \right)^2 \right] - \left[\sum_{i=0}^{K-1} \frac{1}{p_i} \frac{\partial p_i}{\partial x} \frac{\partial p_i}{\partial y} \right]^2 \right)^{-1} \times \sum_{i=0}^{K-1} \frac{1}{p_i} \begin{bmatrix} \left(\frac{\partial p_i}{\partial y} \right)^2 & -\frac{\partial p_i}{\partial x} \frac{\partial p_i}{\partial y} \\ -\frac{\partial p_i}{\partial x} \frac{\partial p_i}{\partial y} & \left(\frac{\partial p_i}{\partial x} \right)^2 \end{bmatrix} \quad (S25)$$

$$\tilde{\sigma}_{CRB} = \sqrt{\frac{1}{2N} \frac{\sum_{i=0}^{K-1} \frac{1}{p_i} \left[\left(\frac{\partial p_i}{\partial y} \right)^2 + \left(\frac{\partial p_i}{\partial x} \right)^2 \right]}{\left[\sum_{i=0}^{K-1} \frac{1}{p_i} \left(\frac{\partial p_i}{\partial x} \right)^2 \right] \left[\sum_{i=0}^{K-1} \frac{1}{p_i} \left(\frac{\partial p_i}{\partial y} \right)^2 \right] - \left[\sum_{i=0}^{K-1} \frac{1}{p_i} \frac{\partial p_i}{\partial x} \frac{\partial p_i}{\partial y} \right]^2}} \quad (S26)$$

Combining these expressions with the doughnut beam definition from eq. (S17) and the beam displacements of eq. (S24), it is possible to calculate analytically a performance metric of the proposed localization scheme. The functional dependence of $\tilde{\sigma}_{CRB}$, obtained by solving eq. (S26) at the origin, is given by:

$$\tilde{\sigma}_{CRB}(\bar{r} = \bar{0}) = \frac{L}{2\sqrt{2N}} \left(1 - \frac{L^2 \ln(2)}{fwhm^2} \right)^{-1} \quad (S27)$$

For a beam separation that is smaller than the overall beam size ($L \ll fwhm/\sqrt{\ln 2}$) the CRB scales linearly with the beam separation L . Therefore, as in the one-dimensional case, we have the possibility of increasing the photon efficiency of the localization process by means of geometrical degrees of freedom in the excitation pattern.

Figure S3A visualizes this through the spatial dependence of $\tilde{\sigma}_{CRB}$ for different pattern diameters L and a total number of $N = 100$ photons. Analogously to the one dimensional case (see figure S1H), the CRB is lowest around the origin and retains a low value within a diameter L . Figure S3B shows the covariance matrix Σ_{CRB} as a quadratic form for a discrete grid of positions separated 9 nm for $N = 1000$. Note, that the localization precision is not isotropic. Figure S3C illustrates the dependence of $\tilde{\sigma}_{CRB}$ as a function of the total number of photons N for different pattern sizes L at the origin.

The result in eq. (S27) is obtained in the limit of an infinitesimally small background contribution. In order to evaluate the expected performance of our method in a more realistic manner, the influence of the background – which is unavoidable in every realistic implementation of the method – on the CRB will be discussed. Most importantly, the definition of the parameter vector $\bar{p}^{(0)}$ stated in eq. (S4) has to be adapted. We assume that all relevant background contributions (including the usually negligible detector dark counts) follow Poissonian statistics such that they can be described by a Poissonian distribution with mean λ_{bi} . The modified parameter vector \bar{p} is then given by:

$$p_i(\bar{r}_m) = \frac{\lambda_i + \lambda_{bi}}{\sum_{j=0}^{K-1} (\lambda_j + \lambda_{bj})} \approx \frac{c_e q_e \sigma_a I_i(\bar{r}_m) + \lambda_{bi}}{\sum_{j=0}^{K-1} (c_e q_e \sigma_a I_j(\bar{r}_m) + \lambda_{bj})} \text{ with } i \in [0, K-1] \quad (\text{S28})$$

The convenient cancellation of the intrinsic brightness of the molecule, as in eq. (S4), is not applicable in this situation anymore. We can define a signal-to-background ratio SBR as

$$SBR(\bar{r}_m) = \frac{\sum_{j=0}^{K-1} \lambda_j}{\sum_{j=0}^{K-1} \lambda_{bj}} \approx \frac{c_e q_e \sigma_a \sum_{j=0}^{K-1} I_j(\bar{r}_m)}{\sum_{j=0}^{K-1} \lambda_{bj}} \quad (\text{S29})$$

with the definitions as above. It should be noted that the signal-to-background ratio SBR is a function of the set of exposures $\{I_i(\bar{r})\}$ and the position of the emitter \bar{r}_m . With this definition, $p_i(\bar{r}_m)$ can be written as

$$\begin{aligned} p_i(\bar{r}_m) &= \frac{SBR(\bar{r}_m)}{SBR(\bar{r}_m) + 1} \frac{\lambda_i}{\sum_{j=0}^{K-1} \lambda_j} + \frac{1}{SBR(\bar{r}_m) + 1} \frac{\lambda_{bi}}{\sum_{j=0}^{K-1} \lambda_{bj}} \\ &\approx \frac{SBR(\bar{r}_m)}{SBR(\bar{r}_m) + 1} \frac{I_i(\bar{r}_m)}{\sum_{j=0}^{K-1} I_j(\bar{r}_m)} + \frac{1}{SBR(\bar{r}_m) + 1} \frac{\lambda_{bi}}{\sum_{j=0}^{K-1} \lambda_{bj}} \\ &\approx \frac{SBR(\bar{r}_m)}{SBR(\bar{r}_m) + 1} p_i^{(0)}(\bar{r}_m) + \frac{1}{SBR(\bar{r}_m) + 1} \frac{1}{K} \end{aligned} \quad (\text{S30})$$

Again, the beam definition and its displacements are taken from eq. (S17) and eq. (S24), respectively. It is assumed that the background contributions depend on the excitation power only, such that these are rendered equal for the respective beams, i.e., $\lambda_{bi} = \lambda_b$ for all $i \in [0, K-1]$. Using eq. (S26) with eq. (S30), the functional dependence of $\tilde{\sigma}_{CRB}$ at the origin is now given by:

$$\tilde{\sigma}_{CRB}(\bar{r} = \bar{0}) = \frac{L}{2\sqrt{2N}} \left(1 - \frac{L^2 \ln(2)}{fwhm^2}\right)^{-1} \sqrt{\left(1 + \frac{1}{SBR(\bar{0})}\right) \left(1 + \frac{3}{4 SBR(\bar{0})}\right)} \quad (\text{S31})$$

Note, that the $SBR(\bar{0})$ as defined in (S29) has a dependence on L . In an experimental realization, the actual background λ_b will normally not depend on the beam displacement. The emitter at $\bar{r} = \bar{0}$ does see a reduced total intensity when L is diminished, though. It follows then that $SBR(\bar{0})$ decreases with a reduction of L . Assuming, that $SBR(\bar{0}, L_0)$ (the SBR at the origin for a beam displacement L_0) is known, $SBR(\bar{0}, L)$ can be written as:

$$SBR(\bar{0}, L) = \frac{L^2}{L_0^2} \exp\left(\frac{\ln(2)}{fwhm^2} (L_0^2 - L^2)\right) \cdot SBR(\bar{0}, L_0) \quad (\text{S32})$$

Figure S3D visualizes the influence that the SBR has on $\tilde{\sigma}_{CRB}$ at the origin for different $SBR(\bar{0}, L_0 = 100 \text{ nm})$ values. The presence of background increases the value of $\tilde{\sigma}_{CRB}(\bar{0})$ and introduces a limit to the usable beam separation parameter L . The localization precision cannot be improved further by reducing the beam separation over that limit. The situation is analogous to a non-perfect intensity zero (of a doughnut) in STED microscopy, where a finite contrast between the central minimum and the doughnut maximum (i.e. a non-perfect zero) puts an upper limit to the applicable STED-beam intensity and hence to the smallest fluorescence region attainable by STED. Using different beam shapes, however, might enable flexibility on this aspect.

3 Position estimators

Section 2 investigated the Cramér-Rao bounds (CRBs) for different excitation geometries. It was shown that the information content on the position estimation can be increased significantly in a region of interest, and be far higher compared to the information content present in ideal camera localization (no amplification noise, no readout noise, see section 4). This theoretical limit provides the best possible performance that any unbiased estimator can achieve, but not how the actual estimators perform. Section 3.1 investigates the performance of the maximum likelihood estimator (MLE), for different excitation geometries. Section 3.2 investigates further estimators optimized for low photon applications.

3.1 Maximum likelihood estimators

In order to estimate the position \bar{r}_m of an emitter, we use the maximum likelihood estimator \hat{r}_m^{MLE} defined as:

$$\hat{r}_m^{MLE} = \arg \max \mathcal{L}(\bar{r}|\bar{n}) \quad (\text{S33})$$

where $\mathcal{L}(\bar{r}|\bar{n}) = P(\bar{n}|N, \bar{r}_m)$ is the likelihood function dependent on the conditional probability distribution of the measured set of photons $P(\bar{n}|N, \bar{r}_m)$ defined in eq. (S3).

Three instances of MLE calculations follow in this section: (i) the one-dimensional localization with two parabolas, (ii) the two-dimensional localization with four doughnut exposures including background as described in section 2.2, which is the main method presented in this work, and (iii) the one-dimensional localization with two Gaussian exposures, which is used in the intermediate localization step for tracking described in section 6.4.1.

3.1.1 MLE for 1D position with two parabolas

In the vicinity of intensity zeros, a standing wave as well as a doughnut can be approximated in first relevant order by a parabola. It is therefore enlightening to state the MLE for the position in the case of displaced parabolic excitation intensities. Following the definition in eq. (S15), two beams $I_0(x)$ and $I_1(x)$ with parabolic excitation profile (S16) are placed at positions $x_{b_0} = -L/2$ and $x_{b_1} = L/2$, i.e. separated by a distance L . In this excitation geometry, the parameter vector \bar{p} defined in eq. (S4) is given by:

$$p_0 = \frac{\left(1 + \frac{2x}{L}\right)^2}{2\left(1 + \frac{4x^2}{L^2}\right)}, \quad p_1 = \frac{\left(1 - \frac{2x}{L}\right)^2}{2\left(1 + \frac{4x^2}{L^2}\right)} \quad (\text{S34})$$

and leads to the following likelihood function $\mathcal{L}(x|\bar{n})$:

$$\mathcal{L}(x|\bar{n}) = \frac{N!}{n_0! n_1!} \prod_{i=0}^1 p_i^{n_i} = \frac{N!}{n_0! n_1!} \frac{\left(1 + \frac{2x}{L}\right)^{2n_0} \left(1 - \frac{2x}{L}\right)^{2n_1}}{\left(2 + \frac{8x^2}{L^2}\right)^N} \quad (\text{S35})$$

The maximum likelihood estimator \hat{x}_m^{MLE} is then calculated:

$$\begin{aligned} \frac{d}{dx} \mathcal{L}(\hat{x}_m^{MLE}|\bar{n}) &= 0 \\ \Rightarrow \hat{x}_{m,1}^{MLE}(n_0, N) &= -\frac{L}{2} + \frac{L}{1 + \sqrt{\frac{n_1}{n_0}}} \\ \hat{x}_{m,2}^{MLE}(n_0, N) &= -\frac{L}{2} + \frac{L}{1 - \sqrt{\frac{n_1}{n_0}}} \end{aligned} \quad (\text{S36})$$

The MLE has two solutions, where $\hat{x}_{m,1}^{MLE}$ is the result for the region defined by $-\frac{L}{2} < x < \frac{L}{2}$ and therefore the one of interest. Additional exposures can make the MLE unique.

3.1.2 MLE for 2D position with four doughnut-shaped beams

The geometry employed in the two-dimensional localization scheme using four doughnut shaped excitation profiles is depicted in fig. 2 and explained in more detail in section 2.2.

As it is classicly done, we maximize the log-likelihood function and drop the multiplicative factors that only depend on \bar{n} . Thus, starting from eq. (S33) we obtain a simpler function $\ell(\bar{r}|\bar{n})$ to maximize

$$\ln \mathcal{L}(\bar{r}|\bar{n}) \propto \ell(\bar{r}|\bar{n}) = \sum_{i=0}^{K-1} n_i \ln p_i \quad (\text{S37})$$

where the multinomial distribution definition from eq. (S3) was used. An intricate expression is obtained for this simplified likelihood function $\ell(\bar{r}|\bar{n})$ when the parameter vector \bar{p} defined in eq. (S28) is evaluated for the four doughnut beam case defined in eq. (S17) and eq. (S24) (see fig. S7D-G for a visualization). Given the complexity of finding and analytical solution to the maximization of $\ell(\bar{r}|\bar{n})$, the problem is solved numerically.

Given a measured count quartet \bar{n} , the simplified likelihood function $\ell(\bar{r}|\bar{n})$ is evaluated in successive grid searches that approximate to the maximum by subsequently reducing the grid spacing.

The performance of the MLE compared to the CRB is shown in fig. S9A. Different separations L and multiple molecule positions on the x axis, given a signal-to-background

ratio of $SBR = 10$, are depicted. At the origin, the performance of the estimator converges to the CRB for about 100 photons for all shown beam separations L . The further away the molecule is, the more photons are needed for the MLE to achieve the CRB. For $x = 50$ nm, $y = 0$ nm and $L = 75$ nm for example, the MLE converges only starting from about $N = 500$ photons.

3.1.3 MLE for 1D position with two Gaussian-shaped beams

The maximum likelihood estimator for the case of two Gaussian shaped excitation profiles (S19) can be calculated following the same procedure used in section 3.1.1. The parameter vector \bar{p} is:

$$p_0 = \frac{e^{\frac{4 \ln 2 xL}{fwhm^2}}}{2 \cosh \frac{4 \ln 2 xL}{fwhm^2}}, \quad p_1 = \frac{e^{-\frac{4 \ln 2 xL}{fwhm^2}}}{2 \cosh \frac{4 \ln 2 xL}{fwhm^2}} \quad (S38)$$

and leads to the following likelihood function $\mathcal{L}(x|\bar{n})$:

$$\mathcal{L}(x|\bar{n}) = \frac{N!}{n_0! n_1!} \prod_{i=0}^1 p_i^{n_i} = \frac{N!}{n_0! (N - n_0)!} \frac{e^{4 \ln 2 \frac{Lx(2n_0 - N)}{fwhm^2}}}{\cosh \left(\frac{4 \ln 2 xL}{fwhm^2} \right)^N} \quad (S39)$$

Differentiation of this likelihood function yields the maximum likelihood position estimator \hat{x}_m^{MLE} :

$$\begin{aligned} \frac{d}{dx} \mathcal{L}(\hat{x}_m^{MLE} | \bar{n}) &= 0 \\ \Rightarrow \hat{x}_m^{MLE}(n_0, N) &= \frac{fwhm^2}{8 \ln(2) L} [\ln(n_0) - \ln(N - n_0)] \\ \hat{x}_m^{MLE}(n_0, N) &= \frac{fwhm^2}{8 \ln(2) L} \ln \left(\frac{n_0}{n_1} \right) \end{aligned} \quad (S40)$$

Thus, the MLE takes a surprisingly simple form which is easily implemented for live position estimation in the FPGA board. From eq. (S40), we infer that position estimation makes sense only if at least one photon is collected from each exposure. In the case of $n_i = 0$, the emitter would be located at infinity. In the presented tracking application, these photon combinations are neglected by defining $\hat{x}_m^{MLE}(n_i = 0|N) \triangleq 0$.

It is of particular interest not only to compute the ML-estimate, but also the localization bias and precision for different emitter positions x_m . They can be calculated as:

$$\text{bias}_{MLE}(x_m) = \langle \hat{x}_m^{MLE} \rangle - x_m \quad (S41)$$

$$\tilde{\sigma}_{MLE}^2(x_m) = \sum_{n_0=0}^N (\hat{x}_m^{MLE} - \langle \hat{x}_m^{MLE} \rangle)^2 P(n_0|N) \quad (\text{S42})$$

$$\text{with } \langle \hat{x}_m^{MLE} \rangle = \sum_{n_0=1}^{N-1} \hat{x}_m^{MLE} P(n_0|N)$$

Both quantities are visualized in fig. S4B,D for different photon numbers N . It is evident that, for a given L and $fwhm$ combination, a minimal N is needed in order to have an acceptable bias in a given field of view around the origin. The standard deviation $\tilde{\sigma}_{MLE}(x_m)$ asymptotically approaches the CRB $\tilde{\sigma}_{CRB}(x_m)$ with growing N in that region. Consequently, the localization precision reproduces (i) the $\propto L^{-1}$ dependence of eq. (S23b) and (ii) the faster divergence of the $\tilde{\sigma}_{CRB}(x_m)$ with increasing beam separation L . However, the standard deviation $\tilde{\sigma}_{MLE}(x_m)$ falls below the CRB at positions where the bias is not negligible. This is acceptable, as the CRB is a lower bound for unbiased estimators only.

3.2 Other estimators

In addition to the MLE, we investigated further position estimators. The single emitter tracking application, which employs the excitation beam pattern (EBP) introduced in section 2.2, requires an estimator that is suitable for low photon numbers, as well as for live position estimation in the FPGA board. In particular, the position estimation needs to be fast enough, such that the molecule does not leave the high sensitivity region surrounding the EBP origin.

We start this section presenting the solution of the least-mean-squared (LMS) estimator, for a linearization of the problem around the origin. We then present a modified LMS estimator (mLMSE) which has a better performance, but keeps the computational simplicity. Finally, we present the position estimator used in post processing, which is a numerically unbiased version of the mLMSE.

\bar{r}	Molecule position
$\bar{p}(\bar{r})$	Multinomial parameter
\bar{n}	Collected counts
$\hat{p}(\bar{n})$	Estimator of the multinomial parameter
$\hat{r}(\hat{p}(\bar{n}))$	Estimator of the molecule position
$\bar{R}(\bar{r}) = E(\hat{r})$	Expectation of the position estimator

3.2.1 Linearized least mean squared (LMS) estimator

The vicinity of the EBP origin is the area with the lowest CRB values and also the place in which the molecule is kept in the tracking application. Therefore, it makes sense to construct an estimator that is especially suited for that region. A first order approximation of the parameter vector \bar{p} (see eq. (S4)) at the beam pattern origin is used:

$$p_i(\bar{r})|_{r=0} \cong p_i(\bar{0}) + \sum_{j=1}^d r_j \frac{\partial p_i}{\partial r_j} \quad (\text{S43})$$

where d depicts the dimensionality of the \bar{r} -space and $\bar{0}$ is its null vector.

Given a set of measured counts \bar{n} , the position estimate of the molecule \hat{r} can be obtained by making use of the invariance property of the MLE:

$$\bar{p}(\hat{r}) = \hat{p} \quad (\text{S44})$$

where \bar{p} is the parameter vector defined in eq. (S43) and \hat{p} is the MLE, which, in the case of a multinomial distribution, is given by:

$$\hat{p}_i = \frac{n_i}{\sum_j n_j} = \frac{n_i}{N} \quad (\text{S45})$$

In other words, the estimated \hat{p} values are mapped into the position space \bar{r} through the function $\bar{p}(\bar{r})$. A visualization of this concept for a one dimensional example can be found in fig. 1 of the main text.

Inserting eq. (S43) into eq. (S44) yields:

$$J\hat{r} = \hat{p} - \bar{p}(\bar{0}) \quad (\text{S46})$$

with $J = \bar{\nabla}\bar{p}$ the Jacobian matrix of the transformation from the \bar{r} -space to the \bar{p} -space (this is not J^* , the Jacobian of the \bar{r} -space to *reduced* \bar{p} -space transformation, introduced in eq. (S9)).

