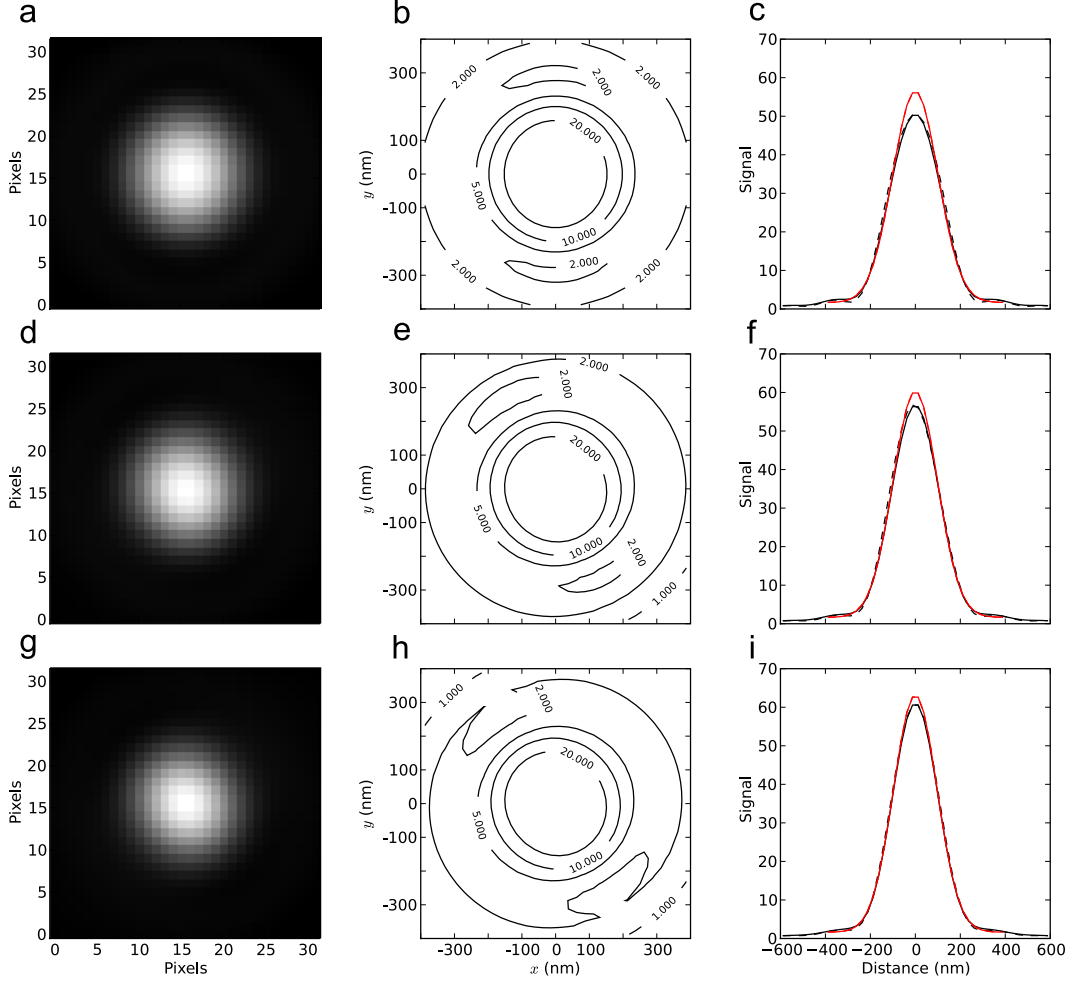


Supplementary Figure 1: PSF of TIRF-excited fluorescent bead

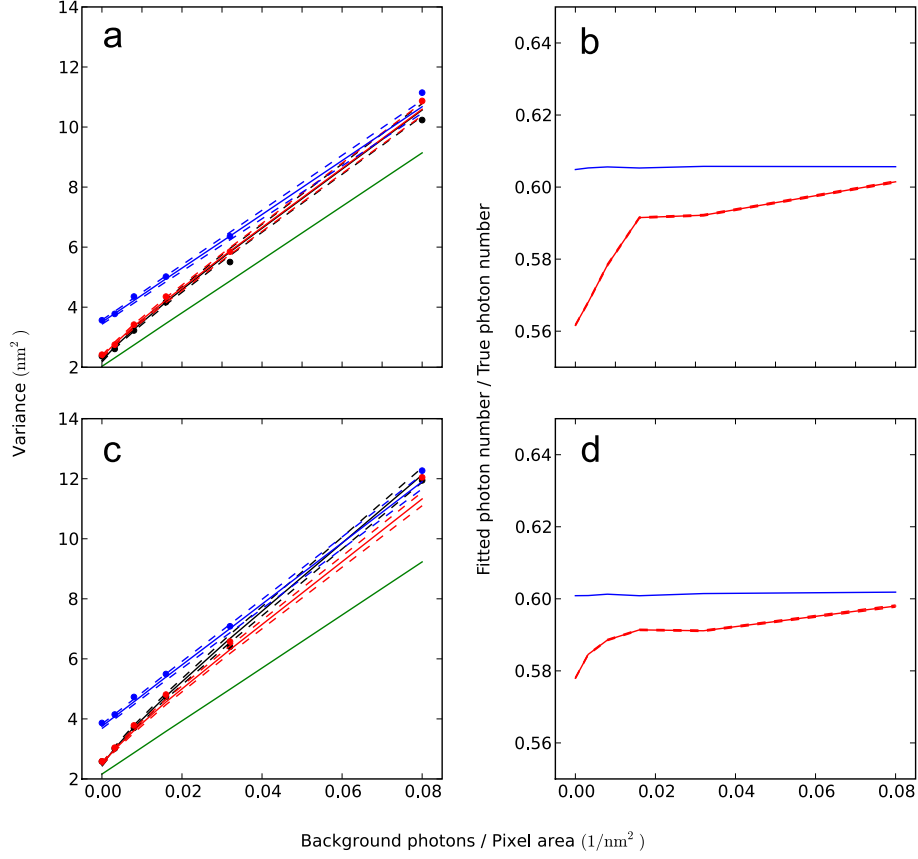


Supplementary Figure 1. PSF of fluorescent bead in TIRF at different angles of incident light. We used the following physical parameters: peak emission wavelength $\lambda = 605$ nm, refractive index of the optics $n = 1.78$, refractive index of the medium $n_0 = 1.33$, and numerical aperture $NA = 1.65$. Furthermore, we assumed right-handed circular polarized light. **(a):** The PSF of a fluorescent bead using light incident at the critical angle. We normalized the theoretical PSF to 10^4 photons and integrated it over each pixel in a 32×32 pixel array with 25 nm pixel width. **(b):** The theoretical PSF used in **(a)** is shown as lines of constant contour for the values 20, 10, 5, 2, and 1. The plot reveals that the PSF is not circular symmetric w.r.t. rotation in the plane. The PSF is symmetric, however, with respect to change of sign in both

Supplementary Figure 1: PSF of TIRF-excited fluorescent bead

coordinates, and its maximum here coincides with its center. (c): Cross sections through the PSF used in (a) and (b) for $y = 0$ (full black line) and $x = 0$ (dashed black line). Close inspection reveals that the PSF is wider along the y -axis than along the x -axis. Note that the PSF vaguely resembles the Airy distribution [Ref. 1 in the paper], but is not. The PSF lacks the diffraction rings of the Airy distribution. Instead, it has a shoulder of almost constant intensity surrounding the central peak. The Airy distribution is the diffraction-limited image of a plane wave [Ref. 1 in the paper], not of a dipole and not of any superposition of dipoles. The Airy distribution is symmetric with respect to rotation in the plane, and while this is generally not the case for PSFs of dipoles, the latter do decrease as $1/r^3$ like the Airy distribution, hence have infinite variance, because $\langle \vec{r}^2 \rangle = \int_0^\infty r^2 p(r) 2\pi r dr \sim \int^\infty = \infty$. This is the reason why the simple average of photon arrival coordinates, the centroid of the photon distribution (**Supplementary Note**), is a very bad estimator for the location of the photon source: this estimator inherits infinite variance from the experimental PSF. A rotationally symmetric 2D Gaussian plus a constant background was fitted (MLEwG) to the distribution in (a). This Gaussian's cross-sections for $y = 0$ (full red line) and for $x = 0$ (dashed red line) are shown. These two curves are indistinguishable, as expected, due to the rotational symmetry of the Gaussian. (d–f): Same as respective (a–c) for light incident at an angle $\theta_i = 1.0 > \bar{\theta}_i$. Notice the larger degree of asymmetry apparent in (e). By close inspection of (f), it is seen that the maximum of the distribution along the y -axis is shifted slightly towards the negative values, why the maximum and the center of the distribution do not coincide. The Gaussian fit is unbiased on the x -coordinate of the center, while a small bias (~ 1 nm) is found on the y -coordinate. This scale of (f) does not resolve this bias. (g–i): Same as respective panels above, but for light incident at $\theta_i \simeq \pi/2$. Notice that the 2D Gaussian is an increasingly better approximation with increasing angle of incidence.