The solution of the overdetermined linear system in eq. (S46) is obtained by a least mean square projection. Consequently, the following cost function S needs to be minimized:

$$S = \|\hat{p} - \bar{p}(\bar{0}) - J\hat{r}\|^2 \quad (\text{S47})$$

The solution is given by:

$$\hat{r}_{LMS}(\hat{p}) = (J^T J)^{-1} J^T (\hat{p} - \bar{p}(\bar{0})) \quad (\text{S48})$$

For the two dimensional localization with four doughnut shaped excitation profiles introduced in section 2.2, we obtain:

$$\hat{r}_{LMS}(\hat{p}) = \frac{L}{2} \cdot \frac{1}{1 - \frac{L^2 \log(2)}{fwhm^2}} \begin{bmatrix} -\hat{p}_3 + \frac{1}{2}(\hat{p}_1 + \hat{p}_2) \\ \frac{\sqrt{3}}{2}(\hat{p}_2 - \hat{p}_1) \end{bmatrix} \quad (S49)$$

Comparing this solution with the beam positions \bar{r}_{b_i} defined in eq. (S24) we can rewrite eq. (S49) into:

$$\hat{r}_{LMS}(\hat{p}) = -\frac{1}{1 - \frac{L^2 \log(2)}{fwhm^2}} \sum_{i=1}^3 \hat{p}_i \bar{r}_{b_i} \quad (S50)$$

This is because the gradient of each function $p_i(\bar{r})$ points in the direction of the corresponding beam position \bar{r}_{b_i} .

3.2.2 Modified least mean squared (mLMS) estimator

The LMS solution in eq. (S50) takes a simple form that can be implemented in the FPGA board for live position estimation. Unfortunately, it does not make use of the photons n_0 collected from the first doughnut exposure, as the function $p_0(\bar{r})$ has no linear term in its polynomial expansion at the origin. These photons do not hold directional information but they do hold radial information. Especially in the vicinity of the EBP origin, an increase of the \hat{p}_0 value indicates an increase of the radial coordinate. To include this property in the position estimator, eq. (S50) is expanded in orders of \hat{p}_0 with parameters β_j :

$$\hat{r}_{mLMS}^{(k)}(\hat{p}, \bar{\beta}) = -\frac{1}{1 - \frac{L^2 \log(2)}{fwhm^2}} \left(\sum_{j=0}^k \beta_j \hat{p}_0^j \right) \sum_{i=1}^3 \hat{p}_i \cdot \bar{r}_{b_i} \quad (S51)$$

Note that an expansion in arbitrary orders $\hat{p}_0^k \hat{p}_1^l \hat{p}_2^m \hat{p}_3^n$ with the respective directional vectors could be conducted. This was not studied further in this work, though. In the case of the live position estimator in tracking, the estimator $\hat{r}_{mLMS}^{(k=1)}$ was used.

3.2.3 Numerically unbiased mLMS (numLMS) estimator

Though the mLMS is fast to calculate, it has the drawback of being biased. This is especially true for experimental beam shapes that might slightly deviate from their ideal counterparts (for which the mLMS was calculated). Consequently, the measured trajectories in the tracking application, which rely on the mLMS, have to be corrected in post processing. Unfortunately, the MLE introduced in section 3.1.2 does not converge to

the CRB in the photon range employed in tracking, where an average of $\langle N \rangle = 9$ photons were used per localization.

In comparison to the MLE, the mLMSE has the advantage that performance metrics such as the covariance and the bias can be calculated analytically. The knowledge of the bias enables the possibility to unbiased the estimator. We first present the definition and derivation of the analytical bias of the mLMSE, followed by the routine that finds the optimal numerically unbiased mLMSE (numLMSE).

Bias calculation

The bias of the mLMSE is given by:

$$\text{bias}_{mLMS}^{(k)}(\bar{r}, \bar{\beta}) = E\left(\hat{r}_{mLMS}^{(k)}(\hat{p}, \bar{\beta})\right)_{(\bar{r})} - \bar{r} \quad (\text{S52})$$

where the explicit dependence on the molecule position \bar{r} is written as a subscript. Though $\hat{r}_{mLMS}^{(k)}$ is a random variable that depends on the *individual realization* of the collected photon counts \bar{n} (eq. (S51)) and the expansion parameter $\bar{\beta}$, its expectation depends on the actual molecule position.

The expectation of the position estimator $\bar{R}_{mLMS}^{(k)} \triangleq E\left(\hat{r}_{mLMS}^{(k)}\right)_{(\bar{r})}$ is

$$\bar{R}_{mLMS}^{(k)}(\bar{r}, \bar{\beta}) = -\frac{1}{1 - \frac{L^2 \log(2)}{fwhm^2}} \sum_{j=0}^k \sum_{i=1}^3 \beta_j E(\hat{p}_0^j \hat{p}_i) \cdot \bar{r}_{b_i} \quad (\text{S53})$$

A closed expression can be obtained by making use of the identity $E(\hat{p}_0^j \hat{p}_i) = E(n_0^j n_i) / N^{j+1}$ and the expression taken from (40) for the generalized factorial moments of the multinomial distribution. The mean localization $\bar{R}_{mLMS}^{(k=2)}$, for a molecule at position \bar{r} in the case that N photons were collected, is given by:

$$\bar{R}_{mLMS}^{(k=2)}(\bar{r}, \bar{\beta}) = E\left(\hat{r}_{mLMS}^{(k=2)}\right)_{(\bar{r})} = \frac{-1}{1 - \frac{L^2 \log(2)}{fwhm^2}} \sum_{i=1}^3 \left(\beta_0 p_i(\bar{r}) + \frac{(N-1)}{N} \left[\beta_1 + \frac{\beta_2}{N} \right] p_0(\bar{r}) p_i(\bar{r}) + \frac{\beta_2 (N-1)(N-2)}{N^2} p_0^2(\bar{r}) p_i(\bar{r}) \right) \bar{r}_{b_i} \quad (\text{S54})$$

The order k in eq. (S51) was found to increase the area (around the origin) in which unbiaseding is possible. This is related to the function $\bar{R}_{mLMS}^{(k)}(\bar{r}, \bar{\beta})$ being injective up to a maximal radius $|\bar{r}|$ only. The size of this radius depends on the choice of $\bar{\beta}$ and, in particular, of k . In this work, we chose $k = 2$ as a compromise, which enables to unbiased the mLMSE in a radius of about $L/2$ surrounding the origin.

Unbiasing

In order to unbiased the estimator $\hat{\bar{r}}_{mLMS}^{(k=2)}$, the following numerical optimization routine is employed:

1. Choose a value $\mathcal{R}_2 \in \mathbb{R}$, a set of positions $\bar{r} \in \mathbb{R}^2$ and a set of query points $\bar{r}_q \in \mathbb{R}^2$ with $|\bar{r}_q| \leq \mathcal{R}_2$. Minimize the loss function $\mathbb{L}(\mathcal{R}_1)$, with $\mathcal{R}_1 \in \mathbb{R}$, $\mathcal{R}_1 \leq \mathcal{R}_2$.

The loss function $\mathbb{L}(\mathcal{R}_1)$ is calculated as follows:

- a. Obtain optimal mLMSE in \mathcal{R}_1 .

Optimize $\bar{\beta}$ such that $\langle \text{bias}_{mLMS}^{(k=2)}(\bar{r}, \bar{\beta}) \rangle$ is minimized, where the average is taken over all positions \bar{r} , with $|\bar{r}| \leq \mathcal{R}_1$. Let the optimal $\bar{\beta}$ -vector be $\bar{\beta}_{opt}$.

- b. Unbias the optimal mLMSE by generating an interpolant function $\mathcal{F}_{\mathcal{R}_1}$ such that:

$$\begin{aligned} \mathcal{F}_{\mathcal{R}_1}: \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ \mathcal{F}_{\mathcal{R}_1} \left(\bar{R}_{mLMS}^{(k=2)}(\bar{r}, \bar{\beta}_{opt}) \right) &= \bar{r}, \quad \forall \bar{r}: |\bar{r}| \leq \mathcal{R}_1 \end{aligned} \quad (\text{S55})$$

- c. Test the generated unbiased estimator.

- i. For each query point \bar{r}_q , generate a set $\hat{P}_{\bar{r}_q} = \{\hat{p}_0, \dots, \hat{p}_M\}_{\bar{r}_q}$ of M \bar{p} -parameter estimates, and calculate the mLMSE $\hat{\bar{r}}_{mLMS}^{(k=2)}(\hat{p}, \bar{\beta}_{opt})$ for all \hat{p} -vectors.

- ii. Calculate the mean square error (MSE):

$$\begin{aligned} MSE_{\hat{p}_{\bar{r}_q}} &= \frac{1}{M} \sum_{i=1}^M \left(\mathcal{F}_{\mathcal{R}_1} \left(\hat{\bar{r}}_{mLMS}^{(k=2)}(\hat{p}_i, \bar{\beta}_{opt}) \right) - \bar{r}_q \right)^2 \\ &\text{with } \hat{p}_i \in \hat{P}_{\bar{r}_q} \end{aligned} \quad (\text{S56})$$

- iii. Calculate the loss function $\mathbb{L}(\mathcal{R}_1)$:

$$\mathbb{L}(\mathcal{R}_1) = \left\langle \left| \sqrt{MSE_{\hat{p}_{\bar{r}_q}}} - CRB(\bar{r}_q) \right| \right\rangle \quad (\text{S57})$$

where the average is taken over the set of query points \bar{r}_q .

2. Let the region \mathcal{R}_1 minimizing the loss function \mathbb{L} be \mathcal{R}_{opt} . The numerically unbiased estimator is then given by:

$$\hat{r}_{numLMS}^{(k=2)}(\hat{p}) = \mathcal{F}_{\mathcal{R}_{opt}} \left(\hat{r}_{mLMS}^{(k=2)}(\hat{p}, \bar{\beta}_{opt}) \right) \quad (\text{S58})$$

In this work we chose $\mathcal{R}_2 = \frac{L}{2}$. The set of positions \bar{r} was a rectangular grid with spacing of 1nm. The query points were equally spaced positions on the x and the y axis with $x, y \leq L/2$, respectively. To calculate $\bar{\beta}_{opt}$ the optimization function “fmincon” of Matlab was used, where β_1 was set to $\beta_1 = 0$ per default. As interpolant \mathcal{F} , the “scatteredInterpolant” function from Matlab was employed. The set size M was chosen to be $M = 10^4$. The loss function was optimized using a grid search. Note that the mean localization $\bar{R}_{mLMS}^{(k=2)}(\bar{r}, \bar{\beta})$ depends on N and on the vector parameter \bar{p} (eq. (S53)), and that \bar{p} depends on the SBR (eq. (S30)). To reduce the computation time of the numLMS calculation, the parameters \mathcal{R}_{opt} and $\bar{\beta}_{opt}$ were calculated only for some N and SBR values. Estimates of the optimal parameter for other N and SBR combinations were then inferred by 2D interpolation.

The localization performance of the MLE, compared to the numLMSE, is depicted in fig. S9. It can be seen that the MLE converges to the CRB for $N \gtrsim 100 - 500$ photons only, depending on the emitter position \bar{r} and on L . For smaller N , the MLE deviates considerably from its information theory limit. In comparison, the divergence of the numLMSE from the CRB is strongly reduced. This is especially true for low photon numbers N . Note that the pictured numLMSE was designed to work in a region $|\bar{r}| \lesssim L/2$. The dependence of the localization performance as a function of the SBR is not explicitly shown, but it was found to performs well for the evaluated range of $SBR = 2 - 60$.

4 Camera localization

Camera localization is a widely employed method to image or track molecules labeled with a fluorescent emitter. In this work, the localization performance of MINFLUX is frequently compared to the information theory limit and the CRB or MLE performance of camera localization. In this section, both quantities are calculated for an ideal camera, neglecting influences of gain, readout noise and motion blurring.

4.1 Static localization precision

When a static single point-like emitter is imaged onto a camera, the collected photons generate an intensity distribution on the camera that is described by the point spread function (PSF) of the imaging system. The position of the emitter can be estimated with substantially better precision than the spread of the PSF.

For characterization of the achievable localization precision, we calculate the Fisher information and Cramér-Rao bound on the localization precision for a perfect camera. Currently available cameras have additional sources of noise (e.g. read noise, EM excess noise) which make the localization precision worse. Taking all these noise sources into account complicates the model and become less relevant as camera technologies are improved. The CRB for camera localization is derived in the same way as presented in section 2, with the parameter vector \bar{p} being the photon detection probabilities in each pixel of the camera (or rather a small camera region [e.g. $K = 9 \times 9$ pixels] around the emitter, where the fluorescence photons will be concentrated). The entries p_i of the parameter vector \bar{p} are given by the probability of detecting a photon in pixel i of the camera. Approximating the PSF by a symmetric Gaussian function with the width σ_{PSF} yields

$$I_{PSFgauss}(x, y) = A_{gauss} e^{-\frac{1}{2} \left(\frac{x-x_m}{\sigma_{PSF}} \right)^2} e^{-\frac{1}{2} \left(\frac{y-y_m}{\sigma_{PSF}} \right)^2} \quad (S59)$$

with the emitter at position (x_m, y_m) . For a square shaped pixel with side length a and center at (x_i, y_i) we get:

$$\begin{aligned} p_i^{(0)}(x_m, y_m) &= \frac{\int_{x_i - \frac{a}{2}}^{x_i + \frac{a}{2}} \int_{y_i - \frac{a}{2}}^{y_i + \frac{a}{2}} I_{PSFgauss}(x, y) dx dy}{\int_{x_{min} - \frac{a}{2}}^{x_{max} + \frac{a}{2}} \int_{y_{min} - \frac{a}{2}}^{y_{max} + \frac{a}{2}} I_{PSFgauss}(x, y) dx dy} \\ &\approx \frac{1}{4} \left(\operatorname{erf} \left(\frac{x_i + a/2 - x_m}{\sqrt{2} \sigma_{PSF}} \right) - \operatorname{erf} \left(\frac{x_i - a/2 - x_m}{\sqrt{2} \sigma_{PSF}} \right) \right) \\ &\quad \cdot \left(\operatorname{erf} \left(\frac{y_i + a/2 - y_m}{\sqrt{2} \sigma_{PSF}} \right) - \operatorname{erf} \left(\frac{y_i - a/2 - y_m}{\sqrt{2} \sigma_{PSF}} \right) \right) \end{aligned} \quad (S60)$$

with the emitter position (x_m, y_m) and the coordinates of the camera pixels ranging from (x_{min}, y_{min}) to (x_{max}, y_{max}) . Eq. (S60) is valid as long as the full PSF (eq. (S59)) is

covered by the K pixel-sized region of the camera. The superscript (0) indicates that no background contributions have been taken into account so far. To include Poissonian background contributions we define a camera signal-to-background ratio:

$$SBR_c = \frac{\lambda_{signal}}{\lambda_{background}/K} \quad (S61)$$

where λ_{signal} is the average total fluorescence photons, $\lambda_{background}$ is the mean total background counts and K is the number of pixels of the corresponding camera region. Using this definition, we get the following success probability \bar{p} :

$$p_i(x, y) = \frac{1}{K + SBR_c} + \frac{SBR_c}{K + SBR_c} \cdot p_i^{(0)}(x, y) \quad (S62)$$

Inserting $p_i^{(0)}$ and solving the integral yields:

$$p_i(x_m, y_m) = \frac{1}{K + SBR_c} + \frac{SBR_c}{K + SBR_c} \cdot \frac{1}{4} \cdot \left(\operatorname{erf}\left(\frac{x_i + a/2 - x_m}{\sqrt{2} \sigma_{PSF}}\right) - \operatorname{erf}\left(\frac{x_i - a/2 - x_m}{\sqrt{2} \sigma_{PSF}}\right) \right) \cdot \left(\operatorname{erf}\left(\frac{y_i + a/2 - y_m}{\sqrt{2} \sigma_{PSF}}\right) - \operatorname{erf}\left(\frac{y_i - a/2 - y_m}{\sqrt{2} \sigma_{PSF}}\right) \right) \quad (S63)$$

with the pixel center coordinate (x_i, y_i) of pixel i , the emitter position (x_m, y_m) , the pixel size a , the number of pixels K and the camera signal-to-background ratio SBR_c . The Fisher information and CRB can be calculated using eq. (S11) and eq. (S13), respectively. CRBs with relevant background levels are shown in fig. S6A.

The CRB gives a lower limit on the variance of any unbiased estimator but it does not describe the actual performance of an estimator. Therefore, we investigated the performance of the MLE for single emitter position estimation with a camera. The MLE is known to be asymptotically unbiased and efficient (i.e. it converges to the CRB). The likelihood function is given by

$$\mathcal{L}(x|\bar{n}) = \frac{N!}{\prod_{j=0}^{K-1} n_j!} \prod_{i=0}^{K-1} p_i^{n_i} \quad (S64)$$

where p_i is the probability of detecting a photon in pixel i as given by eq. (S63) evaluated at (x, y) and $\{n_i\}$ is the set of collected photons in the pixels of the camera and the total number of detected photons $N = \sum_{i=0}^{K-1} n_i$. The performance of the MLE on simulated images is shown in fig. S6A. Deviations of the MLE from the CRB are visible for low N and SBR_c . This is in agreement with the asymptotic convergence of the MLE to the CRB. A representative collection of simulated images for different SBR_c and N is shown in fig. S6B-F. It should be stressed that the MLE requires precise knowledge of the PSF shape and noise of the system (especially the camera) (41).

In this work, the following parameters were used for calculation of the localization performance using a camera: width of the detection PSF σ_{PSF} : 100 nm for static localization of a single emitter (fig. 3) and imaging (fig. 4), and 87 nm for tracking (fig. 5); pixel size of the camera a : 100 nm; camera ROI around emitter: 9×9 pixels, i.e. $K = 81$; signal-to-background ratio SBR_c : 50 or 500 for static localization of a single emitter (fig. 3), 500 for imaging (fig. 4), and 20 or 40 for tracking (fig. 5).

5 Diffusion coefficient estimation

Different algorithms to estimate the apparent diffusion coefficient D , as well as the localization precision σ , from single-particle trajectories can be found in the literature. This section starts by giving a brief overview and motivates the choice of an optimized least-square fit (OLSF) diffusion estimation (32) for this work. In the second part, it is evaluated whether the OLSF is applicable to the trajectories obtained using MINFLUX by using computer simulations. It is found that the OLSF from (32) needs to be slightly adapted. Doing so enables D estimation with a performance close to its information-based theoretical limit. Furthermore, the OLSF extracts the tracking error ϵ from the measured trajectory.

5.1 Algorithm overview

A maximum likelihood estimation (MLE) approach for the diffusion, based on the statistics of the observed molecule displacements, was developed in (42). In this approach, the calculation of the exact value of the likelihood function can be very time consuming, especially for trajectories with large number of positions S . An approximation based on the discrete Fourier transform was employed, rendering the MLE computationally more efficient. In the case that molecular blinking is present, some time-lapse displacements are missing in the particle trajectory. In that situation, the MLE can still be used, as shown in (43). Unfortunately, non-equally spaced time-lapse displacements do not allow the use of the mentioned approximation. This renders the MLE time consuming, especially for long trajectories. In this work, blinking as well as long trajectories are present, such that the MLE could not be employed.

Nevertheless, the likelihood function from (42) enables the computation of the CRB for the D and σ^2 parameter estimation. For $S \gg 1$, the theoretical limit on the relative standard deviation σ_D/D is given by:

$$\frac{\sigma_D}{D} \geq \sqrt{\frac{2}{d(S-1)}} \cdot [1 + 2\sqrt{1 + 2X}]^{\frac{1}{2}} \quad (\text{S65})$$

where d is the dimensionality of the trajectory space. Here $X = \sigma^2/D\Delta t - 2R$ is the *reduced square localization error*, where Δt is the time interval between trajectory points and R is the motion blur coefficient, a dimensionless variable characterizing the duration and type of the emitter exposure:

$$R = \frac{1}{\Delta t} \int_0^{\Delta t} I_p(t) (1 - I_p(t)) dt \quad (\text{S66})$$

where $I_p(t)$ is the illumination percentage occurring between the start of the exposure ($t = 0$) until time t . In the case of uniform exposure during the whole time interval Δt , R

is $1/6$, for the multiplex cycle employed in MINFLUX tracking (see 6.4.3) it is $R \approx 1/6.2$.