Supplementary Figure 2: Simulation results for estimation methods



Supplementary Figure 2. Simulation results for estimation methods. **(a):** Variances of the different estimation methods as a function of the number of background photons per area. Data was simulated for an image area of 32×32 square pixels, each of 25 nm width, with the PSF centered in the area. The expected number of photons per pixel varies from 0 to 50 for this pixel size, while the expected number of photons in each image is $N = 10,000$. The variances of 3,000 localization estimates done with MLEwT, MLEwG, and GME at each of six different background values are plotted as, respectively, black, red, and blue dots. The theoretical predictions for the same variances are shown as full lines in corresponding colors, with dashed lines representing $\pm 1 \text{ s.d.}$ of the sample variance distributions. For the Gaussian-based estimators, the mean values of the fitted parameters σ^2 , N_G , and b_G^2 were used to calculate their variances. The variance given in Eq. (17) of Ref. 15 is shown as the full green line. **(b):** The fitted number of photons relative to its true number for MLG and GME. From the simulation for $(25 \text{ nm})^2$ pixels from above, the mean of

Supplementary Figure 2: Simulation results for estimation methods

the fitted number of photons obtained for MLG and GME are shown as full red, respectively blue lines. Dashed lines represent SEM. Both Gaussian-based estimators are seen to estimate the number of photons to $\sim 60\%$ of its true value. **(c)**: Same as **(a)**, except here 8×8 pixels each of $(100 \text{ nm})^2$ size was used to bin the exact same photon detection events in the exact same image area. **(d)**: Same as **(b)**, except here $(100 \text{ nm})^2$ pixels were used to image the simulated photons as in **(c)**.

SUPPLEMENTARY NOTE: THEORETICAL TREATMENT

Contents

1	Effect of finite pixel width a	2
1.1	Case of general point-spread function	2
1.2	Case of 2D Gaussian point-spread function	2
1.3	From sums to integrals	3
2	Estimation with photon statistics	3
2.1	Poisson statistics	4
2.2	Diffraction limited image of a point-source	4
2.3	Unweighted least-squares fit: Gaussian Mask Estimator (GME) . . .	5
2.3.1	Unweighted Fitting	5
2.3.2	Variance of Gaussian Mask Estimator	6
2.4	Weighted least-squares fit (WLS)	7
2.5	Weighted least-squares fit with offset	8
2.6	Maximum Likelihood Estimation (MLE)	11
2.6.1	MLE	11
2.6.2	Why MLE is better than WLS	11
2.6.3	Variance of MLEwT	12
2.6.4	Variance of MLEwG	12
2.6.5	The function $\mathcal{I}_1(\tau)$	13
2.6.6	The centroid estimator	13
2.7	A too-good-to-be-true interpolation formula for precision	13
3	The theoretical point spread function	14
3.1	Fixed-dipole case	14
3.2	In-plane dipoles and 2D Gaussians	16
3.3	Isotropic case	17
3.4	Polarized illumination in TIRF	18
3.5	PSF for isotropic distribution of dipoles in TIRF	20
3.6	Comparison with 2D Gaussian	23
3.7	Analytical approximation	23
3.8	Covariance matrix for parameters of PSF	26
4	What we <i>really</i> see with an EMCCD	26
4.1	Expected input, stochastic output	26
4.2	Mistakes to avoid	27
4.3	Expected localization errors in the real world of electron statistics .	27
5	Effects on precision from background and finite pixel width	31

1 Effect of finite pixel width a

1.1 Case of general point-spread function

With $p(\vec{r})$ the point-spread function (PSF) and p_i its integral over pixel $\#i$,

$$p_i = \int_{\text{pixel}\#i} p(x, y) dx dy , \quad (1)$$

one can, for moderate pixel width a compared to the characteristic radius of the diffraction limited spot, approximate p_i well with a systematic expansion of it in (pixel area)/(spot area), keeping only terms to two leading orders. This is done by Taylor series expansion of $p(\vec{r})$ about the midpoint \vec{r}_i of pixel $\#i$, and after a small calculation one has that

$$p_i = a^2 \left(p(\vec{r}_i) + \frac{a^2}{24} \Delta_{\vec{r}} p(\vec{r}_i) \right) , \quad (2)$$

where $\Delta_{\vec{r}}$ is the Laplacian differential operator, and terms of $\mathcal{O}(a^6)$ have been left out.

1.2 Case of 2D Gaussian point-spread function

With $p(\vec{r})$ a 2D normalized Gaussian with center at $\vec{r} = \vec{\mu}$ and variance σ^2 ,

$$p(\vec{r}|\vec{\mu}, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp \left(-\frac{(\vec{r} - \vec{\mu})^2}{2\sigma^2} \right) , \quad (3)$$

$$p_i = a^2 p(\vec{r}_i) \left(1 + \frac{a^2}{12} \left(\frac{(\vec{r}_i - \vec{\mu})^2}{2\sigma^2} - 1 \right) \right) . \quad (4)$$

This result can, to the same order in a^2/σ^2 , also be written

$$p_i = p_i(\vec{\mu}, \sigma^2) = a^2 p(\vec{r}_i|\vec{\mu}, \sigma_a^2) \quad (5)$$

with

$$\sigma_a^2 = \sigma^2 + \frac{1}{12} a^2 , \quad (6)$$

i.e., to order a^2/σ^2 the effect of finite pixel size on a Gaussian PSF is to broaden it without changing its general shape.

The last expression for p_i is a slightly better approximation, and guaranteed to be positive, as it must be. One can derive it in a more direct manner by treating the integral over pixel $\#i$ as an average, $\int_{\text{pixel}\#i} \dots dxdy = a^2 \langle \dots \rangle_i$, and *cumulant expand* $p_i = a^2 \langle p \rangle_i = a^2 \langle \exp(\ln p) \rangle_i$ to two leading orders in $\ln p$.

1.3 From sums to integrals

We need the results of this subsection below.

Application of Eq. (2) to

$$\int_{\text{pixel } \#i} \left(1 - \frac{a^2}{24} \Delta_{\vec{r}}\right) (\vec{r} - \vec{\mu})^2 p^2(\vec{r}|\vec{\mu}, \sigma_a^2) dx dy$$

gives that it equals

$$a^2(\vec{r}_i - \vec{\mu})^2 p^2(\vec{r}_i|\vec{\mu}, \sigma_a^2)$$

to two leading orders in a^2/σ_a^2 , and similarly

$$\int_{\text{pixel } \#i} \left(1 - \frac{a^2}{24} \Delta_{\vec{r}}\right) (\vec{r} - \vec{\mu})^2 p^3(\vec{r}|\vec{\mu}, \sigma_a^2) dx dy = a^2(\vec{r}_i - \vec{\mu})^2 p^3(\vec{r}_i|\vec{\mu}, \sigma_a^2)$$

After summing these two identities over pixels, the term with the Laplacian contributes nothing to the two integrals it occurs in,

$$\sum_i \int_{\text{pixel } \#i} \Delta_{\vec{r}}(\dots) dx dy = \int_{\text{all pixels}} \Delta_{\vec{r}}(\dots) dx dy = 0 . \quad (7)$$

Thus we have proven that to two leading orders in a^2/σ_a^2

$$\sum_i \left(\frac{\partial p_i}{\partial \mu_x} \right)^2 = \frac{a^2}{2\sigma_a^4} \int_{\text{all pixels}} (\vec{r} - \vec{\mu})^2 p^2(\vec{r}|\vec{\mu}, \sigma_a^2) = \frac{a^2}{8\pi\sigma_a^4} \quad (8)$$

$$\sum_i p_i \left(\frac{\partial p_i}{\partial \mu_x} \right)^2 = \frac{a^2}{2\sigma_a^4} \int_{\text{all pixels}} (\vec{r} - \vec{\mu})^2 p^3(\vec{r}|\vec{\mu}, \sigma_a^2) = \frac{a^4}{36\pi^2\sigma_a^6} . \quad (9)$$

2 Estimation with photon statistics

An EMCCD camera does not output photon counts; see Sect. 4. But the Poisson statistics of photon counts is simpler to discuss without the added stochastic element of the electron-multiplication process of the EMCCD. Since the statistics of the EMCCD output is easily derived from that of photon counts, we discuss the latter here.

Photon statistics is also what is considered in Ref. [1], with a 2D Gaussian point-spread function (PSF). The main Monte Carlo simulation results in Ref. [1] are derived analytically here. We also replace its widely used approximate interpolation formula for the localization error [1, Eq. (17)], with three formulas, which are exact in their treatment of background:

1. An error formula for the so-called Gaussian Mask Estimator (GME), which explains the “30% excess error” observed for GME in Ref. [1].

2. An error formula for WLS, which includes the effect of a constant offset—with such an offset one can, with a smaller sacrifice of precision, avoid a larger loss of precision that WLS suffers from on data sets with low expected counts in some pixels—which is what one has in the best data sets in super-resolution microscopy. In the literature, this problem of WLS has been patched with other ad hoc methods. We will not address them, since they all are sub-optimal compared to MLE.
3. An error formula for *Maximum Likelihood Estimation* with 2D Gaussian PSF (MLEwG). It demonstrates the superiority of MLE, a product of 20th century statistics.

2.1 Poisson statistics

Assume that we have recorded integer photon counts n_i in each pixel i in an image. These counts are interpreted theoretically as having expectation value $E_i(\theta)$, which depends on parameters θ to be estimated with GME, WLS, and MLEwG. The count n_i is assumed drawn from a Poisson distribution P_n . We do not know its expectation value $\langle n_i \rangle$, but if our theory is correct, then a set of *true values* θ^* exists such that

$$\langle n_i \rangle = E_i(\theta^*) = E_i^* , \quad (10)$$

where the last identity defines a shorthand notation. With this notation

$$P_{n_i} = e^{-E_i^*} \frac{E_i^{*n_i}}{n_i!} . \quad (11)$$

In particular, the variance is $\text{Var}(n_i) = E_i^*$.