Vestergaard et. al (44) proposed a regression-free covariance-based estimator (CVE) that is computationally very fast. It converges to the CRB of eq. (S65) for small reduced square localization errors $X \lesssim 1$. However, for $X \gtrsim 1$ the CVE departs considerably from the CRB. Given that this work employs time intervals of $\Delta t = 125\mu\text{s}$, the reduced square localization error exhibits large values $X \gg 1$ (for typical D and σ^2 values), such that the CVE is not well-suited.

Michalet and Berglund (32) developed an algorithm for an optimized least-square fit (OLSF) to the mean square displacement (MSD) curve. In the case of free isotropic Brownian Motion in d dimensions, for a lag s the MSD is given by:

$$\rho_s = 2d(\sigma^2 - 2RD\Delta t) + 2dDs\Delta t \quad (\text{S67})$$

The experimental MSD curve is fitted to this model, using time lags from $s = 0$ to a given value s . There is an optimal number points $s = s_{opt}$ to fit, in order to extract the diffusion coefficient estimate \hat{D} with maximum precision. The value of s_{opt} for D estimation depends solely on the *reduced square localization error* X and on the trace length S . It was shown that, if $s_{opt}(X, S)$ is employed, the OLSF performance approaches optimality for a wide parameter range. This, in addition to its computational simplicity, makes the OLSF a suitable candidate for extracting diffusion coefficients from the measured trajectories.

5.2 Applicability and performance of the OLSF estimator

In the presented single emitter tracking application, the emitter is blinking and large X values are present. Furthermore, the localization precision has a dependence on the relative position of the emitter to the beam pattern origin. This leads to non-uniform localization precisions for different trajectory points (see section 2.2 and fig. S3). Therefore, the prerequisites that allow the utilization of the $s_{opt}(X, S)$ relationship developed in (32) are not met. To evaluate whether the OLSF is still applicable, we resorted to an *in silico* evaluation.

5.2.1.1 Simulation

Two dimensional trajectories with 1200 localizations were generated for free isotropic Brownian motion with diffusion constants in the range of $D \in [0.01, 1] \mu\text{m}^2/\text{s}$ and total count rates of $\Gamma \in [20, 200]$ kHz. A total of 100 trajectories were generated for each parameter combination. Blinking on- and off-times were assumed to be exponentially distributed with parameters $t_{on} = 2$ ms and $t_{off} = 0.6$ ms, which are in agreement with what was found experimentally. The time sampling was set to $\Delta t = 125 \mu\text{s}$, where a 12-fold subsampling was used to account for motion blurring during

illumination. The excitation geometry was set to that described in section 2.2, with a separation of $L = 130$ nm. The *fwhm* of the excitation PSF (eq. (S17)) was set to $fwhm = 450$ nm. The signal-to-background ratio was set to $SBR = 3.7$ in the center of the EBP. Poissonian photon count realizations were drawn in every subsampling interval. Emitter localizations employed the live position estimator $\hat{r}_{mLMS}^{(k=1)}$ with $\beta_0 = 1.27$ and $\beta_1 = 3.8$. The chosen simulation parameters covered the range of the employed values in the single emitter tracking measurement.

In the case of blinking, a MSD curve can still be defined as

$$\langle \rho_s \rangle = \frac{1}{\sum_{i=1}^{S-s} q_i q_{i+s}} \sum_{i=1}^{S-s} (\bar{r}_{i+s} - \bar{r}_i)^2 q_i q_{i+s} \quad (\text{S68})$$

with

$$q_i = \begin{cases} 1 & \text{emitter on at sample } i \\ 0 & \text{emitter off at sample } i \end{cases} \quad (\text{S69})$$

In order to identify blinking states (on and off) from the simulated tracks, the Viterbi algorithm was employed as explained in section 6.4.4. The raw trajectory \hat{r} of the emitter is estimated with the live position estimator $\hat{r}_{mLMS}^{(k=1)}$ defined in eq. (S51). This estimator is biased and works well in a region surrounding the beam pattern origin, as explained in section 3.2.3. This causes the MSD curves to deviate from the expected linear relationship of eq. (S67). Therefore, the raw trajectories are corrected in post processing using the numLMS estimator (see section 3.2.3).

5.2.1.2 MSD curves

Figure S10A visualizes both, the MSD for a raw trajectory as well as the MSD for the corrected trace. It can be seen that the linear behavior of eq. (S67) is retrieved after the correction. At the same time, the offset of the corrected trace is higher, which suggests that the localization error σ is increased after the numLMS correction. This is counterintuitive at first glance. The reason is related to the finite working region of the mLMSE. Outside of it, position estimations are biased, especially radially, which reduces the spread of localizations.

5.2.1.3 Optimal fit length

For each numLMS-corrected trajectory, the MSD curve (eq. (S68)) was fitted to eq. (S67) for different lengths $s \in [2, 1200]$ (logarithmically spaced). Subsequently, the s values minimizing σ_D/D were identified and found to follow the relationship:

$$s_{opt}(X) = 2 + 4.6X^{0.62} \quad (\text{S70})$$

The identified s_{opt} values as well as eq. (S70) are visualized in fig. S10C.

Given that the X value of a trajectory is not known a priori, the recursive algorithm proposed in (45) is used:

Start by setting $s_0 = S/10$ and extract $\hat{\sigma}_1$ and \hat{D}_1 from eq. (S67). These estimates allow the calculation of a reduced square localization error estimate \hat{X}_1 . Using the latter, a new length s_1 can be obtained from eq. (S70). Repeating this procedure results in a fast convergence to the optimal length value. Non-convergence is an exceptional case, and can be spotted by reoccurrences of s -values (i.e. optimization is trapped within the same two values). (S71)

This algorithm was applied to trajectories with lengths S of 100, 1200 and 5000 localizations, respectively. Note that in case that $s_i > S$, the number of points used in the MSD fit is set to $s_i = S$. The parameter choices (i.e. D, Γ, \dots) were equal to the ones stated before in ‘‘Simulation’’. After initialization of the recursive algorithm (using $s_0 = S/10$), a total of 10 iterations were conducted. Figure S10B visualizes the median and the standard deviation of the values $s(\hat{X}_{11})$ for the three respective trace lengths. They are plotted against the estimation \hat{X}_{12} to visualize the convergence (i.e., all the points are next to the optimal curve of eq. (S70)).

5.2.1.4 Diffusion coefficients

The apparent diffusion coefficients were extracted using the last fit of the recursion. Subsequently, the variance σ_D^2 was calculated from the 100 repetitions of each parameter combination. The resulting values of σ_D/D are visualized in fig. S10D. Although it is close to the CRB, the estimator is not optimal and could be improved. It deviates from its theoretical limit (eq. (S65)) by a factor of about 1.4. It should be noted that this CRB corresponds to the non-blinking case and should only be taken as a rough guide. The derivation of the likelihood function (and CRB) for the parameters of a trajectory of a blinking emitter surpasses the scope of this work.

5.2.1.5 Estimation of the localization precision

Equation (S67) does not only permit to extract the diffusion coefficient, but also the localization precision σ . Figure S10E shows that the estimated localization precision $\hat{\sigma}$, obtained from eq. (S67), coincides very well with the ground truth tracking error ϵ (see section 6.4.6). Note that in (32) it is shown that the optimal trajectory length s_{opt} for diffusion coefficient D estimation differs from the one for localization precision σ estimation. Nevertheless, the $[S, N]$ -parameter region spanned in this work is well suited for precise σ estimation.

In conclusion, the OLSF is well suited to extract the diffusion coefficient D of trajectories with (i) blinking, (ii) non-uniform localization precisions and (iii) large values of the *reduced square localization error* X for the trajectory lengths that are present in this work.

6 Materials and methods

6.1 PSF and parameter vector \bar{p}

The PSF of the employed optical system was measured by scanning the excitation beam over a small fluorescent microsphere and detecting the emitted fluorescent photons. The sample consisted of immobilized fluorescent microspheres and immobilized Au nanorods to allow for a live drift measurement and compensation. Live drift measurement and compensation, by repositioning the piezo stage, was done with a custom LabView program, as described in section 6.5.2.

The measurement procedure is as follows:

- A fluorescent microsphere is placed at the center of the beam scanners by moving the manual and piezo stage.
- The stage lock system is set to lock the position of a nearby Au nanorod in three dimensions.
- 100-500 frames are recorded, in order to allow for additional drift correction in post processing, and also compensation of fluorescent brightness changes of the microsphere during the measurement (i.e. blinking and photobleaching).

Scanning is performed with the non-descanned electro-optical beam scanners. Typically, a pixel size of 1 nm and a dwell time of 40 μ s are used. The laser power is set to about 1 μ W in the back focal plane of the objective.

ANALYSIS

A functional expression for the relevant area of the experimental PSF is obtained from a fit of a model function to the centered frames.

Centering. Centering (and correction for residual sample drift) is performed by finding the center of the excitation beam in each frame, and shifting the frames with respect to each other, such that their centers overlap.

Second order 2D polynomial functions are fitted to the individual frames. The positions of the minimum of the fitted functions serve as estimates for the centers, as the doughnut beams in use have zeros of intensity at their center. The obtained center positions are smoothed in time (by using a moving average filter of typically 10 frames) to reduce the influence of noise. The frame coordinates are then shifted accordingly – rounded to integer pixel units, to avoid any interpolation of Poissonian counts. Finally, the centered frames are averaged, providing the (single) image of the PSF, which will be fit to a model function.

Fitting. For both imaging and tracking applications, the excitation beam pattern is placed on the fluorescent emitter, such that it is in proximity of the beam center.

Therefore, a functional expression describing the beam shape in the central region is sufficient for any further analysis. The PSFs were fitted in a region of $200 \times 200 \text{ nm}^2$ (or $300 \times 300 \text{ nm}^2$ for EBP size $L > 100 \text{ nm}$) around the beam center. In this region, the PSF has mainly a parabolic profile. Deviations from a parabolic profile were accounted for by fitting a 2D polynomial up to order 4

$$I_{PSF}(\vec{r}) = I_{PSF}(x, y) = \sum_{i=0}^4 \sum_{j=0}^i a_{ij} (x - x_0)^i (y - y_0)^{i-j} \quad (\text{S72})$$

with the coefficients $\{a_{ij}\}$, x_0 and y_0 . Figure S7A-C shows a typical measured PSF and the fitted model function.

The parameter functions p_i , corresponding to a set of 4 doughnut shaped exposures, are obtained from the fitted PSF functions in a straight forward way. For displacements of the four doughnut beams as given by eq. (S24), the four beams can be written as

$$I_i(\vec{r}) = I_{PSF}^*(\vec{r} - \vec{r}_{b_i}) \quad (\text{S73})$$

where I_{PSF}^* is an offset subtracted version of eq. (S72). This intensity is used to calculate the parameter vector $\vec{p}^{(0)}$, as defined in eq. (S4). For the localization of each molecule, its empirical SBR (eq. (S29)) has to be used in combination with $\vec{p}^{(0)}$ to calculate the parameter vector \vec{p} , as defined in eq. (S30). The four components of the parameter vector $\vec{p}^{(0)}$ for a beam separation of $L = 50 \text{ nm}$ are shown in fig. S7D-G.

6.2 Localization performance map

The experimental localization performance was characterized by repeatedly estimating the positions of single ATTO 647N molecules in ROXS buffer. The measurements were performed at different (known) positions with respect to the excitation beam pattern (EBP), in order to obtain the spatial dependence of the standard deviation and bias of the position estimation.

For characterizing the localization performance in a region of size L (excitation beam separation), the multiplexed EBP was scanned in a 35×35 pixel grid (referred to as “frame” in the following) with a pixel spacing of 3 nm (fig. 3A-D). For characterization of the localization error at the center of the EBP (fig. 3E), we scanned the EBP in a 9×9 grid with a spacing of 2 nm with a change of L after a full 9×9 grid scan.

In both cases, the scan region was centered on a single ATTO 647N molecule and scanning was performed with the electro-optical beam scanning system using the 125 kHz bandwidth amplifiers. To ensure a stable beam position, multiplexing of the beam position was performed at 5 kHz with a waiting time of $20 \mu\text{s}$ after each beam

movement. The average power of the modulated excitation laser was set to about 20 μW resulting in a mean of about 0.3 to 2 collected photons per multiplexing cycle (dependent on L and on the distance from EBP center).

ANALYSIS

Obtaining the localization performance from the measured data requires a few processing steps. First, the recorded frames have to be centered. Due to the low number of photons recorded in each frame, noise is reduced by applying a sliding average filter on the data. The sliding window was 31 frames for the 35×35 pixel frames and 61 frames for the 9×9 pixel frames. A 2D polynomial function is fitted to the filtered frames to retrieve the beam center. In a next step, the individual unfiltered frames are shifted laterally to have their centers aligned. Shifting was performed in integer pixel units, to avoid combining or interpolating photon counts. The sum of all centered frames was used to obtain the experimental PSF. The PSF was modeled by a 2D polynomial function up to order four given by eq. (S72), whose coefficients were obtained from a fit to the data.

For a quantification of the localization performance for a specific number of detected photons, the data in each pixel is binned such that the photon number of interest is achieved. In this step, it is assured that every data point is only used once. Hence, many independent measurements with the same number of photons are obtained. This allows to localize the emitter many times using the same number of photons. Localization was performed with both, the MLE (see section 3.1.2) and the numLMSE (see section 3.2.3). Figure S8 shows the experimental localization performance like in fig. 3 but for an extended range of detected photons and beam separations L .

The experimental localization performance is characterized by the mean 1D error defined as:

$$\sigma_{exp}^{(1D)} = \sqrt{\frac{1}{2M} \sum_{i=1}^M \left((\hat{x}_m^{(i)} - x_0)^2 + (\hat{y}_m^{(i)} - y_0)^2 \right)} = \sqrt{\frac{1}{M} \sum_{i=1}^M (\varepsilon^{(i)})^2} \quad (\text{S74})$$

with the true emitter position (x_0, y_0) , the number of independent localizations M and the estimated emitter position $(\hat{x}_m^{(i)}, \hat{y}_m^{(i)})$. The uncertainty of $\sigma_{exp}^{(1D)}$ (i.e. the $\pm 1\sigma$ error bars in fig. 3D,E) was obtained by:

- Taking the sample covariance of the set of points $(\hat{x}_m^{(i)}, \hat{y}_m^{(i)})$.
- Assuming it to be the parameter of a bivariate Gaussian distribution.

- Calculating the cumulative radial integral, centered on (x_0, y_0) , and its derivative with respect to $(\sigma_{exp}^{(1D)})^2$, thus obtaining the distribution for $(\varepsilon^{(i)})^2$.
- Convolution M times this distribution with itself, to obtain the distribution of $(\sigma_{exp}^{(1D)})^2$ and through a variable transformation, we obtain the distribution of $\sigma_{exp}^{(1D)}$.
- The reported uncertainty of $\sigma_{exp}^{(1D)}$ is the width of that distribution, which is fairly symmetric for $M > 10$.

6.3 MINFLUX nanoscopy

6.3.1 Data acquisition

In order to acquire an image of small objects, the excitation beam pattern (EBP) has to be centered on the object of interest. This is performed by first taking a faint widefield image in which bright diffraction limited spots are selected manually. The centers of the bright spots are obtained by Gaussian fitting. Next, the EBP and detection volume are placed onto a spot center. Fine adjustment of the center of the four doughnut shaped excitation profiles is performed by a PI feedback loop that estimates the emitter location with the mLMS estimator (section 3.2.2). This adjustment is performed with the piezoelectric tip-tilt mirror. Typically, a beam separation of $L = 100$ nm, a multiplex cycle rate of 8 kHz, and 10 – 20% of the laser intensity applied during imaging was used. After a short time (~ 500 ms), the feedback loop is turned off and the parameters are changed to the parameters for the actual imaging measurement (e.g. laser power, beam separation L). The count traces for the four multiplexed excitation beams are then recorded. After the molecules in the illuminated region are bleached or enough data is recorded, the next region of interest is measured. The EBP is placed onto the next selected bright spot in the widefield image, and the whole procedure is repeated.

Separation of fluorescent molecules within the area illuminated by the four doughnut EBP is performed by stochastic switching of the molecules between fluorescent and non-fluorescent state. The off-switching of Alexa Fluor 647 is mainly driven by the excitation light of 642 nm. Transitions from the non-fluorescent state to the fluorescent state can be induced by illuminating with activation light of 405 nm wavelength. For minimizing both, the time with no molecule emission and the probability of having more than one molecule emitting at the same time, a conditional activation scheme was implemented. The 405 nm laser is switched on or off depending on the lowpass-filtered fluorescent count rate: it is switched on when the total count rate is below a threshold (~ 20 kHz) for a certain time (~ 50 ms), and it is turned off as soon as the count rate is above the threshold.

For the imaging, multiplexing of the EBP was performed with a period of 125 μs with a gate length of 13.6 μs and a gate delay of 13.6 μs (see fig. S5). The excitation laser power was 300 μW for measurements with $L = 70$ nm and 400 μW for measurements with $L = 50$ nm. The power of the 405 nm activation laser was set to about 1-2 μW . All powers were measured close to the back focal plane of the objective.

The process of obtaining an image from recorded count traces requires a few analysis steps which are described in the following.

6.3.2 Trace segmentation

Fluorescence dye molecule emission events have to be extracted from the recorded count traces. For this purpose, a simple hidden Markov model (HMM) for the sum of the fluorescence signal of the four exposures is employed. The model assumes three emission states (state 1: background emission; state 2: emission from a single emitter; state 3: emission from more than one emitter). State trajectory estimation is performed by applying the Viterbi algorithm.

The emission probability distribution of the three states are estimated in a two-step process. First, a mixture of Poisson distributions is fitted to the measured distribution of total count trace binned to 1 ms bins. The emission probability distribution is approximated by a Poisson distribution for the background with mean of λ_1 , a Poisson distribution for one emitter with a mean of λ_2 and a Poisson distribution for two emitters with a mean of $\lambda_3 = 2\lambda_2$. To refine the emission distributions, an estimation of the most probable state path of the unbinned total count trace is performed using the Viterbi algorithm. We used the Matlab implementation “hmmviterbi” which is part of the Statistics Toolbox. The transition probability matrix (T_{ij}) for a transition from state i to state j is estimated to:

$$T = \begin{pmatrix} 0.9995 & 0.0005 & 0 \\ 0.00025 & 0.9995 & 0.00025 \\ 0 & 0.00025 & 0.99975 \end{pmatrix} \quad (\text{S75})$$

From the most probable state path, an improved estimate for the emission distributions of the three states is obtained. This emission distribution is used in combination with the transition probability matrix (eq. (S75)) to re-run the Viterbi algorithm. The resulting estimation for the most probable state path is used to segment the count trace and extract single molecule emission events and the background level. Data points next to a state transition are discarded (as the emitter may have turned on during the four exposure cycle). Emission events with a total number of photons above a threshold, as indicated in fig. 4, are considered for further analysis. The trace segmentation for the measurement of the larger DNA origami (fig. 4A-G,M) is shown in fig. S11F-G. The histogram of detected photons per emission event in the measurement of the larger origami is shown in fig. S11K.

6.3.3 Localization of emission events

For each single emitter emission event above the total count threshold, the position of the emitter is estimated. The counts for the four beam exposures are accumulated and the emitter position is estimated from the resulting count quartet \bar{n} . The *SBR* is estimated from the data directly. From the HMM segmentation, the average background level λ_{bg} can be estimated very well. For each emission event, the *SBR* is estimated by

$$SBR = \frac{N}{\lambda_{bg} Q} - 1 \quad (S76)$$

where $N = \sum_{j=1}^Q \sum_{i=0}^{K-1} n_{ij}$ is the number of detected photons during the emission event of Q multiplex cycles duration (n_{ij} is the number of photons collected for exposure i in multiplex cycle j) and λ_{bg} is the average background counts per multiplex cycle ($K = 4$ exposures).

Position estimation is performed with the MLE, which is numerically obtained through a grid search, as described in section 3.1.2. Four successive 2D grids with spacings of 5 nm, 1 nm, 0.1 nm and 0.01 nm are used. The first grid spans a region of diameter 240 nm around the EBP center. An example of this grid search is shown in fig. S11E.

6.3.4 Clustering into nano-domains

In order to quantify the experimental localization precision, the estimated positions of the emission events are clustered into nano-domains. Clustering is performed with the k-means clustering algorithm with the squared Euclidean distance as distance metric. 100 sets of initial cluster centroid positions are used to avoid local extrema. Clustering was performed with a number of clusters between 1 and 12. The most likely number of clusters was determined by the highest mean cluster silhouette value.

6.3.5 Splitting events into equal number of photon events

A direct comparison between the experimental localization precision and its CRB requires the use of a fixed number of photons. For this purpose, each emission event is split into shorter emission events with a fixed number of photons. Splitting of the events is performed such that each beam multiplex cycle is used only once and that the total number of photons is exactly the number of interest.

6.3.6 Calculation of experimental covariance for fixed number of photons

The emission events split into equal numbers of photons are localized in the same manner as the original emission events. All localization events within a nano-domain are then used to calculate the *sample* covariance of the estimated emitter position of the nano-domain. This covariance is a measure for the experimental localization precision for a fixed number of photons at the position of the nano-domain and can be compared with

the CRB or the localization performance of single molecule localization using a camera. Details on the camera case can be found in section 4.

6.3.7 Drift correction in post processing

In addition to the active drift compensation with the system lock (see section 6.5.2) during the measurement process, a post processing drift compensation of the DNA origami imaging data can be performed. For post processing drift correction, a common drift function $\bar{x}_d(t)$ and the mean nano-domain centers \bar{x}_{0i} are fitted simultaneously to all localizations:

$$\bar{x}(i, t) = \sum_{j=1}^M \delta_{ij} \bar{x}_{0j} + \bar{x}_d(t - t_0) \quad (\text{S77})$$

with the Kronecker delta δ_{ij} , the number of nano-domains M , the mean time of the measurement t_0 , the time t and nano-domain identification i of an emission event. The drift function $\bar{x}_d(t)$ was approximated by a polynomial up to order 4. Figure S11H-J shows the post processing drift correction of the imaging of the smaller origami (fig. 4H-L,N). No post processing drift correction was applied to data of the larger origami (fig. 4C-G,M).

6.4 MINFLUX tracking

6.4.1 Tracking of predefined trajectory

Before tracking a diffusing molecule within a cell, we tested MINFLUX tracking against ground truth data by tracking predefined trajectories of a 20 nm fluorescent microsphere (see section 6.6.4). For this purpose, a known waveform was sent to the piezoelectric stage and MINFLUX tracking was used to retrieve the induced trajectory. As shown in fig. S14, we chose a Lissajous pattern. The Lissajous figure was realized by applying two sinusoidal translations along the x and y direction of the piezo stage. The peak-to-peak amplitudes were 500 nm and the frequencies were 1.5 Hz and 1 Hz in x and y , respectively. MINFLUX tracking was performed at 8 kHz multiplex and controller update rate. The employed beam separations were $L = 100$ nm (fig. S14A-B) or $L = 50$ nm (fig. S14C).

6.4.2 Tracking initiation: intensity thresholding and Gaussian localizations

In order to track molecules with the excitation beam pattern (EBP) introduced in section 2.2 (and fig. 2B), an initial localization is needed to position the EBP on top of the molecule. This is related to the mLMSE (see section 3.2.2 and 3.2.3) working only in a finite region around the EBP origin.

In a first step, a Gaussian 560 nm excitation beam with $fwhm \approx 320$ nm is scanned over a field of view of $1.9 \times 1.9 \mu\text{m}$. In order to activate mEOS2 molecules, a Gaussian 405 nm activation beam is applied at the center of the scanned region with a

period of 3 Hz and a pulse length of 1 ms. The pixel size of the scan was set to 127 nm and the pixel dwell time to 50 μ s. The scan is repeated until a photon count threshold of $N_{th} = 15$ is surpassed. In that case, the scanning routine as well as the activation are stopped, and a Gaussian localization is launched around the center of the corresponding pixel.

This localization is implemented by four Gaussian beam exposures. Two are placed at $x_{b_0} = -L/2$ and $x_{b_1} = L/2$, in order to estimate the x position of the molecule. The analogue is done with two further beams for the y position estimation. The beam separation L was set to $L = 300$ nm. The multiplex cycle timings (see fig. S5) were: 10 μ s of gate delay, 38 μ s of excitation and detection gate and 2 μ s for the localization window. Thus, the localization was carried out in 198 μ s. Live position estimation employs the MLE introduced in section 3.1.3. Using a number of photons around $N = 60$ ensures a localization precision of at least $\sigma \lesssim 30$ nm in a field of view of about 300 nm surrounding the Gaussian EBP origin. In the experimental conditions, a mean of $\langle N \rangle = 53$ was measured.

Subsequently, the MINFLUX tracking routine is launched (see section 6.4.3). When terminated, the tracking initiation routine is restarted. Note that an additional 405 nm activation timeout of 10 s was introduced, in order to reduce the probability of having multiple activated emitter present.

6.4.3 The MINFLUX tracking routine

This tracking routine uses 560 nm excitation and the EBP introduced in section 2.2. The multiplex cycle rate (see fig. S5) was set to 8 kHz, where a gate delay of 7.4 μ s, an excitation and detection gate of 23.2 μ s and a localization window of 2 μ s were employed. The EBP size was set to $L = 130$ nm. We used between 50-100 μ W excitation powers. Live position estimation is calculated using the mLMS estimator $\hat{r}_{mLMS}^{(k=1)}$ introduced in section 3.2.2. The $\bar{\beta}$ parameters are set to $\beta_0 = 1.27$ and $\beta_1 = 3.8$. Given that the mEOS2 proteins shows prominent blinking (see fig. 5C and S12), a reaction threshold of 6 counts is set for the total number of photons collected during each multiplex cycle. This is to prevent the EBP from being background-driven when the emitter is in a fluorescent off-state, as it would drift away. The threshold was set such that the probability of reacting to background counts is in the order of 2 %.

In order to enable single emitter tracking on a large field of view, a set of scanners and high voltage amplifiers are utilized as explained in section 6.5.1

6.4.4 Trace segmentation and blinking statistics

6.4.4.1 Trace segmentation

Measured trajectories consist of two segments. A time interval in which an emitter is tracked and, after bleaching or losing, a segment that is driven by background counts

only. The respective segments were isolated by manual inspection. To be accepted as a successful trajectory, the average central doughnut ratio $\langle n_0/\sum n_i \rangle$ needs to fulfill: $\langle n_0/\sum n_i \rangle < 0.23$, with n_i the counts obtained in the i^{th} doughnut exposure of a multiplex cycle. Note that in the case of no emitter being present, a value of $\langle n_0/\sum n_i \rangle = 0.25$ is expected. Trajectories with unusually long off-times (the emitter in an off-state on the order of 100 ms) were separated into two independent trajectories. Likewise, long events of unusually high count rates (≥ 150 kHz) with central doughnut ratio values ≥ 0.25 were cut out, to exclude the possibility of multiple emitters being present.

6.4.4.2 *Blinking statistics*

In order to identify molecule blinking events, a two stage hidden Markov model (HMM – see section 6.3.2) is employed on \bar{N} (the vector containing the total number of counts per multiplexing cycle for a whole trajectory).

Two emission states (state 1: emitter on, state 2: emitter off) are assumed. Estimation of the most probable state path employs the Matlab ‘hmmviterbi’ algorithm. In the first stage (noted by (0) superscript), the transition matrix is approximated as:

$$T^{(0)} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 1 - \frac{\Delta t}{\hat{t}_{on}^{(0)}} & \frac{\Delta t}{\hat{t}_{on}^{(0)}} \\ \frac{\Delta t}{\hat{t}_{off}^{(0)}} & 1 - \frac{\Delta t}{\hat{t}_{off}^{(0)}} \end{pmatrix} \quad (S78)$$

with $\hat{t}_{off}^{(0)} = 1$ ms and $\hat{t}_{on}^{(0)} = 3$ ms and Δt the sampling time of the trajectory. The emission probability distribution is assumed to be Poissonian for both states, with parameters $\hat{\lambda}_{on}^{(0)}$ and $\hat{\lambda}_{off}^{(0)}$:

$$\hat{\lambda}_{on}^{(0)} = \langle \bar{N} \rangle \frac{\hat{t}_{on}^{(0)} + \hat{t}_{off}^{(0)}}{\hat{t}_{on}^{(0)} + \frac{\hat{t}_{off}^{(0)}}{\overline{SBR}^{(0)} + 1}}; \quad \hat{\lambda}_{off}^{(0)} = \frac{\hat{\lambda}_{on}^{(0)}}{\overline{SBR} + 1} \quad (S79)$$

and $\overline{SBR}^{(0)} = 4$ is assumed. The resulting state path is then used to improve the *state emission* estimates. From the state path, two vectors containing the on and off times \bar{t}_{on} and \bar{t}_{off} are extracted. These times are assumed to follow truncated exponential distributions (as time intervals shorter than the sampling time are not collected). The updated parameter estimates $\hat{t}_{on}^{(1)}$ and $\hat{t}_{off}^{(1)}$ are obtained from their MLE:

$$\hat{t}_{on}^{(1)} = \langle \bar{t}_{on} \rangle - \Delta t, \quad \hat{t}_{off}^{(1)} = \langle \bar{t}_{off} \rangle - \Delta t \quad (S80)$$

The new estimates are employed to update eq. (S78). Using \bar{N} and the state path, the parameter estimate $\overline{SBR}^{(1)}$ is calculated, as well as the new emission probability distributions for the on and the off states.

The Viterbi algorithm is applied a second time (employing the parameters with superscript (1)), from which the final estimates $\hat{t}_{on}^{(2)}$, $\hat{t}_{off}^{(2)}$ and $\overline{SBR}^{(2)}$ are obtained. This average signal-to-background value is used for the trajectory reconstruction (numLMS correction) in section 6.4.5. This is an approximation, given that the emitter at different positions of the EBP is expected to have different SBR values. For a beam separation of $L = 130$ nm and typical tracking error distribution ($\approx 30 - 40$ nm), the variation in the SBR values is expected to be small. For the tracking application in *E. coli*, we obtain $\langle SBR \rangle \approx 2.7$ and $\sigma_{SBR} \approx 1$ (it should be stressed that the distribution of the SBR is asymmetric, thus it is not 2.7 ± 1).

The state path also allows the determination of the total number of localizations S in each trajectory by counting all multiplex cycles in which the emitter is in the on-state. Note that all localizations with less than 3 total counts were discarded in the further processing steps.

6.4.5 Trajectory reconstruction

As explained in section 3.2 and in section 6.4.3, the recorded tracks are obtained employing the live position estimator $\hat{r}_{mLMS}^{(k=1)}(\hat{p}, \hat{\beta})$, which is biased. In order to obtain an improved trajectory estimation, the trajectories are corrected using the numLMSE $r_{numLMS}^{(k=2)}(\hat{p})$, introduced in section 3.2.3.

The numLMSE requires knowing the $\bar{p}(\bar{r})$ parameters and the SBR. The spatial dependence of the \bar{p} parameter is measured following the procedure in section 6.1. The background λ_b is estimated from the isolated trajectory (second) segment that is driven by background counts only (see section 6.4.4). The \bar{p} parameter estimates \hat{p} are obtained from the measured count quartet using the MLE of the underlying multinomial distribution (see eq. (S45)).

6.4.6 Error definition

In the tracking application, the single emitter is free to move during the multiplexing cycle. This motion blurring leads to a reduction of the localization precision compared to a static emitter localization. The average tracking error ϵ employed in this paper is defined as the distance between the *average emitter position during the multiplexing cycle* and its respective localization. For a track with M multiplex cycle updates, the tracking error ϵ is defined as:

$$\epsilon = \sqrt{\frac{1}{dM} \sum_{i=1}^M \|\hat{r}_i - \langle \bar{r}_i \rangle\|^2}, \quad \text{with } \bar{r} \in \mathbb{R}^d \quad (\text{S81})$$

where $\|\cdot\|$ is the Euclidian norm, \hat{r} is the position estimate and $\langle \bar{r} \rangle$ the average emitter position during the multiplex cycle of duration T :

$$\langle \bar{r} \rangle = \frac{1}{T} \int_0^T \bar{r}(t) dt \quad (\text{S82})$$

6.4.7 Diffusion coefficient estimation

In order to attribute a diffusion constant D to the numLMSE corrected trajectory (see section 6.4.5), the adapted OLSF algorithm described in section 5.2 is employed. The MSD is calculated using eq. (S68), where the emitter on and off states are obtained from the HMM state path (see section 6.4.4). Note that all localizations with less than 3 total counts are discarded.

The apparent diffusion constant plot in fig. 5D was calculated by using a sliding time window of 35 ms with a spacing of 3.75 ms. For every time window, the error bars ($\pm\sigma$) were approximated as the CRB times 1.4. This is because that is the precision obtained from simulations (see section 5.2).

The apparent diffusion histogram in fig. 5F was calculated by estimating D values from a sliding window of 35 ms length every 17.5 ms on the respective tracks. D estimations utilizing less than 100 valid localizations were discarded.

6.4.8 Localization precision

Figure 5H shows the localization precision σ against the mean number of photons per localization N_l . The localization precision σ was estimated for each track using the adapted OLSF algorithm described in section 5.2 (see also 6.4.7). Furthermore, N_l was estimated using the total counts \bar{N} and the state path (see 6.4.4). Subsequently, the precisions were binned according to their N_l values, and the respective mean and standard deviations were calculated and plotted.

6.5 Experimental setup

The setup consists of a custom-built scanning microscope with fast beam scanning and modulation capabilities. A schematic illustration is shown in fig. S13.

6.5.1 Optical setup

Excitation of the sample can be performed by a Gaussian or doughnut-shaped beam or a wide field illumination. Two main lasers are available (Laser 1 and 2 in fig. S13), providing light with a wavelength of 642 nm or 560 nm. The beams are focused to a Gaussian or doughnut-shaped excitation spot with circular polarization. Amplitude modulation and switching between both beams is performed with electro-optical modulators (EOM).

Lateral scanning of the beam position in the sample is performed by two orthogonal electro-optical deflectors (EOD) and a piezoelectric tip/tilt mirror. Each EOD is driven by two high voltage amplifiers (Falco Systems WMA-300 and Trek PZD700A) in a differential arrangement. The first amplifier provides scanning with a bandwidth of 5 MHz, whereas the second provides a bandwidth of 125 kHz. In combination with the piezoelectric tip-tilt mirror, this provides a high-dynamic beam scanning system with a range of about $20 \times 20 \mu\text{m}^2$. The ranges of the EODs are about $\pm 150 \text{ nm}$ and $\pm 1 \mu\text{m}$ for the fast and slow amplifiers, respectively. The sample is mounted on a piezoelectric stage which allows fine positioning in three dimensions.

Multiplexing of the excitation beam pattern (EBP) is done with the EODs and the fast amplifiers. The feedback controller is implemented with a combination EODs driven by the slow controller and the piezoelectric tip-tilt mirror. All scanners are controlled by the field-programmable gate array (FPGA) board. The control signal is split into two (for the EODs using the 125 kHz amplifiers and for the tip-tilt mirror) with a highpass/lowpass filter bank, with a time constant of 10 ms. This assures that the piezoelectric tip-tilt mirror (response time $\sim 1 \text{ ms}$) can follow easily.

The laser beams are focused onto the sample by an oil immersion microscope objective. Fluorescence photons are separated from the excitation light by a dichroic mirror (DM1 in fig. S13), spectrally filtered (F1 or F2 in fig. S13) and detected by an EMCCD camera or an APD. The APDs are coupled to multimode fibers which act as detection pinhole for spatial filtering. The effective pinhole diameter for APD 1 and APD 2 are 420 nm and 2.5 μm , respectively.

APD 1 is used for imaging whereas APD 2 is used for tracking applications. Switching between the camera and the APDs is done with a mirror on a motorized flip mount.

In addition to the main lasers 1 and 2, there are three other lasers (laser 3 to 5 in fig. S13) for wide field or focused illumination of the sample. Switching between the two illumination modes is performed by a lens (FL in fig. S13) on a motorized flip mount. The lasers in the setup are spectrally filtered (if necessary) by dichroic clean-up filters. Polarizers and wave plates are used to ensure a proper polarization.

6.5.2 System lock

The system is equipped with a system lock that measures and corrects drifts of the sample in all three dimensions.

Axial measurement of the sample position is based on measuring the displacements of a total internal reflected (TIR) beam on the coverslip-media interface. For this purpose, an infrared laser beam (laser 6 in fig. S13) is focused off-center into the back focal plane of the objective lens. The TIR signal is detected by a CMOS camera (camera 2 in fig.

S13). The center of mass in the camera image is used as measure for the axial sample position.

Lateral measurement of the sample position is performed by darkfield imaging scattering nanorods onto another CMOS camera (camera 3 in fig. S13). A 2D Gaussian function is fitted to the image of a nanorod and the center position serves as measure for the lateral sample position.

The axial and lateral positions of the sample are kept constant by commanding the xyz piezo stage with a PI feedback loop written in LabView. Camera images are acquired at a framerate of about 90 fps with camera 2 and with about 160 fps with camera 3. The images are exponentially averaged and the position of the stage is updated every 100 ms.

6.5.3 Measurement control software

Measurements are performed with custom programs written in LabView (National Instruments, Austin, TX, USA). These programs control the devices directly, via the data acquisition (DAQ) boards or the FPGA board. The LabView programs perform the measurements as described in the previous sections. The main components of the programs are: control the laser beam position and modulation, multiplex the EBP, acquire the counts from the APD, perform a live position estimation using the mLMSE, reposition the beams and save the recorded data. In addition, the sample position is kept stable by a custom LabView program as described in section 6.5.2.

6.6 Sample preparation

6.6.1 Buffers

Reducing and Oxidizing System buffer (ROXS): ROXS buffer is prepared according to (18).

Imaging buffer (IB): Imaging buffer for Alexa Fluor 647 is prepared according to (29) with slightly different final concentrations of the ingredients: 0.4 mg/ml glucose oxidase, 64 μ g/ml catalase, 50 mM TRIS/HCl pH 8.0, 10 mM NaCl, 10 mM MEA (Cysteamine) and 10 % (w/v) glucose. Additionally, 10 mM MgCl₂ is added.

Folding buffer (FB): DNA origami folding buffer is prepared from 1xTAE by addition of 10 mM MgCl₂.

6.6.2 Cleaning of coverslips and objective slides

In order to reduce background from unwanted fluorescent particles on the coverslips, all coverslips are cleaned prior to sample preparation. Cleaning is performed with a 2 % dilution of Hellmanex® III (Hellma GmbH & Co. KG, Müllheim, Germany) in Milli-Q water. The coverslips are sonicated inside the cleaning solution two times for 15 min. After the first 15 min, they are rinsed with Milli-Q water and the cleaning solution is replaced. After the second 15 min sonication step, the coverslips are rinsed

with Milli-Q water and dried with compressed N₂. Objective slides are cleaned in the same manner but with only a single 15 min sonication step instead of two.

For measurements with drift compensation via the system lock, Au nanorods with their scattering peak at 980 nm (A12-25-980; Nanopartz Inc., Loveland, CO, USA) are immobilized on the cleaned coverslips. Prior to use, the nanorod stock solution is sonicated for 5 min, vortexed for 10 s and diluted in Milli-Q water (1: 3 to 1: 5). 10 µl of this solution are placed on a cleaned coverslip. After 1 min of incubation, the coverslip is rinsed with Milli-Q water and dried with compressed N₂.

6.6.3 Flow channel

Several samples are prepared with self-assembled flow channels. Flow channels are formed by gluing a coverslip with double sided tape (Scotch®, 3M France) to an objective slide.