2.2 Diffraction limited image of a point-source

Now consider a set of pixels, which cover little more than the diffraction limited image of *one* point-source. Our theory for the expected photon count in pixel $\#i$ among these pixels is

$$E_i(\theta) = N p_i + b^2 \quad (12)$$

with parameters $\theta = (N, \vec{\mu}, \text{“width parameter”}, b^2)$, where N is the total expected number of photons with origin in the point source and detected in the whole image plane, p_i is their probability distribution across the whole image plane, b^2 is the expected number of photons per pixel from a background with spatially constant expectation value across the image.

Obviously, $\sum_{\text{all } i} p_i = 1$, where we have summed over all pixels in an infinite image plane. We assume that the restricted set of pixels, which we fit to in order to localize the center of an isolated point source, cover its diffraction limited image well enough to satisfy this identity, or, at least, the milder condition

$$\frac{\partial}{\partial \vec{\mu}} \sum_i p_i = \frac{\partial}{\partial \vec{\mu}} \int_{\text{pixels in } \Sigma_i} p(x, y) dx dy = \vec{0} . \quad (13)$$

This last condition is satisfied well even if the set of pixels does not cover the tail of the PSF well, as long as the set of pixels in the sum is chosen fairly symmetrically about the spot, or has its edges where the tail decreases slowly or not at all. This last situation is realized when we fit a 2D Gaussian to real data and treat the shoulders of the true PSF as a constant background.

We will use also the closely related conditions

$$\int_{\text{pixels in } \Sigma_i} \Delta_{\vec{r}} ((\vec{r} - \vec{\mu})^2 p^2(\vec{r})) \, dx \, dy = \vec{0} \quad (14)$$

$$\int_{\text{pixel } \#i} \Delta_{\vec{r}} ((\vec{r} - \vec{\mu})^2 p^3(\vec{r})) \, dx \, dy = \vec{0} \, , \quad (15)$$

which follow from the divergence theorem and the vanishing/antisymmetry of the gradients of $(\vec{r} - \vec{\mu})^2 p^2(\vec{r})$ and $(\vec{r} - \vec{\mu})^2 p^3(\vec{r})$ on the edge of the area integrated over.

2.3 Unweighted least-squares fit: Gaussian Mask Estimator (GME)

Unweighted least-squares fitting results in the so-called Gaussian Mask Estimator (GME), when the fitted PSF p is a 2D Gaussian.

2.3.1 Unweighted Fitting

In general one minimizes the unweighted sum of squares

$$\chi^2 = \sum_i (n_i - E_i)^2 \, . \quad (16)$$

The minimizing parameter set $\hat{\theta}((n_i)_i)$ is our least-squares estimate for the truth θ^* . With the notation

$$E_{i,a} = \frac{\partial E_i}{\partial \theta_a} \, ; \quad E_{i,a,b} = \frac{\partial^2 E_i}{\partial \theta_a \partial \theta_b} \, , \quad (17)$$

the estimate $\hat{\theta}((n_i)_i)$ is the solution to the stationarity conditions

$$\sum_i (n_i - E_i) E_{i,a} = 0 \quad \text{for all } a. \quad (18)$$

Here $\sum_i E_i E_{i,a} = \frac{\partial}{\partial \theta_a} \sum_i \frac{1}{2} E_i^2$, which vanishes for E_i as in Eq. (12) and $\theta_a = \mu_x$ or μ_y . Thus the stationarity condition in this case reads

$$\sum_i n_i \frac{\partial p_i}{\partial \vec{\mu}} = \vec{0} \, , \quad (19)$$

which, using Eq. (5), can be rewritten

$$\vec{\mu} = \frac{\sum_i n_i \vec{r}_i p_i}{\sum_j n_j p_j} \, . \quad (20)$$

This is the formula for the centroid, except the sums are “masked” with the Gaussian p_i , hence its name *Gaussian Mask Estimator*. Since p_i depends on $\vec{\mu}$, this is an implicit equation for $\vec{\mu}$, which can be solved only numerically. The variance of the resulting estimate can be found explicitly, however.

2.3.2 Variance of Gaussian Mask Estimator

With reasonable statistics, the estimate $\hat{\theta}((n_i)_i)$ is close to the truth, θ^* , for most images $(n_i)_i$. To study its fluctuations about this truth, we approximate χ^2 by its second order Taylor polynomial in $\Delta\theta = \theta - \theta^*$, or, mathematically equivalently, linearize the stationarity condition in $\Delta\theta$. To this end, it is convenient to introduce the shorthand $E_{i,a}^* = E_{i,a}(\theta^*)$, $E_{i,a,b}^* = E_{i,a,b}(\theta^*)$, and the notation

$$\Delta n_i = n_i - \langle n_i \rangle = n_i - E_i^* . \quad (21)$$

Note, however, that we do *not* assume that Δn_i is small compared to its expectation value E_i^* : In case of low background b^2 , Δn_i is comparable to or even larger than E_i^* in the tail of the PSF, which makes WLS an unreliable estimator, as we shall see below, but does not destabilize unweighted least-squares fitting, as we shall see now.

Linearizing the stationarity equations (18) in $\Delta\theta$ gives

$$\sum_b \left(\sum_i E_{i,a}^* E_{i,b}^* - \sum_i \Delta n_i E_{i,a,b}^* \right) \Delta\theta_b = \sum_i \Delta n_i E_{i,a}^* . \quad (22)$$

Here the left-hand-side describes multiplication of the vector $(\Delta\theta_b)_b$ with a matrix I with elements $I_{a,b}$. Unless the sum over i includes only few pixels with only low photon counts, the second term in this matrix, $-\sum_i \Delta n_i E_{i,a,b}^*$, makes a small contribution to the matrix in comparison with the first term, because the sign of Δn_i is a random function of i . It also helps that in pixels with $E_i^* \gg 1$, $\Delta n_i = \mathcal{O}(\sqrt{E_i^*}) \ll E_i^*$. We consequently leave out the second term, and have

$$\Delta\theta_a = \sum_b (I^{-1})_{a,b} \sum_i \Delta n_i E_{i,b}^* \quad (23)$$

with

$$I_{a,b} = \sum_i E_{i,a}^* E_{i,b}^* , \quad (24)$$

hence the covariance matrix

$$\langle \Delta\theta_a \Delta\theta_b \rangle = \sum_{a',b'} (I^{-1})_{a,a'} \left(\sum_i E_i^* E_{i,a'}^* E_{i,b'}^* \right) (I^{-1})_{b',b} , \quad (25)$$

where we have used that $\langle \Delta n_i \Delta n_j \rangle = \delta_{i,j} E_i^*$.

For E_i as in Eq. (12), $E_{i,a}^*$ is an even, respectively odd, function of $\vec{r}_i - \vec{\mu}^*$ for subscript a denoting derivative w.r.t. N , b^2 , and σ_a^2 , respectively μ_x and μ_y . Consequently, the matrix I is block diagonal. A similar argument gives that the block for which subscript a denotes derivative w.r.t. μ_x and μ_y , is diagonal, i.e.,

$$\langle \Delta\mu_x \Delta\mu_y \rangle = 0 \quad (26)$$

$$\langle (\Delta\mu_x)^2 \rangle = \langle (\Delta\mu_y)^2 \rangle = \frac{\sum_i E_i^* (E_{i,a}^*)^2}{\left(\sum_j (E_{j,a}^*)^2 \right)^2} = \frac{\sum_i \left(p_i^* + \frac{b^{*2}}{N^*} \right) \left(\frac{\partial p_i^*}{\partial \mu_x} \right)^2}{N^* \left(\sum_j \left(\frac{\partial p_j^*}{\partial \mu_x} \right)^2 \right)^2} \quad (27)$$

Here $p_i^* = p_i(\vec{\mu}^*, \sigma_a^{*2})$ with p_i given in Eq. (5), so

$$\sum_i \left(\frac{\partial p_i}{\partial \mu_x} \right)^2 = \frac{a^4}{\sigma_a^4} \sum_i (x_i - \mu_x)^2 p^2(\vec{r}_i | \vec{\mu}, \sigma_a^2) = \frac{a^4}{2\sigma_a^4} \sum_i (\vec{r}_i - \vec{\mu})^2 p^2(\vec{r}_i | \vec{\mu}, \sigma_a^2) \quad (28)$$

and similarly

$$\sum_i p_i \left(\frac{\partial p_i}{\partial \mu_x} \right)^2 = \frac{a^6}{2\sigma_a^4} \sum_i (\vec{r}_i - \vec{\mu})^2 p^3(\vec{r}_i | \vec{\mu}, \sigma_a^2) . \quad (29)$$

Using the results of Subsection 1.3, we thus arrive at

$$\text{Var}(\hat{\mu}_x) = \frac{\sigma_a^{*2}}{N^*} \left(\frac{16}{9} + \frac{8\pi\sigma_a^{*2}b^{*2}}{N^*a^2} \right) . \quad (30)$$

This formula is our result for the variance of the Gaussian Mask Estimator. It is used in practice by substituting our estimate for the parameters (N, σ_a^2, b^2) for their true values, see **Fig. 4** in the paper. Our result differs from Eq. (17) in Ref. [1] in two ways. First, it accounts systematically for the dependence on finite pixel area to two leading orders: in [1, Eq. (17)] the a^2 -dependence of the term proportional to b^2 was left out. Second, it has the factor $16/9$ on $\frac{\sigma_a^{*2}}{N^*}$ where [1, Eq. (17)] has a factor 1.

Our formula describes the results that were obtained in Ref. [1] by Monte Carlo simulation of photon counts distributed with Gaussian PSF in 1D. The 30% “excess error,” “excess noise,” and “excess uncertainty of localization” discussed in Ref. [1] after observing it in Figs. 3, 4, and 5 there, is explained by $\sqrt{16/9} = 4/3$ being 30% in excess of 1. The reason [1, Eq. (17)] underestimates the true error on the GME is exactly the one conjectured in Ref. [1, p.2781]: Failure to model the noise correctly.

2.4 Weighted least-squares fit (WLS)

Some practitioners consider this method more correct than unweighted least-squares fitting, hence an improvement of GME. It is, as long as one only fits to large pixel counts. For the desirable and very relevant case of low background, however, it suffers a loss of precision without signalling that it does so, unless one specifically looks for warning signs. This problem of WLS is well known in other contexts [2]; we explain it, and how it can be patched with a zero-offset at the cost of some precision. We give a formula for the variance of this estimator. It shows how precision decreases with increasing zero-offset. At very large zero-offset one recovers the GME. At vanishing zero-offset, one recovers the precision of MLEwG, with said caveat against instability. Less blunt fixes of this problem exist, but they are as ad hoc by nature, cause almost the same loss of precision, are analytically less tractable, and as sub-optimal compared to MLE. We consequently do not treat them here, but give only this one, very tractable example of an ad hoc patch of WLS.

2.5 Weighted least-squares fit with offset

In Ref. [1] GME is improved by introducing experimental weights $1/n_i$ in the sum of squares, i.e., by minimizing

$$\chi_{\text{WLS}}^2 = \sum_i \frac{(n_i - E_i)^2}{n_i} . \quad (31)$$

This is referred to as “full least-squares.” In the literature it is also called “modified least-squares” [2], the modification being use of n_i in the denominator in Eq. (31) instead of the expectation value of n_i . It has the stationarity conditions

$$\sum_i \frac{n_i - E_i}{n_i} E_{i,a} = 0 \quad \text{for all } a. \quad (32)$$

Obviously, pixels with $n_i = 0$ are infinitely important in this sum, and will make a program crash due to division with zero.

Pixels in an EMCCD camera, however, output an integer representing an electron count, which for zero photon-counting events is distributed as a narrow Gaussian around an offset that for our camera is ~ 1000 . By looking at outputs from the darkest pixels in an image, we obtain this offset as the center of a Gaussian that is a few units wide.

The offset ensures positive output, but should be subtracted in order to make WLS conform to its name. If one does not, one has a hybrid method with a nature that falls between WLS and GME, depending on E_i^* , i.e., depending on the values of N and b^2/a^2 in an imaged point source. If they are so small that the pixel output is comparable to 1000 for all pixels in a fit, one is actually using GME. If one subtracts the offset, say 1000, from all pixel outputs, one will have non-positive counts in some pixels. If one subtracts some of the offset, but leaves enough to ensure sufficiently positive counts everywhere to patches the problem of division with zero or very small numbers, one has a hybrid between GME and WLS. When only the tails of the PSF have low photon counts, and its core has large counts, this patch costs only a small decrease in precision. We give formulas for this below.

Note, however, that these formulas are only valid if the offset is much larger than the expected output for one photon-counting event. Low precision is not ensured against merely by an offset that ensures positive output. That only ensures against division by zero. We have illustrated in **Fig. 4** in our paper what this leads to: A method with substantially *less* precision than GME and *no warnings* to its user that it has failed in this manner. The formula for its variance given below is consequently not valid, but seriously underestimates the actual error on estimates.

We cannot recommend this hybrid, nor WLS in any form, when precision is desired and low pixel counts occur. We recommend MLE as optimally precise. However, old habits tend to linger, so we now detail the properties of WLS, including its problem with low photon counts. We provide a formula for its precision, which is valid when it does not have that problem, either because of low counts, or because

a constant has been added to all counts, the off-set. Note that such an additive constant can be chosen to avoid critically low counts in pixels where they may occur, while it simultaneously is a negligible addition in pixels with large expected counts. So its effect on precision differs little from that of methods which only add counts ad hoc in pixels with realized low counts.

Returning to photon statistics, assume that the pixel output is $n_i + b_{\text{offset}}^2$ when n_i photons are recorded. Here b_{offset}^2 can with advantage be arranged to be less than the vendor-set value near 1000, simply by subtracting a constant value from all pixel outputs. (Which value to subtract depends on the gain, the expected photon count per pixel, and one's temper, since one must choose between high stability paid for by reduced precision, or higher precision paid for with risk of instability. Our only recommendation is to avoid this choice altogether by using MLEwG instead of WLS.

Then our theory for the expectation value of this pixel output is $E_i + b_{\text{offset}}^2$ with $E_i^* = \langle n_i \rangle$ unknown and b_{offset}^2 known. We minimize

$$\chi_{\text{offset}}^2 = \sum_i \frac{(n_i - E_i)^2}{n_i + b_{\text{offset}}^2} . \quad (33)$$

The stationarity conditions now are

$$\sum_i \frac{n_i - E_i}{n_i + b_{\text{offset}}^2} E_{i,a} = 0 \quad \text{for all } a. \quad (34)$$

Linearizing them w.r.t. $\Delta\theta$ at the true values θ^* gives

$$\sum_b \left(\sum_i \frac{E_{i,a}^* E_{i,b}^* - \Delta n_i E_{i,a,b}^*}{\Delta n_i + E_i^* + b_{\text{offset}}^2} \right) \Delta\theta_b = \sum_i \frac{\Delta n_i E_{i,a}^*}{\Delta n_i + E_i^* + b_{\text{offset}}^2} . \quad (35)$$

Denominators in both sums here can vanish or turn negative for E_i^* of order 1, unless we introduce the offset b_{offset}^2 . By choosing $b_{\text{offset}}^2 \gg 1$, we avoid this happening in any pixel, and also avoid the bias introduced when a pixel with, e.g., $E_i^* = 4$ by a fluctuation occurs with weight $1/(-3 + 4)$ because $\Delta n_i = -3$.

With $b_{\text{offset}}^2 \gg 1$ one always has $\Delta n_i + E_i^* + b_{\text{offset}}^2 \approx E_i^* + b_{\text{offset}}^2$. Using this in the denominators in Eq. (35), it reads

$$\sum_b \left(\sum_i \frac{E_{i,a}^* E_{i,b}^*}{E_i^* + b_{\text{offset}}^2} - \sum_i \frac{\Delta n_i E_{i,a,b}^*}{E_i^* + b_{\text{offset}}^2} \right) \Delta\theta_b = \sum_i \frac{\Delta n_i E_{i,a}^*}{E_i^* + b_{\text{offset}}^2} . \quad (36)$$

where the second term on the left-hand-side is negligible compared to the first term because of the random signs of Δn_i and because Δn_i is of order $\sqrt{E_i^*} \ll E_i^*$ where $E_i^* \gg 1$. We consequently drop this term from the linearized expression, exactly like we dropped a similar term in the same place in the discussion of unweighted least-squares fitting, and have

$$\sum_b I_{a,b} \Delta\theta_b = \sum_i \frac{\Delta n_i E_{i,a}^*}{E_i^* + b_{\text{offset}}^2} \quad (37)$$

with

$$I_{a,b} = \sum_i \frac{E_{i,a}^* E_{i,b}^*}{E_i^* + b_{\text{offset}}^2} , \quad (38)$$

hence the covariance matrix

$$\langle \Delta \theta_a \Delta \theta_b \rangle = \sum_{a',b'} (I^{-1})_{a,a'} \left(\sum_i \frac{E_i^* E_{i,a'}^* E_{i,b'}^*}{(E_i^* + b_{\text{offset}}^2)^2} \right) (I^{-1})_{b',b} , \quad (39)$$

where we have used that $\langle \Delta n_i \Delta n_j \rangle = \delta_{i,j} E_i^*$.

Like in the case of unweighted least-squares fitting, the matrix I is block diagonal, and the block relating to fluctuations in $\vec{\mu}$ is diagonal, so

$$\begin{aligned} \langle \Delta \mu_x \Delta \mu_y \rangle &= 0 \\ \langle (\Delta \mu_x)^2 \rangle &= \frac{\sum_i \frac{E_i^* (E_{i,a}^*)^2}{(E_i^* + b_{\text{offset}}^2)^2}}{\left(\sum_j \frac{(E_{j,a}^*)^2}{E_j^* + b_{\text{offset}}^2} \right)^2} \\ &= \frac{\sum_i \frac{(p_i^* + \frac{b^{*2}}{N^*}) \left(\frac{\partial p_i^*}{\partial \mu_x} \right)^2}{\left(p_i^* + \frac{b^{*2}}{N^*} + \frac{b_{\text{offset}}^2}{N^*} \right)^2}}{N^* \left(\sum_j \frac{\left(\frac{\partial p_j^*}{\partial \mu_x} \right)^2}{p_j^* + \frac{b^{*2}}{N^*} + \frac{b_{\text{offset}}^2}{N^*}} \right)^2} = \frac{2}{N^*} \frac{\int \frac{(p^* + \frac{b^{*2}}{N^* a^2}) \left(\frac{\partial p^*}{\partial \mu} \right)^2}{\left(p^* + \frac{b^{*2}}{N^* a^2} + \frac{b_{\text{offset}}^2}{N^* a^2} \right)^2} dx dy}{\left(\int \frac{\left(\frac{\partial p^*}{\partial \mu} \right)^2}{p^* + \frac{b^{*2}}{N^* a^2} + \frac{b_{\text{offset}}^2}{N^* a^2}} dx dy \right)^2} . \end{aligned} \quad (41)$$