6.6.4 Fluorescent microspheres

Samples for PSF measurements are prepared on cleaned coverslips with immobilized Au nanorods. The coverslips are coated with Poly-L-Lysine by placing 50 µl 0.01 % Poly-L-Lysine solution (Sigma Aldrich) on the coverslips. After 5 min, the coverslip is rinsed with Milli-Q water and dried with compressed N₂. 20 nm fluorescent microspheres (FluoSpheres®, 0.02 µm, dark red fluorescent; Thermo Fischer Scientific Inc., Waltham, MA, USA) are sonicated for 10 min and diluted in PBS 10⁶ times. 20 µl bead solution is placed on the Poly-L-Lysine coated coverslip and rinsed with Milli-Q water after 1 min. The coverslip is dried with compressed N₂. The coverslip is mounted on an objective slide forming a flow channel. The channel is filled with PBS and sealed with epoxy glue (Hysol®, Loctite).

6.6.5 Immobilized single ATTO 647N molecule

Samples with individual immobilized ATTO 647N molecules are prepared from labeled DNA oligonucleotides. The sample for static localization measurements (fig. 3) is formed by annealing two 31 base long oligonucleotides of single stranded DNA (ssDNA) labeled with a single ATTO 647N molecule. For immobilization via biotin-streptavidin interaction, one end of the DNA is labeled with a biotin molecule. The dye is bound to the base at position 4. Single stranded DNA oligonucleotides were purchased from IBA GmbH (Göttingen, Germany). The sequences (5' to 3') are: ATA A(ATTO647N)TT TCA TTG CCA TAT ACT ACA GGA ATA A and TTA TTC CTG TAT ATG GCA ATG AAA TTA T(Biotin). The parentheses mark the bases labeled with ATTO 647N or Biotin, respectively. The oligonucleotides are mixed at equal concentrations and diluted to 100 nM concentration in 10 mM TRIS (pH 8), 10 mM NaCl and 1 mM EDTA. Annealing is performed by heating to ~95°C and gradually cooling down to room temperature in 45 min.

Samples are prepared according to the following protocol. From a cleaned objective slide and a cleaned coverslips with immobilized Au nanorods a flow channel is formed. The channel is filled with 15 μ l biotinylated BSA (Albumin, biotin labeled bovine, A8549-10MG, Sigma Aldrich) dissolved in PBS, 2 mg/ml. After an incubation time of 15 min the channel is flushed with 400 μ l PBS. The PBS in the flow channel is replaced by 15 μ l Streptavidin (Streptavidin, recombinant, 11721666001, Sigma Aldrich) dissolved in PBS, 0.5 mg/ml. After an incubation time of 15 min the channel is flushed with 400 μ l PBS. The PBS in the channel is replaced by 15 μ l annealed DNA solution (50 pM in PBS). After an incubation time of 15 min the channel is flushed with 200 μ l PBS and afterwards with 200 μ l ROXS. The sample is immediately sealed with epoxy glue (Hysol®, Loctite).

6.6.6 Immobilized labeled DNA origami

Custom DNA origami internally labeled with fluorescent dye molecules were purchased from GATTAquant GmbH (Braunschweig, Germany). The DNA origami have a size of about 60 nm \times 80 nm and are labeled with up to nine Alexa Fluor 647 molecules with arrangements as indicated in fig. 4C,H. Each origami has six biotin attached which allow immobilization on a surface via a biotin-streptavidin interaction. Fluorescence on/off switching of Alexa Fluor 647 is facilitated by embedding the sample in an imaging buffer (IB) containing MEA and an oxygen scavenger system.

Samples are prepared according to the following protocol. From a cleaned objective slide and a cleaned coverslips with immobilized Au nanorods a flow channel is formed. The channel is filled with 15 μ l biotinylated BSA (Albumin, biotin labeled bovine, A8549-10MG, Sigma Aldrich) dissolved in PBS, 2 mg/ml. After an incubation time of 15 min the channel is flushed with 400 μ l PBS. The PBS in the flow channel is replaced by 15 μ l Streptavidin (Streptavidin, recombinant, 11721666001, Sigma Aldrich) dissolved in PBS, 0.5 mg/ml. After an incubation time of 15 min the channel is flushed with 400 μ l PBS. Prior to addition of the DNA origami, the channel is flushed with 200 μ l FB. The FB in the channel is replaced by 15 μ l DNA origami solution (5 pM in FB). After an incubation time of 15 min the channel is flushed with 200 μ l FB and afterwards with 200 μ l IB. The sample is immediately sealed with picodent twinsil® speed 22 (picodent® Dental-Produktions- und Vertriebs-GmbH, Wipperfürth, Germany).

6.6.7 E. coli sample preparation

E. coli BW25993 with mEos2 fused to the ribosomal small subunit protein S2 was used for tracking in living cells (24).

M9 minimal media supplemented with 0.4 % glucose and RPMI 1640 amino acids (R7131, Sigma-Aldrich) was used for all cultures and agarose pads. Cultures were inoculated from overnight cultures and incubated shaking at 37°C until an OD600 of 0.15

was reached. Cells were pelleted by centrifugation at 4000 rpm for 4 min, and then resuspended in the same volume of fresh media.

For MINFLUX tracking measurements, these cells are placed on 2.5 % agarose pads (SeaPlaque GTG Agarose, Lonza). Immobilization is attained by squeezing the agarose pad between a clean microscope slide and a clean cover slip (see section 6.6.2). The latter is attached using a frame seal (Frame-Seal Slide Chambers SLF0601, Bio-Rad Laboratories GmbH). Tracking was performed at room temperature.

Supplementary Figures

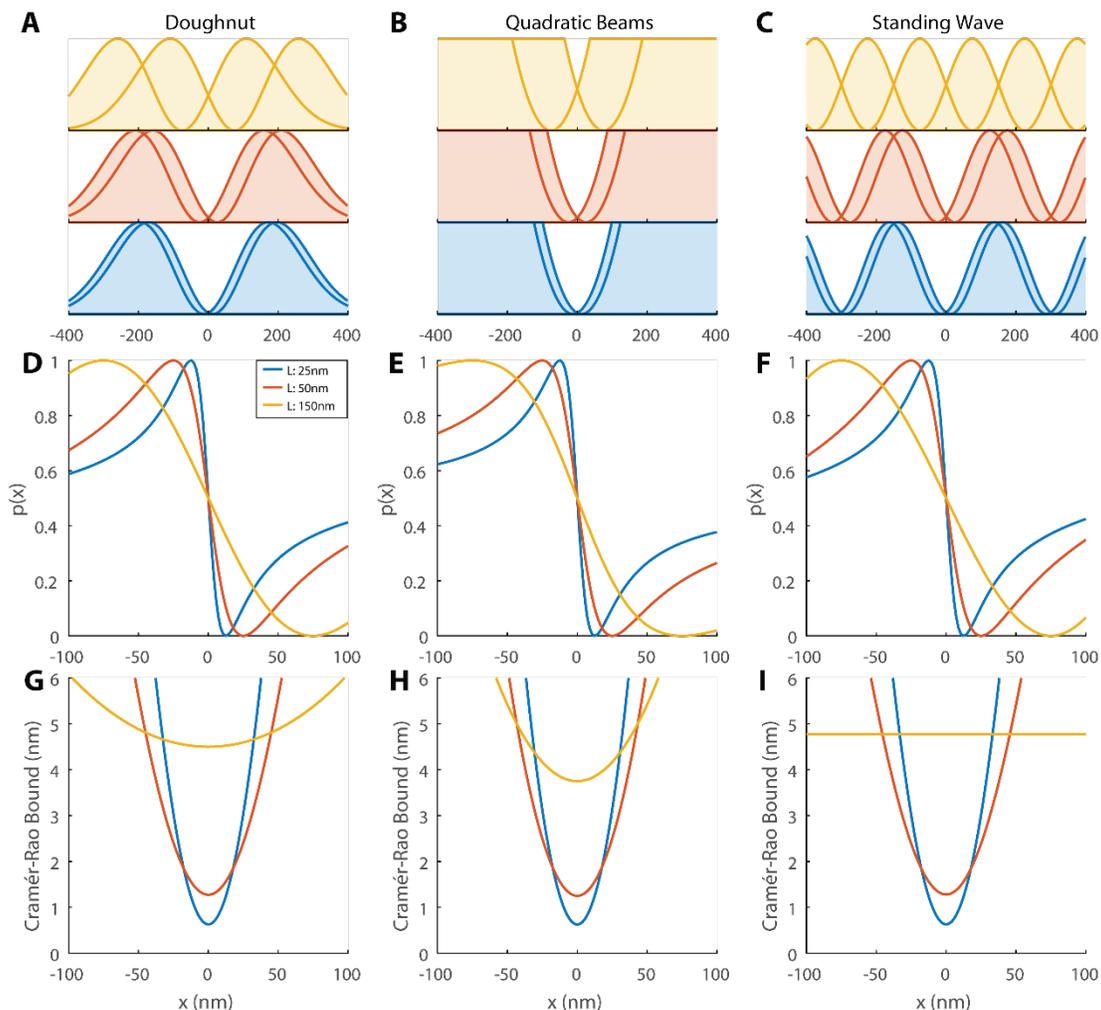


Fig. S1

One dimensional p functions and CRB. One-dimensional localization of an emitter with two (A, D, G) doughnut-, (B, E, H) quadratic beams and (C, E, I) two standing waves. For each case, the beam intensities $I_i(x)$, the binomial parameters $p(x)$ and the CRBs are shown for a beam separation L of 25 nm (blue), 50 nm (red) and 150 nm (yellow), and the use of $N = 100$ photons. The size parameter $fwhm$ is 300nm. The parameter k is set to $k = 2\pi/\lambda$, with $\lambda = 600$ nm.

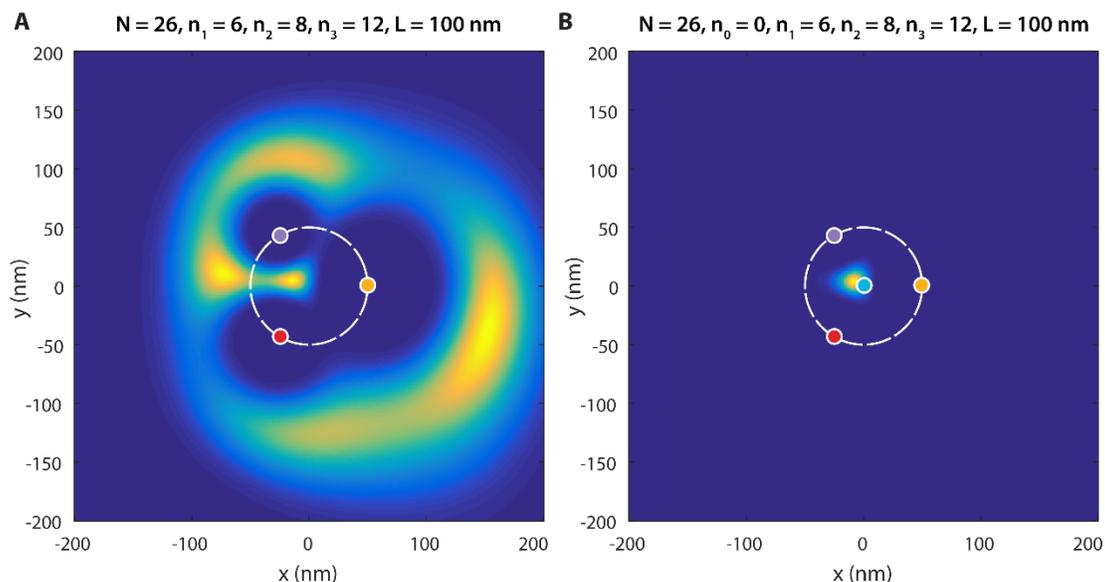


Fig. S2

Visualization of localization indeterminations. (A) Likelihood function $\mathcal{L}(\vec{r}_m | \vec{n})$ for the case of a localization with three exposures to doughnut beams (parameter $fwhm = 200$ nm) with their zeros at the marked colored points and collected photons of $\vec{n} = [6, 8, 12]^T$ with a total number $N = 26$. The likelihood function is badly behaved for creating a maximum likelihood estimator. (B) Same function for the case of a four doughnut localization and collected photons $\vec{n} = [0, 6, 8, 12]^T$. Though the same number of photons is collected, the radial information encoded in $n_0 = 0$ concentrates the relevant support of the likelihood function to a much smaller region with a clearly defined maximum; which has a better behavior in terms of defining a maximum likelihood estimator.

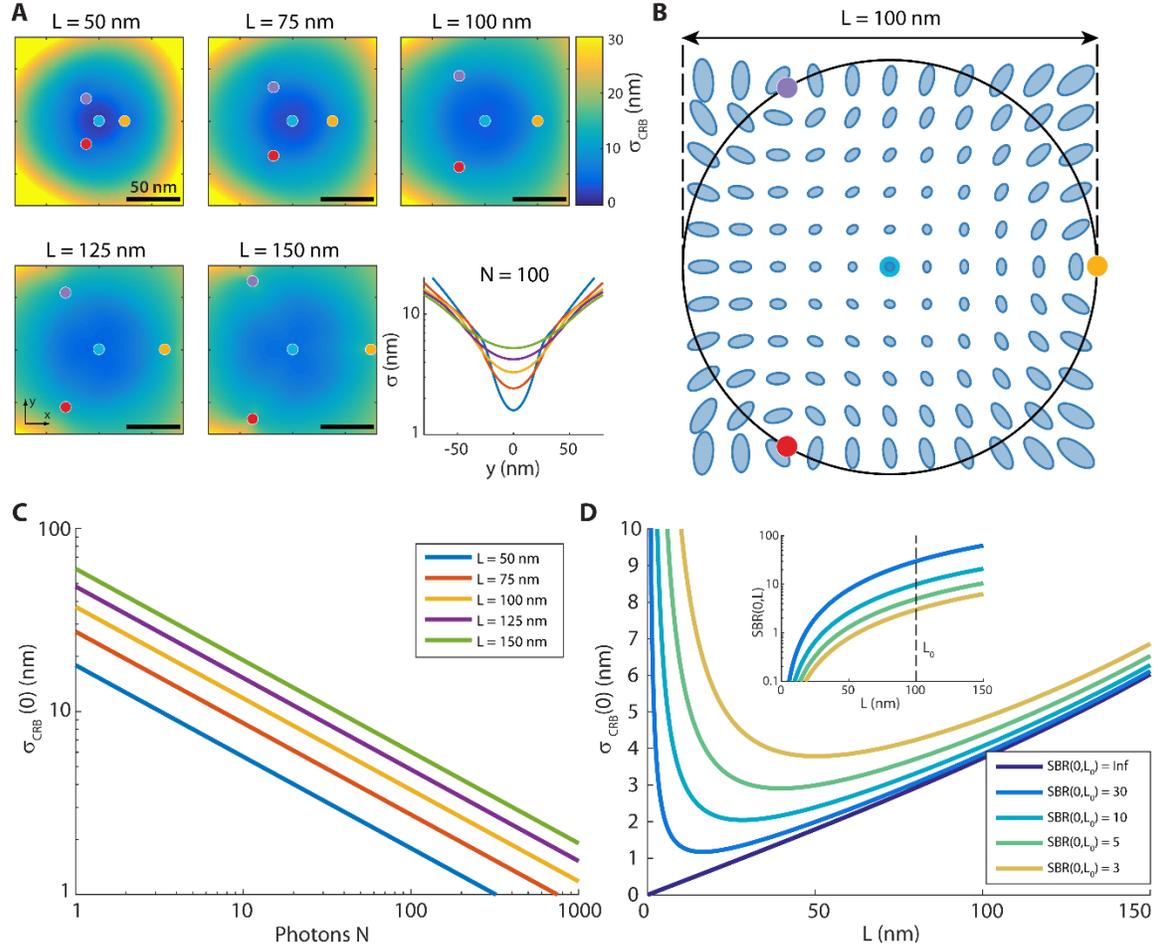


Fig. S3

CRB for the two-dimensional case. (A) Position dependent CRB $\tilde{\sigma}_{CRB}(\vec{r})$ for $N = 100$ photons. The employed EBP is depicted in section 2.2. The doughnut shaped excitation profiles are set to have $fwhm = 360$ nm. Colored dots indicate the respective excitation beam origins. The peripheral beams have triangular symmetry and are inscribed in a circle of diameter L . The CRB is minimal at the EBP origin and grows with increasing distance to it. Reduction of the parameter L enables to increase the localization precision, especially in the vicinity of the origin. The last panel shows a y cut through $\tilde{\sigma}_{CRB}(\vec{r})$, where the coloring indicates different L values (see legend in (C)). (B) The covariance matrix $\Sigma(\vec{r}_m)$ plotted at different positions for a total number of $N = 1000$ photons. The covariance is visualized as ellipses (contour level $e^{-1/2}$). Note, that the localization precision is not isotropic. (C) CRB at the origin of the beam pattern as a function of the total photon number N for different beam pattern diameters L and an infinite SBR. (D) Influence of the $SBR(\vec{0}, L)$ (see eq. (S31)) on the CRB at the origin $\vec{r} = \vec{0}$, for $N = 100$ photons and for different $SBR(\vec{0}, L_0 = 100 \text{ nm})$ values. Depending on the value of the latter, a reduction of L improves the localization precision only up to a minimal L . A

further diminution results in an increase of $\tilde{\sigma}_{CRB}$. The effective signal-to-background ratio $SBR(\bar{0}, L)$ is visualized in the inset.

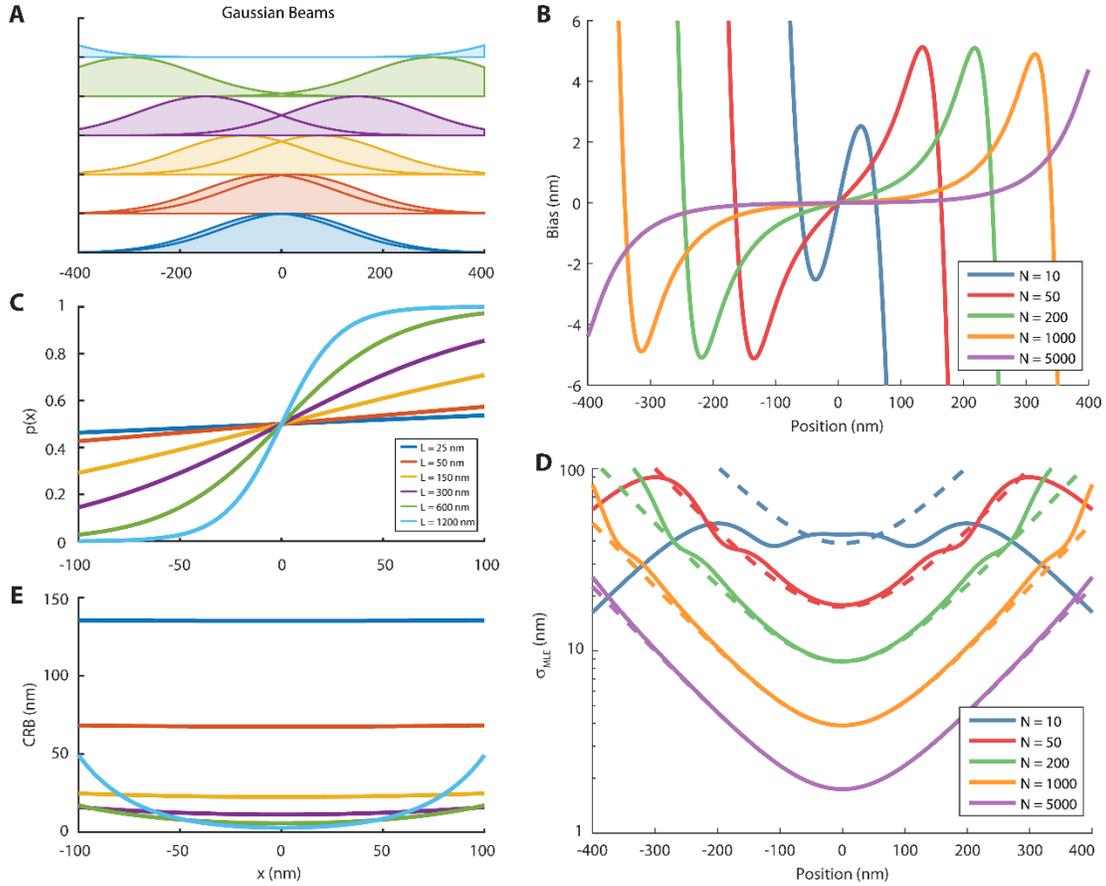


Fig. S4

One-dimensional localization performance with two Gaussian beams. (A, C, E) One-dimensional localization of an emitter with two Gaussian beams. The beam intensities $I_i(x)$, the binomial parameters $p(x)$ and the CRBs are shown for different beam separation L of 25 nm to 1200 nm and the use of $N = 100$ photons. The size parameter $fwhm$ is 300 nm. (B) Bias as a function of the emitter position for different total number of photons N . Excitation geometry: two 1D Gaussian beams with $fwhm = 320$ nm separated by a distance $L = 300$ nm. The region with acceptable bias grows with increasing number of photons, whereas the origin stays unbiased. (D) Comparison of the standard deviation $\tilde{\sigma}_{MLE}$ (solid lines) to the Cramér-Rao bound $\tilde{\sigma}_{CRB}$ (dashed lines). Higher values of N make $\tilde{\sigma}_{MLE}$ approach the information theoretical limit. Especially at outer positions, $\tilde{\sigma}_{MLE}$ falls below the CRB, as the estimator \hat{x}_m^{MLE} is biased.

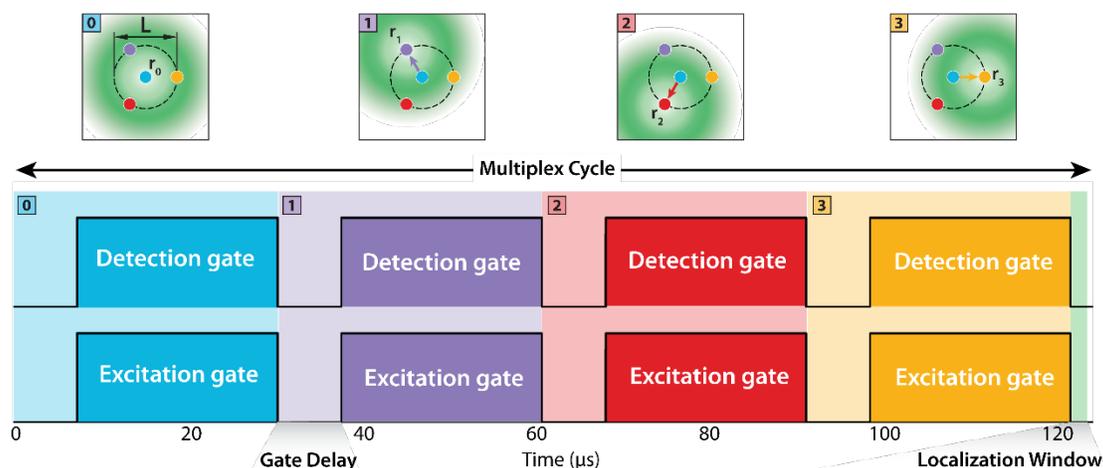


Fig. S5

Visualization of the Multiplex Cycle. An FPGA board controls the modulation and deflection of the laser beam, and the acquisition of the fluorescence photon-counts. The excitation is gated by EOMs in order to assure that the fluorophore is not excited during beam repositioning (see section 6.5). The off-time is given by the *gate delay*. Photons are acquired only if the detection gate is enabled. The delay of the latter is equal to the gate delay. After four successive exposures, a live position estimation can be calculated employing the FPGA board. The calculation time slot is given by the *localization window* (green). Note that the *multiplex cycle* is visualized depicting the excitation pattern described in section 2.2, but is not restricted to it.

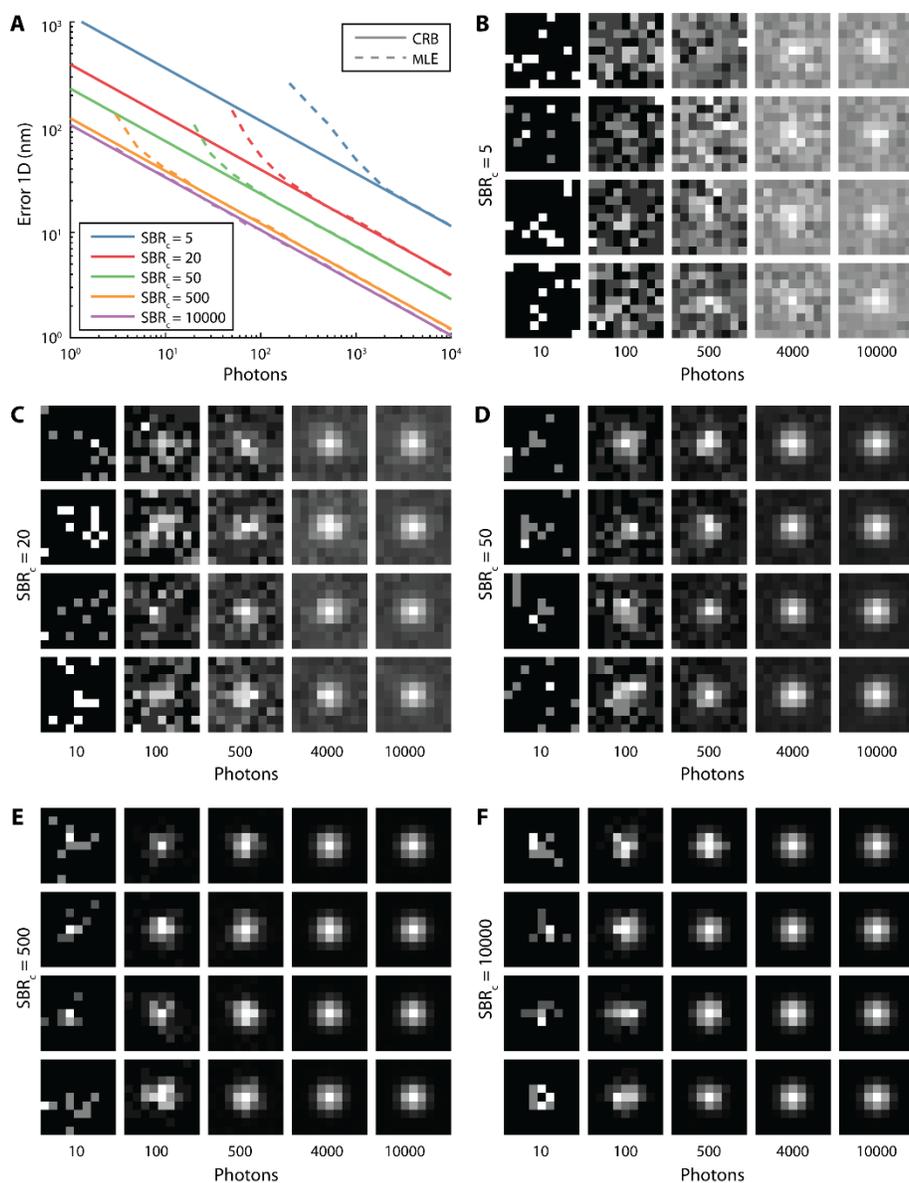


Fig. S6

CRB and MLE performance for an ideal camera. (A) CRB (solid line) and performance of the MLE (dashed line) for single emitter localization using a perfect camera. The SBR_c is define as in eq. (S60). The PSF is modeled as a symmetrical Gaussian function with width $\sigma_{PSF} = 100$ nm. The pixel size of the camera is 100 nm. Camera size 9×9 pixels. (B-F) Simulated camera images for different SBR_c and number of detected photons N . 9×9 pixels, pixel size 100 nm. The emitter is located in the center of the image. All images are normalized to the maximal pixel value.

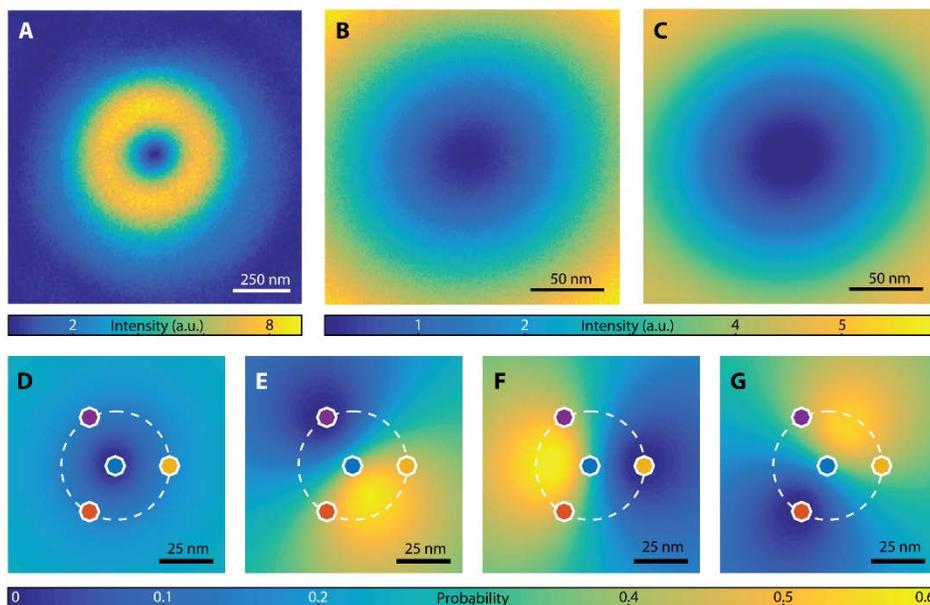


Fig. S7

Experimental PSF. (A) Typical measured PSF of the employed doughnut-shaped excitation beam. (B) Mean image of the drift corrected and centered frames of the central region of doughnut-shaped excitation beam. (C) Fit of the model function to (B). (D-G) $\{p_i^{(0)}\}$ functions of the four multiplexed doughnut-shaped excitation beams for a beam separation of $L = 50$ nm. The positions of the beam centers are indicated in the same way as in fig. 2.

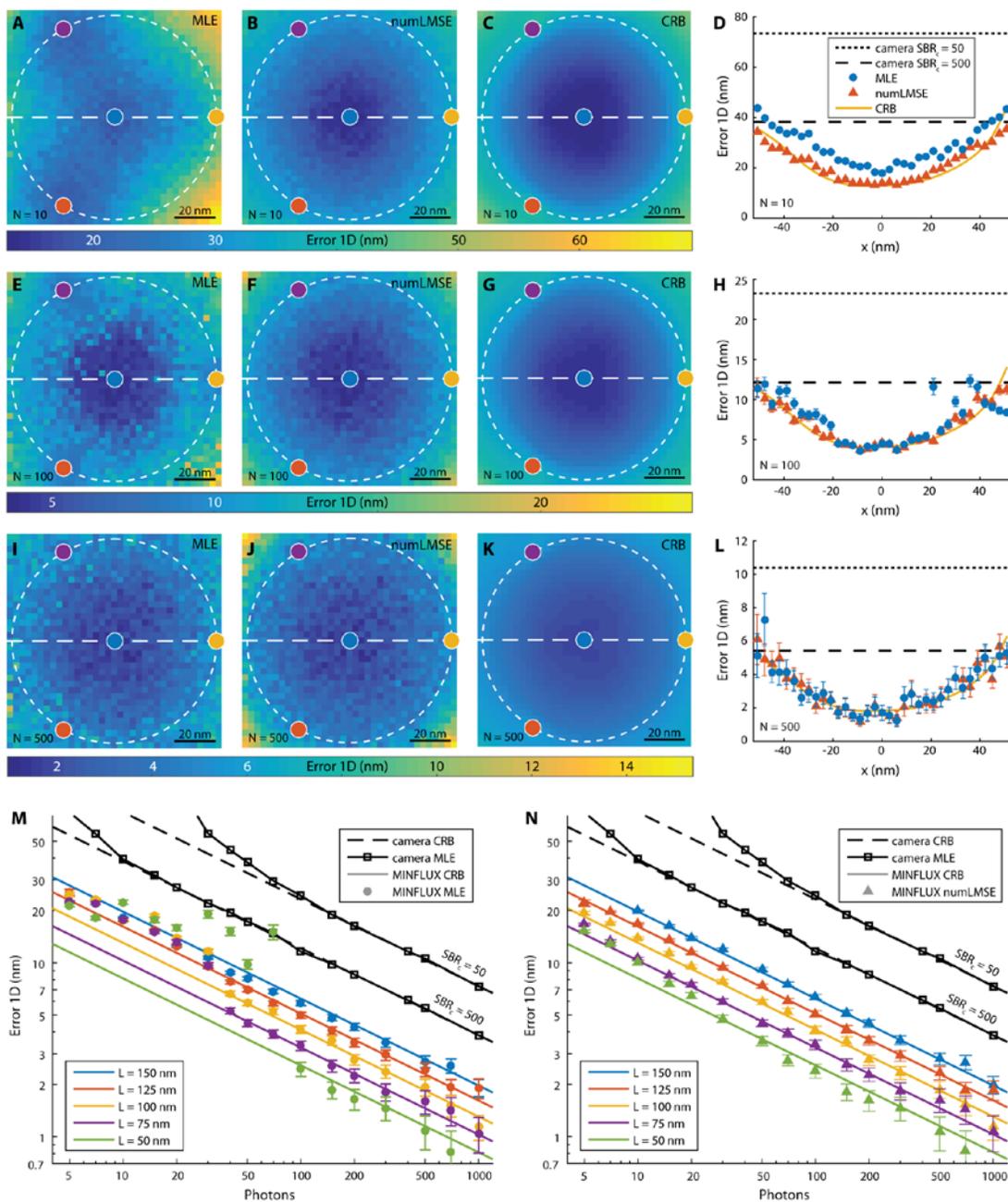


Fig. S8

Details on the localization precision measurements. (A-L) Static localization performance map for a single ATTO 647N molecule. Measurements were performed with $L = 100$ nm. (A), (E) and (I) show the measured performance of the MLE for three different numbers of detected photons. The corresponding performance of the numLMSE is shown in (B), (F) and (J). The CRB is shown in (C), (G) and (K). Profiles along the dashed line in the maps are shown in (D), (H) and (L). The camera localization performance was calculated using the same parameters as for fig. S6 (see also section 4). (M) and (N) show the localization performance of at the center of the beam pattern for

different beam separations L . Position estimation was done with the MLE (see section 3.1.2) and the numLMSE (see section 3.2.3), respectively.

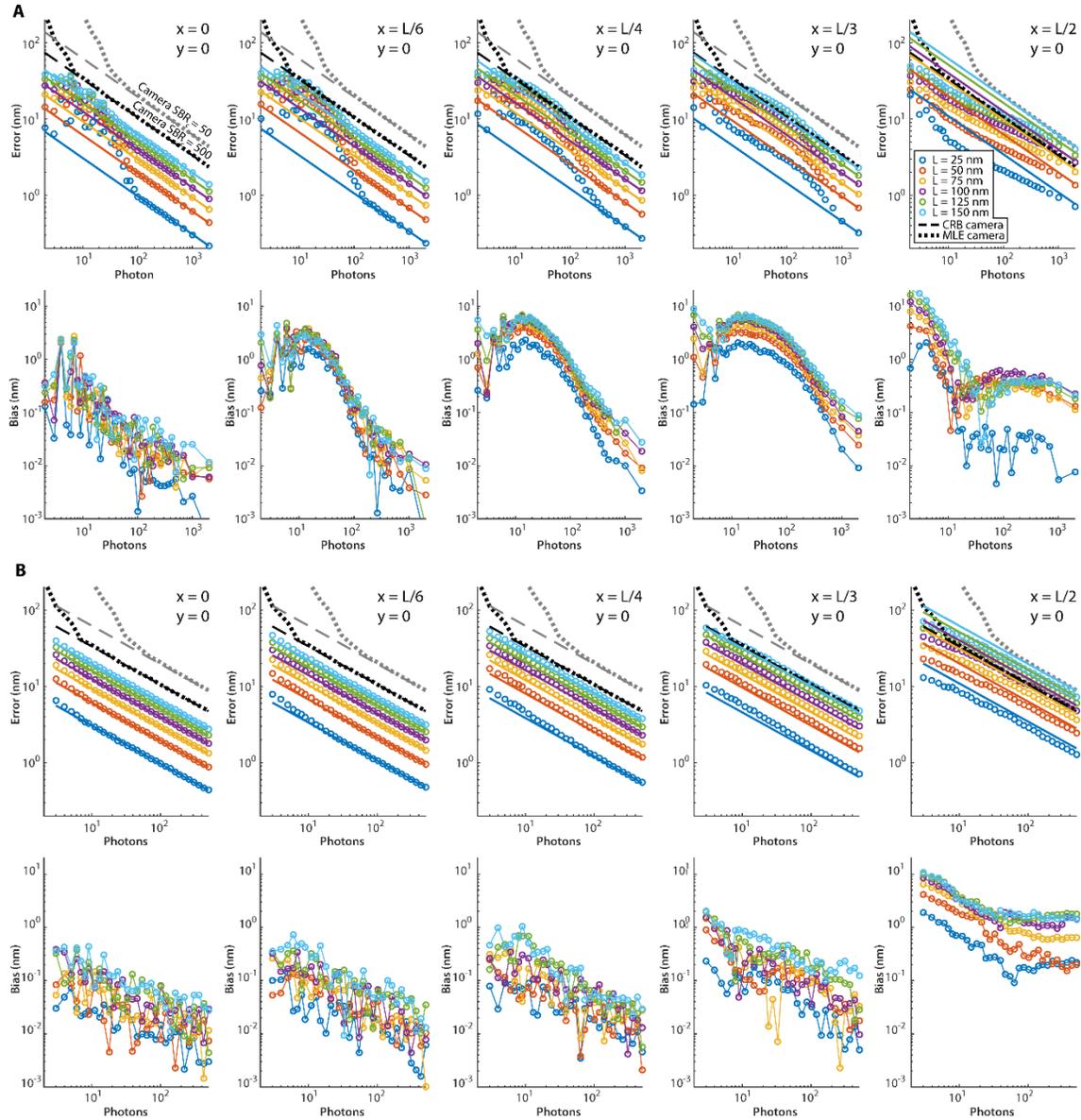


Fig. S9

Convergence of position estimators. Localization performance of the MLE and the numLMSE as a function of the number of photons N evaluated at different positions \vec{r} . The excitation PSF as well as the EBP are the ones introduced in section 2.2. The SBR is set to 10 for every evaluated emitter position in (A) and (B). Note that in an experiment where an emitter is localized a different distances from the EBP center, the SBR increases for emitter positions further away from the center (within the field of view of interest). For every pictured data point, 10^4 $\hat{p}(\vec{r})$ parameter estimates $\{\hat{p}\}_{\vec{r}}$ were generated. The camera CRB and MLE were calculated as depicted in section 4.1. Parameters were set to: $\sigma_{PSF} = 87$ nm, pixel size $a = 100$ nm, number of pixels $K = 9 \times 9$, and two values of SBR_c (50 in gray and 500 in black), respectively. Note

that no read noise is incorporated. **(A)** The MLE of section 3.1.2 was used to localize the respective \hat{p} values. It can be seen that the MLE converges to the CRB for $N \gtrsim 100 - 500$ photons only, depending on \bar{r} as well as on L (units of L in legend: nm). For smaller N the MLE deviates considerably from its information theoretical limit. **(B)** In this case the numLMSE (see section 3.2.3) was used to localize the \hat{p} values of the respective sets $\{\hat{p}\}_{\bar{r}}$. The divergence to the CRB is strongly reduced, compared to the MLE. Especially for low photon numbers N the numLMSE performs better. Note that this estimator is designed to work in a region $|\bar{r}| \lesssim L/2$.