We evaluate this result for the case of p being the 2D Gaussian given in Eq. (5) using

$$\int \frac{\left(\frac{\partial p}{\partial \mu} \right)^2}{(p + c^2)^2} dx dy = \frac{2\pi}{\sigma_a^4} \int_0^\infty \frac{r^3 p^2}{p + c^2} dr = \frac{2}{\sigma_a^2} \mathcal{I}_1(2\pi \sigma_a^2 c^2) , \quad (42)$$

where we have introduced the function

$$\mathcal{I}_1(\tau) = - \int_0^1 \frac{t \ln t}{t + \tau} dt = 1 + \int_0^1 \frac{\ln t}{1 + t/\tau} dt , \quad (43)$$

and

$$\int \frac{\left(\frac{\partial p}{\partial \mu} \right)^2}{(p + c^2)^2} dx dy = \frac{2\pi}{\sigma_a^4} \int_0^\infty \frac{r^3 p^2}{(p + c^2)^2} dr = 4\pi \mathcal{I}_2(2\pi \sigma_a^2 c^2) , \quad (44)$$

where we have introduced the function

$$\begin{aligned} \mathcal{I}_2(\tau) &= - \int_0^1 \frac{t \ln t}{(t + \tau)^2} dt \\ &= \frac{1}{\tau} \left(- \int_0^1 \frac{\ln t}{1 + t/\tau} dt + \int_0^1 \frac{\ln t}{(1 + t/\tau)^2} dt \right) . \end{aligned} \quad (45)$$

Thus we find

$$\text{Var}(\hat{\mu}_x^2) = \text{Var}(\hat{\mu}_y^2) = \frac{\sigma_a^{*2}}{N^*} \left(\frac{1}{\mathcal{I}_1(\tau)} - \frac{2\pi\sigma_a^{*2}b_{\text{offset}}^2}{N^*a^2} \frac{\mathcal{I}_2(\tau)}{\mathcal{I}_1(\tau)^2} \right) \quad (46)$$

with $\tau = 2\pi\sigma_a^{*2}(b^{*2} + b_{\text{offset}}^2)/(N^*a^2)$.

In the case of b^{*2} large, we have large photon counts in all pixels, and optimize precision by choosing $b_{\text{offset}}^2 = 0$. In this case the variance simplifies to

$$\text{Var}(\hat{\mu}_x^2) = \text{Var}(\hat{\mu}_y^2) = \frac{\sigma_a^{*2}}{N^*} \frac{1}{\mathcal{I}_1(\tau)} \quad (47)$$

with $\tau = 2\pi\sigma_a^{*2}b^{*2}/(N^*a^2)$.

This result for the variance of μ_x and μ_y we also find for MLEwG in the next subsection, but in that case, as opposed to the present case, the result is valid also for small values of b^{*2} . **Figure 4i** shows that if WLS with small b_{offset}^2 is applied to images of fluorescent beads with little background fluorescence, it results in estimates that scatter significantly more than is described by a Gaussian with the variance in Eq. (47). **Figure 4e** shows that this larger scatter is larger than that of GME! The scatter of WLS can be reduced to values between the scatter of GME and MLE by a suitable choice of $b_{\text{offset}}^2 \gg 1$ in units of photon counts per pixel (not shown). That, however, is not optimal estimation. Maximum Likelihood estimation is optimal, as we shall see.

2.6 Maximum Likelihood Estimation (MLE)

2.6.1 MLE

MLE is done by maximizing Fisher's so-called *Likelihood* function

$$\mathcal{L}(\theta|(n_i)_i) = \prod_i P_{n_i}(\theta) = e^{-\sum_i E_i} \prod_i \frac{E_i^{n_i}}{n_i!} \quad (48)$$

or its logarithm

$$\ell = \ln \mathcal{L} = \sum_i (-E_i + n_i \ln E_i - \ln(n_i!)) \quad (49)$$

The maximizing parameter set $\hat{\theta}$ solves the stationarity equations

$$\sum_i \frac{n_i - E_i}{E_i} E_{i,a} = 0 \quad \text{for all } a. \quad (50)$$

2.6.2 Why MLE is better than WLS

These stationarity equations differ from those of WLS only by having in the denominator not n_i , but its expectation value E_i . This makes a world of difference for small values of E_i , since E_i is not a fluctuating quantity, while n_i is. In MLE's stationarity equations, n_i occurs only in the numerator, where its fluctuations cause no problems, even if they are larger than the expected value of n_i .

2.6.3 Variance of MLEwT

As above, we find the variance of the estimated parameter values by linearizing the stationarity equations w.r.t. $\Delta\theta$ at the true parameter values θ^* . As above, terms containing the product $\Delta\theta_a \Delta n_i$ are negligible after summing over i , resulting in

$$\sum_b I_{a,b} \Delta\theta_b = \sum_i \frac{\Delta n_i E_{i,a}^*}{E_i^*} \quad (51)$$

with

$$I_{a,b} = \sum_i \frac{E_{i,a}^* E_{i,b}^*}{E_i^*}, \quad (52)$$

which in this case of MLE is called *Fisher's Information Matrix*. Using as before that $\langle \Delta n_i \Delta n_j \rangle = E_i^* \delta_{i,j}$, we have

$$\langle \Delta\theta_a \Delta\theta_b \rangle = (I^{-1})_{a,b}. \quad (53)$$

In particular, the identical variances of μ_x and μ_y can be found for MLE by calculating their corresponding diagonal elements of Eq. (53). In this manner, using the theoretical PSF to calculate the expectation values E_i^* and $E_{i,a}^*$ in Eq. (52), we find the variances of μ_x and μ_y for MLEwT.

In **Fig. 4** we compare experimental results for the variance of MLEwT with *twice* the value given in Eq. (53)—twice because of excess noise; see Sect. 4—and we find perfect agreement.

2.6.4 Variance of MLEwG

For E_i as in Eq. (12) and p as in Eq. (3),

$$\text{Var}(\hat{\mu}_x) = \text{Var}(\hat{\mu}_y) = \frac{\sigma_a^{*2}}{N^*} \frac{1}{\mathcal{I}_1(\tau)} = \frac{\sigma_a^{*2}}{N^*} \left(1 + \int_0^1 \frac{\ln t}{1 + t/\tau} dt \right)^{-1} \quad (54)$$

with $\tau = 2\pi\sigma_a^{*2}b^{*2}/(N^*a^2)$ and $\mathcal{I}_1(\tau)$ defined in Eq. (43).

Excess noise, see Sect. 4, enlarges this variance by a factor two, and for this reason it is *twice* the variance in Eq. (54) that is compared to our experimental results for the variance of MLEwG in **Fig. 4**. They are seen to fully agree. We find also that this variance is indistinguishable in practise from the variance for MLEwT, which gives the information limit. Since MLEwG is as precise as MLEwT, is mathematically vastly simpler to use than MLEwT, and is easier to apply than WLS with zero-offset, everything—except accuracy, see discussion in the paper—favors the use of MLEwG. In MATLAB, e.g., MLE is just another function call replacing a call to least-squares, albeit a more time-consuming one.

2.6.5 The function $\mathcal{I}_1(\tau)$

The function $\mathcal{I}_1(\tau)$ that describes the dependence on b^2 for arbitrary values, cannot be given in terms of any known function. It is easily computed numerically, and this is how to treat it, because it is not easily approximated analytically.

Mathematically, $\mathcal{I}_1(\tau)$ has an essential singularity at $\tau = 0$, i.e., at zero background, $b^2 = 0$. So there is no simple approximation to our result in the case of small background, except an asymptotic formula near vanishing background, which is not useful, since the shoulder of the experimental PSF always contributes to what we call background in the fitted 2D Gaussian plus constant.

Mathematically, $\mathcal{I}_1(\tau)$ has a Taylor expansion in $1/\tau$ at $\tau = \infty$. This expansion unfortunately has finite radius of convergence. It converges only for $\tau > 1$, so it requires $\tau \gg 1$ in order for a few first terms to yield a good approximation. This regime is not of practical interest.

These are two of several reasons why the interpolation formula in Eq. (57) below does not interpolate well in the range of practical interest; see the next subsection.

2.6.6 The centroid estimator

A special case of MLE arises when the fitted function in MLE is assumed a 2D Gaussian with no background. In this case the ML estimator is the so-called centroid (or center-of-intensity) estimator

$$\hat{\mu} = \frac{1}{N} \sum_i \vec{r}_i n_i . \quad (55)$$

Its variance is the familiar result for the variance of an average over N statistically independent measurements, i.e.

$$\text{Var}(\hat{\mu}_x) = \text{Var}(\hat{\mu}_y) = \frac{\sigma_a^{*2}}{N^*} , \quad (56)$$

assuming the PSF has finite variance. However, if the PSF has infinite variance, as the case is for the PSF for an isotropic dipole distribution, (see paper and its **Fig. 3**, Sect. 3, and **Supplementary Fig. 1**) the centroid inherits its infinite variance. In practice, the variance of the centroid is limited by finite image area and various clipping methods. Several variations are considered in Ref. [3], where they are tested on 560 nm fluorescent beads, and it was found experimentally that MLE using a 2D Gaussian outperforms the various centroid estimators.