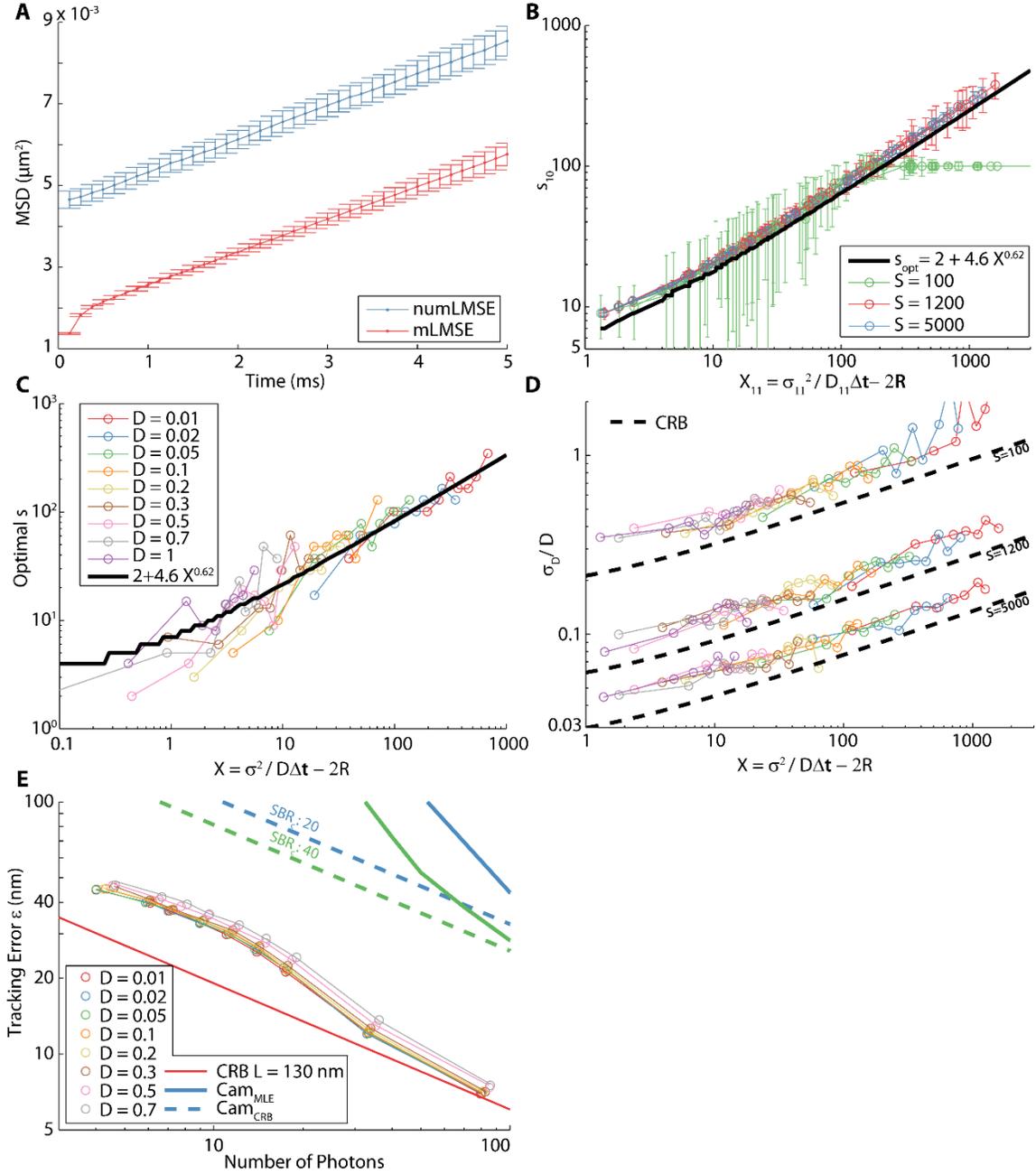


Fig. S10

Estimation of the apparent diffusion coefficient and tracking error. Data in this figure is based on a tracking simulation using MINFLUX (see section 5.2). The single emitter movement follows free isotropic Brownian motion in two dimensions. **(A)** Comparison of the mean of 100 MSD curves for trajectories obtained using the mLMS estimator (red) and the numLMS estimator (blue). The total width of the error bars represents twice the standard deviation. **(B)** Convergence of s to s_{opt} using the recursive algorithm in eq. (S67) for different number of trace length S . Each data point represents

the median of 100 trajectories. The total width of the error bars is twice the standard deviation. The pictured s values s_{10} were obtained after 10 iterations, and are plotted against the estimated reduced square localization error \hat{X}_{11} . The latter was obtained from the results of the 10th iteration. **(C)** Optimization of the number of MSD points s used in the fit to eq. (S63). Optimization was conducted by minimizing the deviation of the estimated \hat{D} value to its ground truth D value (legend in $\mu\text{m}^2/\text{s}$). The optimal s values can be described by the relationship $s_{opt}(X) = 2 + 4.6X^{0.62}$. **(D)** Comparison of the relative standard deviation σ_D/D of the estimated D values to the CRB. Each data point is calculated from 100 trajectories. The results of 3 different trajectory length S are visualized. 10 iterations of eq. (S67) were conducted. **(E)** Comparison of the real tracking error ϵ (thin colored lines), to the value extracted by the OLSF method (colored circles) for different diffusivities (legend in $\mu\text{m}^2/\text{s}$). Additionally, the CRB (eq. (S26)) for $L = 130$ nm and $fwhm = 450$ nm is shown as thick red line. The information theoretical limit for camera localization for two SBR_c are shown as thick dotted lines. The MLE camera performance (for the same two SBR_c) is pictured as thick colored lines.

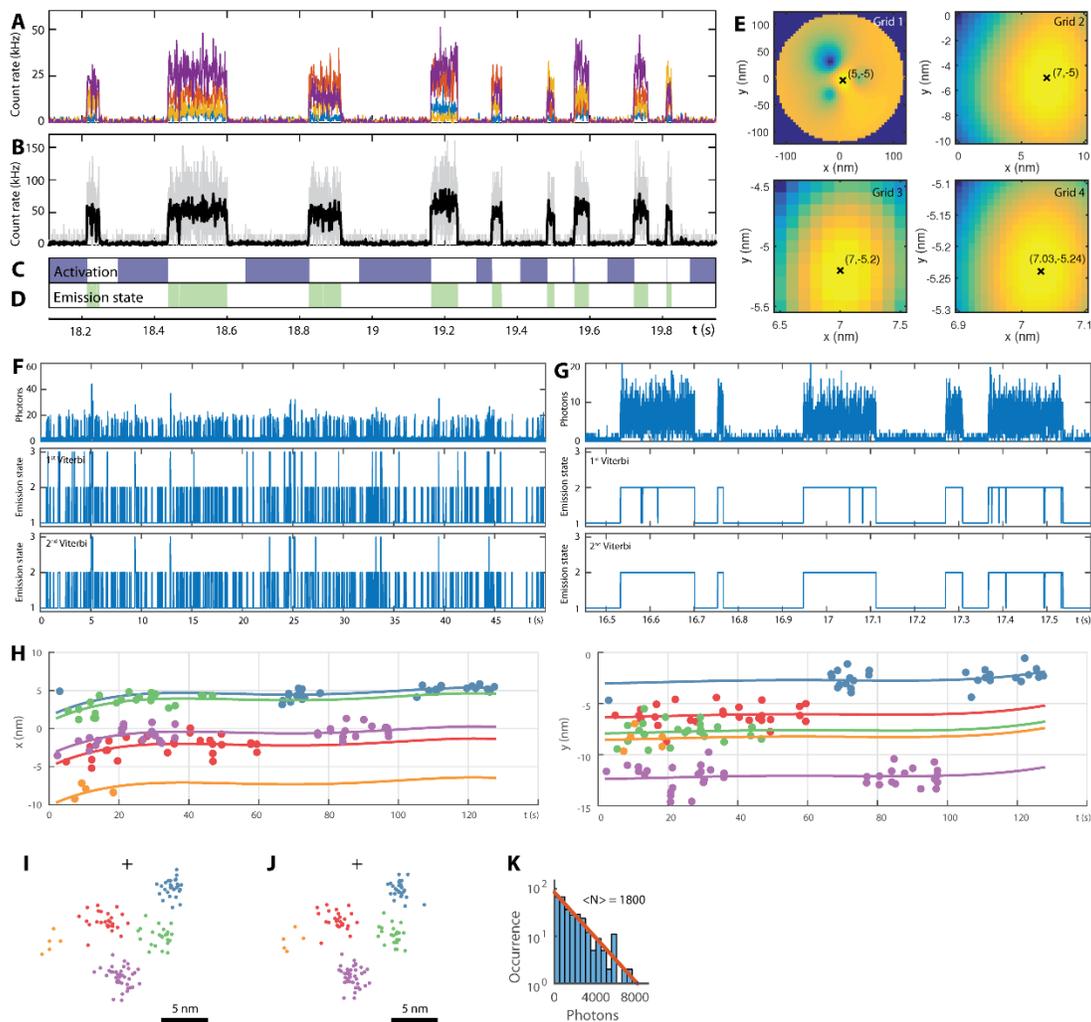


Fig. S11

Details of MINFLUX nanoscopy processing. (A) Typical trace of recorded counts during imaging of a labeled DNA origami. Time bins: 1 ms. (B) Total count rate of recorded photons (gray: raw; black: binned to 1 ms) of the same data as in (A). (C) Conditional activation of molecules with illumination of 405 nm light. The laser is switched on and off according to the count rate of detected photons as described in section 6.3.1. (D) Single molecule emission states detected by an HMM trace segmentation as shown in (F) and (G). (E) Position estimation by maximization of the likelihood function in a successive grid search algorithm. It shows a typical localization in the imaging of the origami. The quartet of detected photons is $\bar{n} = (124, 609, 695, 1382)$ and the *SBR* is 17.75. (F) HMM trace segmentation on the total count trace of detected photons as described in section 6.3.2. (G) Zoom into (F). (H) Post processing drift correction of the 6 nm origami imaging as described in section 6.3.7. The solid lines show the estimated function for the residual sample drift. The colors

are consistent with (I). **(I)** Localizations of 6 nm origami imaging clustered into nano-domains before drift correction. The black cross indicates the position of the central excitation beam. **(J)** Data of (I) after drift correction. **(K)** Histogram of detected photons per emission event in the measurement of the larger origami.

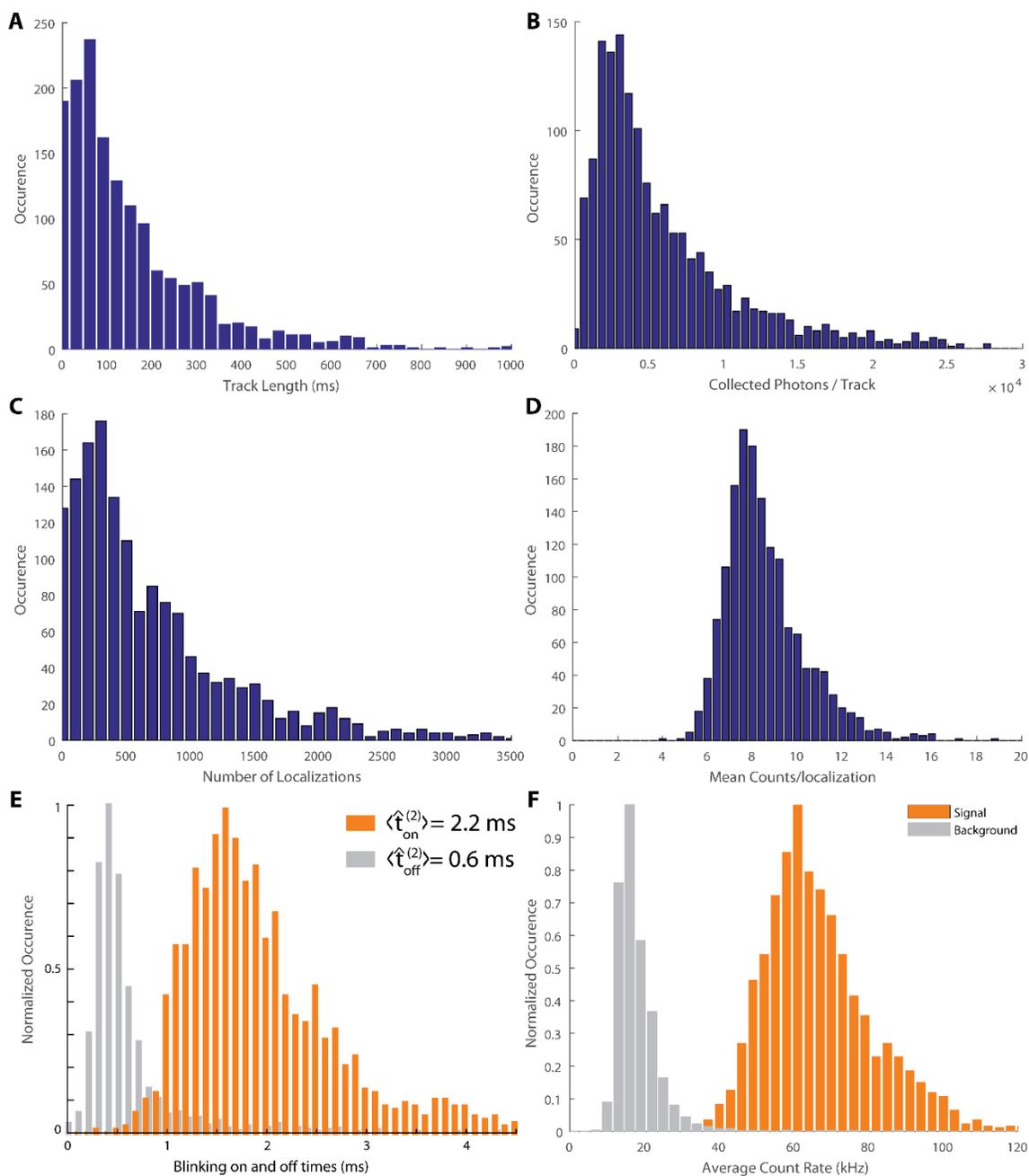


Fig. S12

MINFLUX tracking in living *E. coli*. (A) Histogram of the track length with mean of 157 ms. (B) Collected photons (background corrected) per trace with mean of 5803. (C) Valid localizations per trace with mean of 742 after application of a HMM and discarding of all localizations with less than 3 total counts (see 6.4.3). Note that the distributions in (A-C) extend further from the plot limits. All stated mean values of the distributions in (A-C) are the parameter value (i.e. average) of the respective non-truncated exponential distributions (truncated distributions have higher averages than

reported). The longest measured track was 1444 ms with 86009 photons and 7503 valid localizations. **(D)** Mean counts per localization of the respective tracks, with an ensemble average of 9. **(E)** Blinking on and off times of mEos2 extracted from a two stage HMM on the counts \bar{N} of each track (see 6.4.3). **(F)** Normalized occurrences of the signal and background count rates for the respective traces extracted from \bar{N} and the state path (see 6.4.3). The average *SBR* is 2.7.

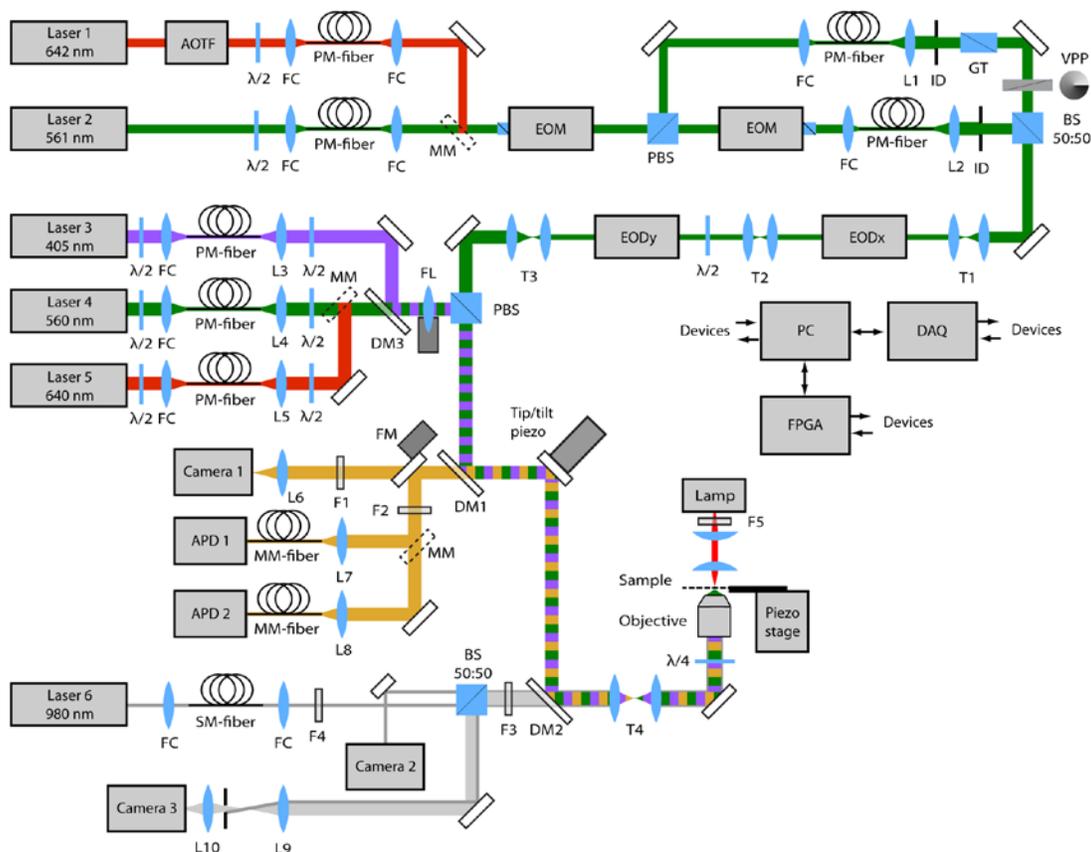


Fig. S13

Schematic illustration of the experimental setup. The following components are used: *Laser 1*: VFL-P-1500-642 (MPB Communications Inc., Pointe-Claire, Quebec, Canada) or Koheras SuperK Extreme (NKT Photonics, Birkerød, Denmark), *Laser 2*: Cobolt Jive™ (Cobolt AB, Solna, Sweden). *Laser 3*: 405-50-COL-004 (Oxxius, Lannion, France), *Laser 4*: VFL-P-1000-560 (MPB Communications Inc., Pointe-Claire, Quebec, Canada), *Laser 5*: LDH-D-C-640 (PicoQuant, Berlin, Germany), *Laser 6*: LP980-SF15 (Thorlabs Inc., Newton, NJ, USA), *AOTF*: AOTFnC VIS (AA Sa, Orsay, France), *EOM*: LM 0202 P 5W + LIV 20 (Qioptiq Photonics GmbH & Co. KG, Göttingen, Germany), *EODx and EODy*: M-311-A (Conoptics Inc., Danbury, CT, USA) + PZD700A (Trek Inc., Lockport, NY, USA) + WMA-300 (Falco Systems BV, Amsterdam, The Netherlands), *Tip/tilt piezo*: PSH-10/2 + EVD300 (both piezosystem jena GmbH, Jena, Germany), *Piezo stage*: P-733.3-DD + E725 (both Physik Instrumente (PI) GmbH & Co. KG, Karlsruhe, Germany), *GT*: Glan-Thompson prism (B. Halle Nachfl. GmbH, Berlin, Germany) *PBS*: polarizing beam splitter cube (B. Halle Nachfl. GmbH, Berlin, Germany), *BS*: beam splitter cube 50:50, *FC*: fiber collimator 60FC-* (Schäfter+Kirchhoff, Hamburg, Germany), $\lambda/2$: half wave plate (B. Halle Nachfl. GmbH, Berlin, Germany), $\lambda/4$: quarter wave plate (B. Halle Nachfl. GmbH, Berlin, Germany), *VPP*: VPP 1a (RPC Photonics, Rochester, NY, USA), *SM-fiber*: single mode fiber (Thorlabs Inc., Newton, NJ, USA), *PM-fiber*: polarization maintaining fiber (Thorlabs

Inc., Newton, NJ, USA or Schäfter+Kirchhoff, Hamburg, Germany), *MM-fiber*: multimode fiber M31L01 (Thorlabs Inc., Newton, NJ, USA), *L*: achromatic lens with VIS or NIR AR coating (Thorlabs Inc., Newton, NJ, USA or Qioptiq Photonics GmbH & Co. KG, Göttingen, Germany), *T*: telescope, *FL*: lens of flip mount, *ID*: iris diaphragm, *MM*: mirror on magnetic mount, *FM*: mirror on motorized flip mount, *DM1*: Z488/633RDC (Chroma Technology Corp., Bellows Falls, VT, USA) or ZT405/488/561 (Chroma Technology Corp., Bellows Falls, VT, USA), *DM2*: BB1-E02P (Thorlabs Inc., Newton, NJ, USA), *DM3*: Z500RDC-XT (Chroma Technology Corp., Bellows Falls, VT, USA), *F1*: FF01-635/LP-25 (Semrock Inc., Rochester, NY, USA) or BLP02-561R-25 (Semrock Inc., Rochester, NY, USA) + FF01-842/SP-25 (Semrock Inc., Rochester, NY, USA), *F2*: ET700/75m (Chroma Technology Corp., Bellows Falls, VT, USA) + ZET642NF (Chroma Technology Corp., Bellows Falls, VT, USA) or BLP02-561R-25 (Semrock Inc., Rochester, NY, USA) + FF01-775/SP-25 (Semrock Inc., Rochester, NY, USA), *F3*: FEL850 (Thorlabs Inc., Newton, NJ, USA), *F4*: LL01-980-12.5 (Semrock Inc., Rochester, NY, USA) + FB980-10 (Thorlabs Inc., Newton, NJ, USA), *F5*: Z635/10 (Chroma Technology Corp., Bellows Falls, VT, USA), *APD 1,2*: SPCM-AQRH-13-FC (Excelitas Technologies, Waltham, MA, USA), *Camera 1*: Luca S, (Andor Technology Ltd., Belfast, UK), *Camera 2,3*: DMK 22BUC02 (The Imaging Source Europe GmbH, Bremen, Germany), *Objective*: HCX PL APO 100x/1.40-0.70 Oil CS (Leica Microsystems GmbH, Wetzlar, Germany), *Lamp*: LQ 1100 (Fiberoptic – Helm AG, Bühler, Switzerland), *PC*: personal computer running Windows 7 (Microsoft Corp., Redmond, WA, USA) and LabView 2013 (National Instruments, Austin, TX, USA), *DAQ*: NI PCIe-6353 + NI PCI-6259 (both National Instruments, Austin, TX, USA) + USB-3103 (Measurement Computing Corporation, Norton, MA, USA), *FPGA*: NI PCIe-7852R (National Instruments, Austin, TX, USA)

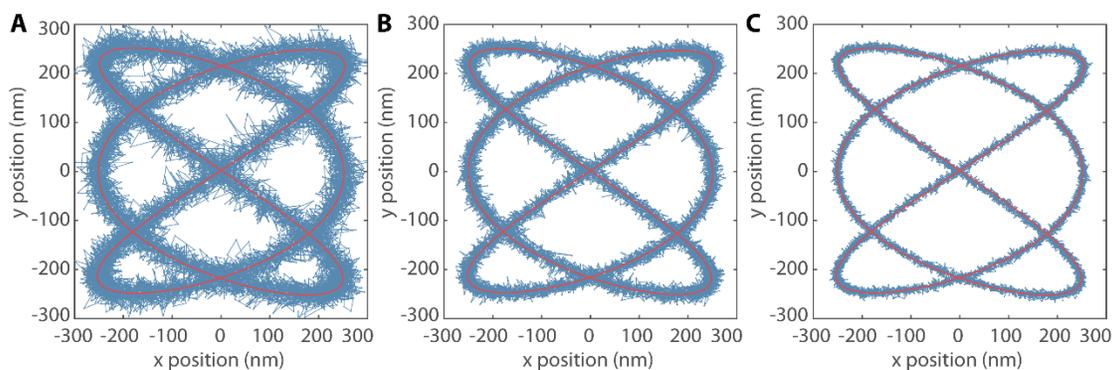


Fig. S14

Tracking of predefined trajectory. (A) A 20 nm fluorescent bead was moved in the xy -plane along a predefined trajectory using the piezoelectric stage. Two sinusoidal trajectories along the x and y direction were injected to the piezoelectric stage (solid red line). The peak-to-peak amplitudes were 500 nm and the frequencies were 1.5 Hz and 1 Hz in x and y , respectively. MINFLUX tracking was performed with a beam separation of $L = 100$ nm and a multiplex cycle rate of 8 kHz. Live position estimation is calculated using the mLMS estimator (see supporting note 3.2.2) with $\bar{\beta}$ parameters set to $\beta_0 = 1.27$ and $\beta_1 = 3.8$. The retrieved trajectory is shown by the blue line. The average photon count rate was 92 kHz, corresponding to a mean number N of photons per localization of 11.5. (B) Same as in (A) but tracking was performed with an average count rate of 263 kHz, corresponding to a mean number N of photons per localization of 32. (C) Same as in (A) but tracking was performed with a beam separation of $L = 50$ nm. The average photon count rate was 327 kHz, corresponding to a mean number N of photons per localization of 40.

Supplementary Tables

d	K	p_i	$F_{\bar{p}}/N$	\mathcal{J}^*	$\tilde{\sigma}_{CRB}$
1	2	$\frac{I_i}{I_0 + I_1}$	$p_0^{-1} + p_1^{-1}$	$\frac{\partial p_0}{\partial r_0}$	$\sqrt{\text{tr}(F_{\bar{r}_m}^{-1})}$
1	3	$\frac{I_i}{I_0 + I_1 + I_2}$	$\begin{bmatrix} p_0^{-1} & 0 \\ 0 & p_1^{-1} \end{bmatrix} + p_2^{-1}\mathbf{1}$	$\begin{bmatrix} \frac{\partial p_0}{\partial r_0} \\ \frac{\partial p_1}{\partial r_0} \end{bmatrix}$	$\sqrt{\text{tr}(F_{\bar{r}_m}^{-1})}$
2	3	$\frac{I_i}{I_0 + I_1 + I_2}$	$\begin{bmatrix} p_0^{-1} & 0 \\ 0 & p_1^{-1} \end{bmatrix} + p_2^{-1}\mathbf{1}$	$\begin{bmatrix} \frac{\partial p_0}{\partial r_0} & \frac{\partial p_0}{\partial r_1} \\ \frac{\partial p_1}{\partial r_0} & \frac{\partial p_1}{\partial r_1} \end{bmatrix}$	$\sqrt{\frac{1}{2} \text{tr}(F_{\bar{r}_m}^{-1})}$
2	4	$\frac{I_i}{I_0 + I_1 + I_2 + I_3}$	$\begin{bmatrix} p_0^{-1} & 0 & 0 \\ 0 & p_1^{-1} & 0 \\ 0 & 0 & p_2^{-1} \end{bmatrix} + p_3^{-1}\mathbf{1}$	$\begin{bmatrix} \frac{\partial p_0}{\partial r_0} & \frac{\partial p_0}{\partial r_1} \\ \frac{\partial p_1}{\partial r_0} & \frac{\partial p_1}{\partial r_1} \\ \frac{\partial p_2}{\partial r_0} & \frac{\partial p_2}{\partial r_1} \end{bmatrix}$	$\sqrt{\frac{1}{2} \text{tr}(F_{\bar{r}_m}^{-1})}$
3	4	$\frac{I_i}{I_0 + I_1 + I_2 + I_3}$	$\begin{bmatrix} p_0^{-1} & 0 & 0 \\ 0 & p_1^{-1} & 0 \\ 0 & 0 & p_2^{-1} \end{bmatrix} + p_3^{-1}\mathbf{1}$	$\begin{bmatrix} \frac{\partial p_0}{\partial r_0} & \frac{\partial p_0}{\partial r_1} & \frac{\partial p_0}{\partial r_2} \\ \frac{\partial p_1}{\partial r_0} & \frac{\partial p_1}{\partial r_1} & \frac{\partial p_1}{\partial r_2} \\ \frac{\partial p_2}{\partial r_0} & \frac{\partial p_2}{\partial r_1} & \frac{\partial p_2}{\partial r_2} \end{bmatrix}$	$\sqrt{\frac{1}{3} \text{tr}(F_{\bar{r}_m}^{-1})}$
3	5	$\frac{I_i}{I_0 + I_1 + I_2 + I_3 + I_4}$	$\begin{bmatrix} p_0^{-1} & 0 & 0 & 0 \\ 0 & p_1^{-1} & 0 & 0 \\ 0 & 0 & p_2^{-1} & 0 \\ 0 & 0 & 0 & p_3^{-1} \end{bmatrix} + p_4^{-1}\mathbf{1}$	$\begin{bmatrix} \frac{\partial p_0}{\partial r_0} & \frac{\partial p_0}{\partial r_1} & \frac{\partial p_0}{\partial r_2} \\ \frac{\partial p_1}{\partial r_0} & \frac{\partial p_1}{\partial r_1} & \frac{\partial p_1}{\partial r_2} \\ \frac{\partial p_2}{\partial r_0} & \frac{\partial p_2}{\partial r_1} & \frac{\partial p_2}{\partial r_2} \\ \frac{\partial p_3}{\partial r_0} & \frac{\partial p_3}{\partial r_1} & \frac{\partial p_3}{\partial r_2} \end{bmatrix}$	$\sqrt{\frac{1}{3} \text{tr}(F_{\bar{r}_m}^{-1})}$

Table S1.

Relevant quantities for characterizing the localization scheme performance. d is the dimensionality of the localization, K is the number of exposures of the emitter, p_i are the components of the multinomial vector parameter \bar{p} , $F_{\bar{p}}$ is the Fisher information matrix on \bar{p} , \mathcal{J}^* is the reduced Jacobian matrix for the change of variables from the reduced \bar{p} space to the molecules position \bar{r}_m space and $\tilde{\sigma}$ is the arithmetic mean of the eigenvalues of the CRB for the covariance matrix of the molecules position estimation. $\mathbf{1}$ is an all-ones matrix. The shaded cases are the most relevant for this work.

Ref.	Probe(s)	D ($\mu\text{m}^2/\text{s}$)		σ (nm)	Δt (ms)	Length (ms)	Average Localizations	Cutoff
a (46)	RNAP-PAmCherry	7-8		40	15	85	5.6	4
b (24)	L1-mEos2	0.055	bound	20	20	180	9	5
	S2-mEos2	0.4	free	60				
c (47)	RelA-YFP	1.52		45	10	90-150	>9-15	6-10
	RelA-mEos2	0.64						
	RelA-Dendra2	0.32						
	S2-mEos2	0.05						
d (48)	S2-YFP	0.04	10-30		30	150-180	5-6	8-13
e (49)	VSVG-EosFP	0.14		25	50	250	4-5	15
	Gag-EosFP	0.11						

Table S2.

Results and parameter values of typical camera tracking experiments in living cells using fluorescent protein labels. The cutoff states the minimum number of localizations used for D estimation.

References and Notes

1. S. W. Hell, J. Wichmann, Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994). doi:10.1364/OL.19.000780 Medline
2. T. A. Klar, S. Jakobs, M. Dyba, A. Egner, S. W. Hell, Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8206–8210 (2000). doi:10.1073/pnas.97.15.8206 Medline
3. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, H. F. Hess, Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006). doi:10.1126/science.1127344 Medline
4. M. J. Rust, M. Bates, X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–796 (2006). doi:10.1038/nmeth929 Medline
5. S. T. Hess, T. P. K. Girirajan, M. D. Mason, Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006). doi:10.1529/biophysj.106.091116 Medline
6. S. W. Hell, Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007). doi:10.1126/science.1137395 Medline
7. H. Deschout, F. Cella Zanacchi, M. Mlodzianoski, A. Diaspro, J. Bewersdorf, S. T. Hess, K. Braeckmans, Precisely and accurately localizing single emitters in fluorescence microscopy. *Nat. Methods* **11**, 253–266 (2014). doi:10.1038/nmeth.2843 Medline
8. W. Heisenberg, *The Physical Principles of the Quantum Theory* (Chicago Univ. Press, Chicago, 1930).
9. N. Bobroff, Position measurement with a resolution and noise-limited instrument. *Rev. Sci. Instrum.* **57**, 1152–1157 (1986). doi:10.1063/1.1138619
10. R. E. Thompson, D. R. Larson, W. W. Webb, Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775–2783 (2002). doi:10.1016/S0006-3495(02)75618-X Medline
11. K. I. Mortensen, L. S. Churchman, J. A. Spudich, H. Flyvbjerg, Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nat. Methods* **7**, 377–381 (2010). doi:10.1038/nmeth.1447 Medline
12. J. Engelhardt, J. Keller, P. Hoyer, M. Reuss, T. Staudt, S. W. Hell, Molecular orientation affects localization accuracy in superresolution far-field fluorescence microscopy. *Nano Lett.* **11**, 209–213 (2011). doi:10.1021/nl103472b Medline
13. M. D. Lew, M. P. Backlund, W. E. Moerner, Rotational mobility of single molecules affects localization accuracy in super-resolution fluorescence microscopy. *Nano Lett.* **13**, 3967–3972 (2013). doi:10.1021/nl304359p Medline

14. A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, P. R. Selvin, Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization. *Science* **300**, 2061–2065 (2003). doi:10.1126/science.1084398 Medline
15. J. Elf, G.-W. Li, X. S. Xie, Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007). doi:10.1126/science.1141967 Medline
16. A. Kusumi, T. A. Tsunoyama, K. M. Hirose, R. S. Kasai, T. K. Fujiwara, Tracking single molecules at work in living cells. *Nat. Chem. Biol.* **10**, 524–532 (2014). doi:10.1038/nchembio.1558 Medline
17. S. J. Sahl, M. Leutenegger, M. Hilbert, S. W. Hell, C. Eggeling, Fast molecular tracking maps nanoscale dynamics of plasma membrane lipids. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6829–6834 (2010). doi:10.1073/pnas.0912894107 Medline
18. J. Vogelsang, R. Kasper, C. Steinhauer, B. Person, M. Heilemann, M. Sauer, P. Tinnefeld, A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem. Int. Ed. Engl.* **47**, 5465–5469 (2008). doi:10.1002/anie.200801518 Medline
19. Q. Zheng, M. F. Juetter, S. Jockusch, M. R. Wasserman, Z. Zhou, R. B. Altman, S. C. Blanchard, Ultra-stable organic fluorophores for single-molecule research. *Chem. Soc. Rev.* **43**, 1044–1056 (2014). doi:10.1039/C3CS60237K Medline
20. S. Weisenburger, B. Jing, D. Hänni, L. Reymond, B. Schuler, A. Renn, V. Sandoghdar, Cryogenic colocalization microscopy for nanometer-distance measurements. *ChemPhysChem* **15**, 763–770 (2014). doi:10.1002/cphc.201301080 Medline
21. A. Sharonov, R. M. Hochstrasser, Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18911–18916 (2006). doi:10.1073/pnas.0609643104 Medline
22. M. Dai, R. Jungmann, P. Yin, Optical imaging of individual biomolecules in densely packed clusters. *Nat. Nanotechnol.* **11**, 798–807 (2016). doi:10.1038/nnano.2016.95 Medline
23. J. V. Pellegrotti, G. P. Acuna, A. Puchkova, P. Holzmeister, A. Gietl, B. Lalkens, F. D. Stefani, P. Tinnefeld, Controlled reduction of photobleaching in DNA origami-gold nanoparticle hybrids. *Nano Lett.* **14**, 2831–2836 (2014). doi:10.1021/nl500841n Medline
24. A. Sanamrad, F. Persson, E. G. Lundius, D. Fange, A. H. Gynnå, J. Elf, Single-particle tracking reveals that free ribosomal subunits are not excluded from the *Escherichia coli* nucleoid. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11413–11418 (2014). doi:10.1073/pnas.1411558111 Medline
25. S. W. Hell, Nanoscopy with focused light (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **54**, 8054–8066 (2015). doi:10.1002/anie.201504181 Medline
26. S. W. Hell, Method of and apparatus for tracking a particle, particularly a single molecule, in a sample. Patent application WO 2013/072273 A1 B2, 23 May 2013.
27. S. W. Hell, High-resolution fluorescence microscopy with a structured excitation beam. Patent application WO 2015/097000 A1, 2 July 2015.

28. J. J. Schmied, A. Gietl, P. Holzmeister, C. Forthmann, C. Steinhauer, T. Dammeyer, P. Tinnefeld, Fluorescence and super-resolution standards based on DNA origami. *Nat. Methods* **9**, 1133–1134 (2012). doi:10.1038/nmeth.2254 Medline
29. G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates, X. Zhuang, Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8**, 1027–1036 (2011). doi:10.1038/nmeth.1768 Medline
30. S. A. McKinney, C. S. Murphy, K. L. Hazelwood, M. W. Davidson, L. L. Looger, A bright and photostable photoconvertible fluorescent protein. *Nat. Methods* **6**, 131–133 (2009). doi:10.1038/nmeth.1296 Medline
31. C. L. Vestergaard, Optimizing experimental parameters for tracking of diffusing particles. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **94**, 022401 (2016). doi:10.1103/PhysRevE.94.022401 Medline
32. X. Michalet, A. J. Berglund, Optimal diffusion coefficient estimation in single-particle tracking. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **85**, 061916 (2012). doi:10.1103/PhysRevE.85.061916 Medline
33. B. Schuler, E. A. Lipman, W. A. Eaton, Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743–747 (2002). doi:10.1038/nature01060 Medline
34. A. Pertsinidis, Y. Zhang, S. Chu, Subnanometre single-molecule localization, registration and distance measurements. *Nature* **466**, 647–651 (2010). doi:10.1038/nature09163 Medline
35. A. H. Voie, D. H. Burns, F. A. Spelman, Orthogonal-plane fluorescence optical sectioning: Three-dimensional imaging of macroscopic biological specimens. *J. Microsc.* **170**, 229–236 (1993). doi:10.1111/j.1365-2818.1993.tb03346.x Medline
36. A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, S. Chu, Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.* **11**, 288–290 (1986). doi:10.1364/OL.11.000288 Medline
37. A. E. Cohen, W. E. Moerner, Method for trapping and manipulating nanoscale objects in solution. *Appl. Phys. B* **86**, 093109 (2005).doi:10.1063/1.1872220
38. S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, New Jersey, 1993), vol. 1.
39. F. Pukelsheim, *Optimal Design of Experiments* (SIAM, 1993), vol. 50.
40. J. E. Mosimann, On the compound multinomial distribution, the multivariate beta-distribution, and correlations among proportions. *Biometrika* **49**, 65–82 (1962).
41. F. Huang, T. M. P. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, M. W. Davidson, D. Toomre, J. Bewersdorf, Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat. Methods* **10**, 653–658 (2013). doi:10.1038/nmeth.2488 Medline
42. A. J. Berglund, Statistics of camera-based single-particle tracking. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **82**, 011917 (2010). doi:10.1103/PhysRevE.82.011917 Medline

43. B. Shuang, C. P. Byers, L. Kisley, L.-Y. Wang, J. Zhao, H. Morimura, S. Link, C. F. Landes, Improved analysis for determining diffusion coefficients from short, single-molecule trajectories with photoblinking. *Langmuir* **29**, 228–234 (2013). doi:10.1021/la304063j Medline
44. C. L. Vestergaard, P. C. Blainey, H. Flyvbjerg, Optimal estimation of diffusion coefficients from single-particle trajectories. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **89**, 022726 (2014). doi:10.1103/PhysRevE.89.022726 Medline
45. X. Michalet, Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **82**, 041914 (2010). Medline
46. M. Stracy, C. Lesterlin, F. Garza de Leon, S. Uphoff, P. Zawadzki, A. N. Kapanidis, Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4390–E4399 (2015). doi:10.1073/pnas.1507592112 Medline
47. W. Li, E. Bouveret, Y. Zhang, K. Liu, J. D. Wang, J. C. Weisshaar, Effects of amino acid starvation on RelA diffusive behavior in live *Escherichia coli*. *Mol. Microbiol.* **99**, 571–585 (2016). doi:10.1111/mmi.13252 Medline
48. S. Bakshi, A. Siryaporn, M. Goulian, J. C. Weisshaar, Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* **85**, 21–38 (2012). doi:10.1111/j.1365-2958.2012.08081.x Medline
49. S. Manley, J. M. Gillette, G. H. Patterson, H. Shroff, H. F. Hess, E. Betzig, J. Lippincott-Schwartz, High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Methods* **5**, 155–157 (2008). doi:10.1038/nmeth.1176 Medline