2.7 A too-good-to-be-true interpolation formula for precision

In [1, Eq. (17)] the variance associated with estimates based on least-squares fitting is approximated with the following interpolation formula,

$$\text{Var}(\hat{\mu}_x) = \frac{\sigma^2}{N} \left(1 + \frac{a^2}{12\sigma^2} + \frac{8\pi\sigma^2 b^2}{Na^2} \right) , \quad (57)$$

where b^2 is the expected number of background photons per pixel, N is the estimated number of photons in an image, and σ^2 is the fitted variance of the 2D Gaussian PSF.

Comparison of Eq.(54) with Eq.(57) shows that they agree for $b = 0$. This case never occurs, however, since the shoulders of the experimental PSF are treated as background.

In the opposite extreme of very large background, $b^2/a^2 \gg N/(8\pi\sigma^2)$, $\tau \gg 1$, and to leading order in an expansion in $1/\tau$, Eq. (54) reads $\text{Var}(\hat{\mu}_x^2) = \text{Var}(\hat{\mu}_y^2) = 4\tau \frac{\sigma_a^{*2}}{N^*}$. This is an improvement of Eq.(57) in the same limit, which is missing a factor $1 + \frac{a^2}{12\sigma^2}$.

Figure 4 compares the variances of the four estimation algorithms and Eq. (17) in [1] for photons distributed according to the theoretical PSF with wavelength $\lambda = 605$ nm. As is clear (**Supplementary Fig. 1**), this PSF is well approximated by a 2D Gaussian plus a constant. The variances of the Gaussian-based estimators were evaluated for the parameter values obtained from their fits of a 2D Gaussian plus a constant. These parameter values varied little between GME and MLEwG.

3 The theoretical point spread function

The experimental PSF is known theoretically with great accuracy: The fluorophore molecule is much smaller than the wavelength of its emitted light, Angstroms compared to hundreds of nanometers. Consequently, the molecule is a point-source to an extremely good approximation—a dipole point-source, since it emits via a molecular dipole transition. The electromagnetic field from an emitting dipole is known analytically from Maxwell’s equations. The Kirchhoff Vector Approximation describes the diffraction of this field in a circular aperture extremely well, because the wavelength of the diffracted light is much smaller than the dimensions of the microscope, sub-micrometer compared to centimeters.

3.1 Fixed-dipole case

The theory for the point spread function (PSF) of a dipole emitter with arbitrary orientation with respect to an interface is summarized in [4] and references therein. The PSF p is proportional to the z -component of the Poynting vector [5],

$$p \propto E_x B_y^* - E_y B_x^* , \quad (58)$$

where \vec{E} and \vec{B} are the electric and magnetic fields respectively, and the superscript $*$ means complex conjugate. In general, this quantity depends on the orientation of the emitting dipole. The point spread function for a dipole located close to the coverslip, in focus, and oriented with azimuth α and polar angle β with respect to

the optical axis, writing all angular dependencies explicitly, is

$$\begin{aligned}
p(x', y' | \alpha, \beta) \propto & \frac{\sin^2 \beta}{4} (|\mathcal{I}[A]|^2 + |\mathcal{I}[B]|^2 - 2 \cos(2\phi' - 2\alpha) \Re(\mathcal{I}[A] \mathcal{I}^*[B])) \\
& + \sin \beta \cos \beta \cos(\phi' - \alpha) (\Im(\mathcal{I}[B] \mathcal{I}^*[C]) - \Im(\mathcal{I}[A] \mathcal{I}^*[C])) \\
& + \cos^2 \beta |\mathcal{I}[C]|^2,
\end{aligned} \tag{59}$$

where \Re and \Im denote, respectively, the real and imaginary part, while

$$A \equiv (E_p^\parallel - E_s^\parallel) J_0(k' \rho' \eta') , \tag{60}$$

$$B \equiv (E_p^\parallel + E_s^\parallel) J_2(k' \rho' \eta') , \tag{61}$$

$$C \equiv E_p^\perp J_1(k' \rho' \eta') . \tag{62}$$

Here J_n is the Bessel function of the first kind and order n , k' is the magnitude of the wave vector in the imaging medium (air), η' is the wave vector's angle with the optical axis in the imaging medium, and ρ' is the radial distance in the image plane from the optical axis to the point at which p is being evaluated, e.g. $(x', y') = \rho'(\cos \phi', \sin \phi')$. The electric field components $E_p^\parallel(\eta)$, $E_s^\parallel(\eta)$, and $E_p^\perp(\eta)$ are given in [4]. The integration operator \mathcal{I} is

$$\mathcal{I}[\dots] \equiv \int_0^{\eta'_{\max}} \frac{\eta'}{\sqrt{\cos \eta}} [\dots] d\eta' , \tag{63}$$

where η is the angle between the optical axis and the wave vector inside the objective. The angles η and η' are related through Abbe's sine condition. The maximal value of η' is given by

$$\eta'_{\max} = \arcsin(NA/Mn') \approx \frac{NA}{Mn'} , \tag{64}$$

where NA is the numerical aperture of the microscope, M is the magnification of the composite optics (the objective and relay lens gives $M = 250$ in our case), and n' is the refractive index of the imaging medium ($n' = 1$ in air, our case). Even for high-aperture objectives, this angle remains much smaller than one—less than 0.01 in our case. This simplifies functions of η' , allowing us to arrive at Eq. (63).

Normalization of p in Eq. (59) at any given values of α and β gives

$$\begin{aligned}
p(x', y' | \alpha, \beta) = & \mathcal{N}(\beta) (\sin^2 \beta I_\parallel (p_\parallel(\rho') + \cos(2\phi' - 2\alpha) \Delta p_\parallel(\rho')) \\
& + \sin \beta \cos \beta \cos(\phi' - \alpha) I_\times p_\times(\rho') \\
& + \cos^2 \beta I_\perp p_\perp(\rho')) .
\end{aligned} \tag{65}$$

where we have introduced the normalized distributions

$$p_\parallel \equiv \frac{1}{4} (|\mathcal{I}[A]|^2 + |\mathcal{I}[B]|^2) / I_\parallel \tag{66}$$

$$p_\perp \equiv |\mathcal{I}[C]|^2 / I_\perp , \tag{67}$$

and the functions

$$\begin{aligned}\Delta p_{\parallel} &\equiv -\frac{1}{2}\Re(\mathcal{I}[A]\mathcal{I}^*[B])/I_{\parallel} \\ p_{\times} &\equiv \Im((\mathcal{I}[B] - \mathcal{I}[A])\mathcal{I}^*[C])/I_{\times} \ ,\end{aligned}\tag{68}$$

all normalized but Δp_{\parallel} ,

$$I_{\parallel} \equiv 2\pi \int_0^{\infty} \rho' d\rho' \frac{1}{4} (|\mathcal{I}[A]|^2 + |\mathcal{I}[B]|^2) \tag{69}$$

$$I_{\perp} \equiv 2\pi \int_0^{\infty} \rho' d\rho' |\mathcal{I}[C]|^2 \tag{70}$$

$$I_{\times} \equiv 2\pi \int_0^{\infty} \rho' d\rho' \Im((\mathcal{I}[B] - \mathcal{I}[A])\mathcal{I}^*[C]) \ , \tag{71}$$

and $p(x', y'|\alpha, \beta)$ is normalized because

$$\mathcal{N}(\beta)^{-1} \equiv I_{\parallel} \sin^2 \beta + I_{\perp} \cos^2 \beta \ . \tag{72}$$

Note that for a dipole orthogonal to the image plane, the polar angle $\beta = 0$, and consequently, because $\mathcal{N}(0)I_{\perp} = 1$,

$$p(x', y'|\alpha, 0) = p_{\perp}(\rho') \ . \tag{73}$$

Thus we see that p_{\perp} is the PSF of a fixed dipole orthogonal to the image plane. Similarly, for a fixed dipole parallel to the image plane, $\beta = \pi/2$. Since $\mathcal{N}(\pi/2)I_{\parallel} = 1$, one has

$$p(x', y'|\alpha, \pi/2) = p_{\parallel}(\rho') + \cos(2\phi' - 2\alpha) \Delta p_{\parallel}(\rho') \ . \tag{74}$$

Thus we see that p_{\parallel} is the PSF of an isotropic distribution of dipoles parallel to the image plane and located at the same point in that plane. We also see that Δp_{\parallel} describes the radial dependence of the anisotropic part of the PSF of a fixed dipole parallel to the image plane.

3.2 In-plane dipoles and 2D Gaussians

For dipoles that appear approximately parallel to the image plane, a warning is appropriate: Their images are fitted fairly well by a 2D Gaussian with two different variances. Attempts to localize fluorophores in this manner will result in large biases, however. The diffraction-limited image is hundreds of nanometers wide, while we pursue localization with nanometer accuracy and precision. Only for a dipole exactly parallel to the image plane does the maximum in image intensity coincide with the dipole's location. The Gaussian fit locates the former, while the latter easily is tens of nanometers away for a dipole only approximately parallel to the image plane. Such a dipole's intensity distribution is slightly skewed. Equation (65) accounts for this skew, a Gaussian does not. Reference [6] demonstrates the danger of Gaussian fits to skewed distributions.

3.3 Isotropic case

We here consider the special case of a dipole source that is either an isotropically illuminated collection of dipole emitters located at the same point in space with isotropic distribution of orientations—e.g., a fluorescent bead with negligible radius compared to the wavelength emitted—or a single dipole that rotates so fast during its excited state that it in practice emits with any orientation with the same probability. The intensity distribution of light from both these sources is the intensity distribution of light from a dipole with fixed orientation, averaged over all possible orientations. Consequently, the PSF for both these sources is the PSF for a fixed dipole, i.e., Eq. (65) averaged uniformly over all possible orientations, because the light from the various dipoles superpose incoherently,

$$\begin{aligned} p_{\text{iso}}(x', y') &= \int \frac{d\Omega}{4\pi} p(x', y' | \alpha, \beta) \\ &= \int_0^{2\pi} \frac{d\alpha}{2\pi} \frac{1}{2} \int_0^\pi d\beta \sin \beta p(x', y' | \alpha, \beta) \\ &= \mathcal{N}_{\parallel} p_{\parallel}(\rho') + \mathcal{N}_{\perp} p_{\perp}(\rho') , \end{aligned} \quad (75)$$

where we have introduced the notation

$$\mathcal{N}_{\parallel} \equiv I_{\parallel} \int \frac{d\Omega}{4\pi} \mathcal{N}(\beta) \sin^2 \beta = I_{\parallel} \frac{1}{2} \int_0^\pi d\beta \sin^3 \beta \mathcal{N}(\beta) \quad (76)$$

$$\mathcal{N}_{\perp} \equiv I_{\perp} \int \frac{d\Omega}{4\pi} \mathcal{N}(\beta) \cos^2 \beta = I_{\perp} \frac{1}{2} \int_0^\pi d\beta \sin \beta \cos^2 \beta \mathcal{N}(\beta) . \quad (77)$$

From the definition of $\mathcal{N}^{-1}(\beta)$ in Eq. (72) it follows that

$$\mathcal{N}_{\parallel} + \mathcal{N}_{\perp} = 1 , \quad (78)$$

which also follows from the fact that p_{iso} , p_{\parallel} , and p_{\perp} all are normalized.

From the definition of \mathcal{N}_{\perp} above with $\mu \equiv \cos \beta$ and

$$r \equiv I_{\perp}/I_{\parallel} , \quad (79)$$

it follows that

$$\begin{aligned} \mathcal{N}_{\perp} &= \mathcal{N}_{\perp}(r) = I_{\perp} \frac{1}{2} \int_0^\pi d\beta \sin \beta \cos^2 \beta \mathcal{N}(\beta) = I_{\perp} \frac{1}{2} \int_{-1}^1 d\mu \mu^2 \mathcal{N}(\mu) \\ &= \frac{1}{2} \int_{-1}^1 d\mu \mu^2 \frac{I_{\perp}}{I_{\parallel}(1 - \mu^2) + I_{\perp}\mu^2} = \int_0^1 d\mu \mu^2 \frac{r}{(1 - \mu^2) + r\mu^2} \\ &= \int_0^1 d\mu \mu^2 \frac{r}{(1 - \mu^2) + r\mu^2} = \frac{r}{r-1} \int_0^1 d\mu \frac{(r-1)\mu^2 + 1 - 1}{1 + (r-1)\mu^2} \\ &= \frac{r}{r-1} \left(1 - \int_0^1 d\mu \frac{1}{1 + (r-1)\mu^2} \right) = \frac{r}{r-1} \left(1 - \frac{\arctan \sqrt{r-1}}{\sqrt{r-1}} \right) \end{aligned} \quad (80)$$

where the last integral is based on the fact that $r > 1$ —we found $r \simeq 1.5$ for our objectives. Similarly,

$$\mathcal{N}_{\parallel}(r) = \frac{1}{r-1} \left(2 \frac{\arctan \sqrt{r-1}}{\sqrt{r-1}} - 1 \right) . \quad (81)$$

When the exciting light is non-isotropic, this uniformly averaged PSF is still an accurate description for a freely rotating fluorophore that rotates sufficiently fast during its excited state. For a slowly rotating fluorophore or a fixed collection of fluorophores, such as those in a fluorescent bead, this isotropic PSF is, however, not adequate. Below, we generalize the PSF to describe non-isotropically excited fluorophores.

3.4 Polarized illumination in TIRF

In TIRF microscopy, the light that excites the fluorescent probes is an evanescent wave created under total internal reflection. The polarization of this light is non-isotropic, which means that dipoles of different orientations are excited with different probabilities per unit time. The PSF of a fluorescent bead is therefore not the simple uniform average over orientations that was calculated above. Instead the average should be weighted in accordance with the polarization.

Following Refs. [5, 7], we consider the refraction of a plane wave at an interface between two optically different media. The incoming light, traveling in the xz -plane in medium 1 with velocity \vec{v}_1 , has the angle θ_i with the z -axis and the transmitted light emerges at an angle θ_t at velocity \vec{v}_2 . Let the incoming light have polarized amplitudes A_{\parallel} and A_{\perp} which are both taken to be complex because they contain the constant part of the phase. (For circularly polarized light, e.g., they differ in phase by $\pm\pi/2$, where the sign depends on the handedness of the circular polarization.) The incident electric field thus is [5]

$$E_x^{(i)} = -A_{\parallel} \cos \theta_i e^{-i\tau_i} \quad (82)$$

$$E_y^{(i)} = A_{\perp} e^{-i\tau_i} \quad (83)$$

$$E_z^{(i)} = A_{\parallel} \sin \theta_i e^{-i\tau_i} , \quad (84)$$

where

$$\tau_i = \omega \left(t - \frac{x \sin \theta_i + z \cos \theta_i}{v_1} \right) \quad (85)$$

is the variable part of the phase. The amplitudes of the transmitted light is related to the incident light amplitudes as [5]

$$T_{\parallel} = \frac{2 \cos \theta_i}{n \cos \theta_i + \cos \theta_t} A_{\parallel} \quad (86)$$

$$T_{\perp} = \frac{2 \cos \theta_i}{\cos \theta_i + n \cos \theta_t} A_{\perp} , \quad (87)$$

where $n \equiv n_2/n_1$ with n_1 and n_2 the refractive indices of Medium 1, respectively 2. The transmitted electric field $\vec{E}^{(t)}$ is given by

$$E_x^{(t)} = -T_{||} \cos \theta_t e^{-i\tau_t} \quad (88)$$

$$E_y^{(t)} = T_{\perp} e^{-i\tau_t} \quad (89)$$

$$E_z^{(t)} = T_{||} \sin \theta_t e^{-i\tau_t} , \quad (90)$$

with

$$\tau_t = \omega \left(t - \frac{x \sin \theta_t + z \cos \theta_t}{v_2} \right) . \quad (91)$$

Total internal reflection occurs for $n_2 < n_1$. In this case, Snell's law

$$\sin \theta_t = \frac{\sin \theta_i}{n} \quad (92)$$

gives a complex value for θ_t , because θ_i here exceeds the critical value $\bar{\theta}_i$ given by $\sin \bar{\theta}_i = n$. In this regime

$$\cos \theta_t = \pm i \sqrt{\frac{\sin^2 \theta_i}{n^2} - 1} , \quad (93)$$

where only the positive sign is physically meaningful.

For total internal reflection, the pre-factor to the x -component of the electric field is found by insertion of Eq. (93), and is

$$\begin{aligned} T_{||} \cos \theta_t &= \frac{2 \cos \theta_i \cos \theta_t}{n \cos \theta_i + \cos \theta_t} A_{||} \\ &= \frac{2 \cos \theta_i i \sqrt{\sin^2 \theta_i / n^2 - 1}}{n \cos \theta_i + i \sqrt{\sin^2 \theta_i / n^2 - 1}} A_{||} \\ &= \frac{2 \cos \theta_i (\sin^2 \theta_i - n^2) + i 2 n^2 \cos^2 \theta_i \sqrt{\sin^2 \theta_i - n^2}}{n^4 \cos^2 \theta_i + \sin^2 \theta_i - n^2} A_{||} . \end{aligned} \quad (94)$$

Now write the complex pre-factor to $A_{||}$ as $x + iy$ and introduce polar coordinates (r, φ) in the complex (x, y) -plane: $x + iy = r \exp(i\varphi)$. Then

$$r = \sqrt{x^2 + y^2} = \frac{2 \cos \theta_i \sqrt{\sin^2 \theta_i - n^2}}{\sqrt{n^4 \cos^2 \theta_i + \sin^2 \theta_i - n^2}} \quad (95)$$

and

$$\varphi = \arctan \left(\frac{y}{x} \right) = \arctan \left(\frac{n^2 \cos \theta_i}{\sqrt{\sin^2 \theta_i - n^2}} \right) . \quad (96)$$

The negative sign in Eq. (88) contributes a phase shift of $-\pi$ to the x -component of the electric field. By defining

$$\delta_{\parallel} \equiv \arctan \left(\frac{\sqrt{\sin^2 \theta_i - n^2}}{n^2 \cos \theta_i} \right) \quad (97)$$

$$\delta_{\perp} \equiv \arctan \left(\frac{\sqrt{\sin^2 \theta_i - n^2}}{\cos \theta_i} \right) , \quad (98)$$

and proceeding in the same manner for the y - and z -components, and using that $\arctan(u) = \pi/2 - \arctan(1/u)$ in the x -component, the transmitted electric field becomes

$$E_x^{(t)} = \frac{2 \cos \theta_i \sqrt{\sin^2 \theta_i - n^2}}{\sqrt{n^4 \cos^2 \theta_i + \sin^2 \theta_i - n^2}} A_{\parallel} e^{-i(\kappa + \delta_{\parallel} + \pi/2)} \quad (99)$$

$$E_y^{(t)} = \frac{2 \cos \theta_i}{\sqrt{1 - n^2}} A_{\perp} e^{-i(\kappa + \delta_{\perp})} \quad (100)$$

$$E_z^{(t)} = \frac{2 \cos \theta_i \sin \theta_i}{\sqrt{n^4 \cos^2 \theta_i + \sin^2 \theta_i - n^2}} A_{\parallel} e^{-i(\kappa + \delta_{\parallel})} . \quad (101)$$

This result has been published previously, e.g. in [7].

3.5 PSF for isotropic distribution of dipoles in TIRF

A fluorescent molecule is described by its position and its electric dipole moment $\vec{\mu}$. Using polar angles (α, β) , its dipole moment is

$$\vec{\mu} = |\vec{\mu}| \begin{pmatrix} \cos \alpha \sin \beta \\ \sin \alpha \sin \beta \\ \cos \beta \end{pmatrix} , \quad (102)$$

and is excited by the electric field $\vec{E}^{(t)} \equiv \vec{E}$ with an amplitude

$$|\vec{E} \cdot \vec{\mu}|^2 = |\vec{\mu}|^2 |E_x \cos \alpha \sin \beta + E_y \sin \alpha \sin \beta + E_z \cos \beta|^2 . \quad (103)$$

To calculate the PSF of a fluorescent bead or a single molecule, which rotates slowly compared to the excitation lifetime, the rotational average should take Eq. (103) into account. However, a single image records all emissions that occur during its exposure time t_{msr} , which is much longer than the period of the electric wave. So effectively the weighting factor is the time average of Eq. (103). In general when calculating quantities that involve non-linear operations on the electric field in complex notation, only its real part should be used. However, when only the time average of a quadratic form is needed, the manipulations can be carried out directly on the complex form [5]. For the time average

$$\langle \cdots \rangle_{t_{\text{msr}}} \equiv \frac{1}{t_{\text{msr}}} \int_0^{t_{\text{msr}}} \cdots dt , \quad (104)$$

of Eq. (103), using that $|z|^2 = zz^*$, we find

$$\begin{aligned}
 \langle |\vec{E} \cdot \vec{\mu}|^2 \rangle_T &= |\vec{\mu}|^2 \left(|E_x|^2 \cos^2 \alpha \sin^2 \beta + |E_y|^2 \sin^2 \alpha \sin^2 \beta + |E_z|^2 \cos^2 \beta \right. \\
 &\quad + (E_x E_y^* + E_x^* E_y) \cos \alpha \sin \alpha \sin^2 \beta \\
 &\quad + (E_x E_z^* + E_x^* E_z) \cos \alpha \sin \beta \cos \beta \\
 &\quad \left. + (E_y E_z^* + E_y^* E_z) \sin \alpha \sin \beta \cos \beta \right) .
 \end{aligned} \tag{105}$$

From Eqs. (99)-(101) we obtain

$$|E_x|^2 = \frac{4 \cos^2 \theta (\sin^2 \theta - n^2)}{n^4 \cos^2 \theta + \sin^2 \theta - n^2} |A_{||}|^2 \tag{106}$$

$$|E_y|^2 = \frac{4 \cos^2 \theta}{1 - n^2} |A_{\perp}|^2 \tag{107}$$

$$|E_z|^2 = \frac{4 \cos^2 \theta \sin^2 \theta}{n^4 \cos^2 \theta + \sin^2 \theta - n^2} |A_{||}|^2 , \tag{108}$$

while at the interface at $x = z = 0$, where the fluorescent molecule is located, we find

$$\begin{aligned}
 E_x E_y^* + E_x^* E_y = 2\Re(E_x E_y^*) &= 2\Re \left(|E_x| e^{-i(\omega t + \pi/2 + \delta_{||})} |E_y| e^{i\pi/2} e^{i(\omega t + \delta_{\perp})} \right) \\
 &= 2\Re \left(|E_x| |E_y| e^{-i(\delta_{||} - \delta_{\perp})} \right) \\
 &= 2|E_x| |E_y| \cos(\delta_{||} - \delta_{\perp}) ,
 \end{aligned} \tag{109}$$

$$\begin{aligned}
 E_x E_z^* + E_x^* E_z = 2\Re(E_x E_z^*) &= 2\Re \left(|E_x| e^{-i(\omega t + \pi/2 + \delta_{||})} |E_z| e^{i(\omega t + \delta_{||})} \right) \\
 &= 2\Re \left(|E_x| |E_z| e^{-i\pi/2} \right) \\
 &= 0 ,
 \end{aligned} \tag{110}$$

and

$$\begin{aligned}
 E_y E_z^* + E_y^* E_z = 2\Re(E_y E_z^*) &= 2\Re \left(|E_y| e^{-i\pi/2} e^{-i(\omega t + \delta_{\perp})} |E_z| e^{i(\omega t + \delta_{||})} \right) \\
 &= 2\Re \left(|E_y| |E_z| e^{-i(\pi/2 + \delta_{||} - \delta_{\perp})} \right) \\
 &= 2|E_y| |E_z| \cos(\pi/2 + \delta_{||} - \delta_{\perp}) .
 \end{aligned} \tag{111}$$

Now, the PSF is calculated by the weighted average of $p(x', y' | \alpha, \beta)$, explicitly

$$p_{\text{pol}}(x', y') = \frac{\int_{4\pi} d\Omega \langle |\vec{E} \cdot \vec{\mu}|^2 \rangle_T p(x', y' | \alpha, \beta)}{\int_{4\pi} d\Omega \langle |\vec{E} \cdot \vec{\mu}|^2 \rangle_T} , \tag{112}$$

where $d\Omega = \sin \beta d\alpha d\beta$ is the surface element on the sphere. The normalization can be explicitly calculated as

$$\begin{aligned}
 \int_{4\pi} d\Omega \langle |\vec{E} \cdot \vec{\mu}|^2 \rangle_T &= |\vec{\mu}|^2 \int_0^\pi \sin \beta d\beta \int_0^{2\pi} d\alpha \langle |\vec{E} \cdot \vec{\mu}|^2 \rangle_T \\
 &= |\vec{\mu}|^2 \pi \int_0^\pi \sin \beta d\beta (|E_x|^2 + |E_y|^2) \sin^2 \beta + 2|E_z|^2 \cos^2 \beta \\
 &= |\vec{\mu}|^2 \frac{4\pi}{3} (|E_x|^2 + |E_y|^2 + |E_z|^2) .
 \end{aligned} \tag{113}$$

Using Eq. (65) for the point spread function for a fixed dipole of orientation (α, β) , the integral over α in the numerator of Eq. (112) can be done explicitly, while the integral over β is involved because of the β -dependence of the normalization factor $\mathcal{N}(\beta)$ in Eq. (65). We now integrate each components in Eq. (65) over α , leaving the integration over β for later. Terms containing $p_{||}$ and p_{\perp} do not depend on α , hence only contribute with such terms in the weight-factor in Eq. (105), which are even powers of $\cos \alpha$ or $\sin \alpha$, resulting in the contribution

$$\sin^2 \beta I_{||} p_{||} \pi (|E_x|^2 \sin^2 \beta + |E_y|^2 \sin^2 \beta + 2|E_z|^2 \cos^2 \beta) , \tag{114}$$

respectively

$$\cos^2 \beta I_{\perp} p_{\perp} \pi (|E_x|^2 \sin^2 \beta + |E_y|^2 \sin^2 \beta + 2|E_z|^2 \cos^2 \beta) , \tag{115}$$

to the numerator in Eq. (112).

The term containing $\Delta p_{||}$ has the α -dependence

$$\cos(2\phi' - 2\alpha) = \cos(2\phi')(2\cos^2 \alpha - 1) + \sin(2\phi')(2\sin \alpha \cos \alpha) , \tag{116}$$

which means that the coefficient to $\cos(2\phi')$ also only contributes when multiplying even powers of $\cos \alpha$ or $\sin \alpha$, while the coefficient to $\sin(2\phi')$ contributes only when multiplying the term containing $\sin \alpha \cos \alpha$. Thus the term containing $\Delta p_{||}$ contributes

$$\sin^2 \beta I_{||} \Delta p_{||} \frac{\pi}{2} (\cos(2\phi') \sin^2 \beta (|E_x|^2 - |E_y|^2) + \sin(2\phi') \sin^2 \beta (E_x E_y^* + E_x^* E_y)) . \tag{117}$$

The term containing p_{\times} depends on α

$$\cos(\phi' - \alpha) = \cos \phi' \cos \alpha + \sin \phi' \sin \alpha , \tag{118}$$

where the coefficient to $\cos \phi'$ only contributes when multiplied by terms containing odd powers of $\cos \alpha$ and even powers of $\sin \alpha$, and similarly for the coefficient to $\sin \phi'$. This results in the contribution

$$\begin{aligned}
 \sin \beta \cos \beta I_{\times} p_{\times} \pi &\left(\cos \phi' \sin \beta \cos \beta (E_x E_z^* + E_x^* E_z) \right. \\
 &\left. \sin \phi' \sin \beta \cos \beta (E_y E_z^* + E_y^* E_z) \right)
 \end{aligned} \tag{119}$$

to the numerator in Eq. (112), where the first term vanishes due to Eq. (110).

To simplify notation, we write the remaining integrals over β as

$$I_{\beta}^{(n,m)} = \int_0^{\pi} \cos^n(\beta) \sin^m(\beta) \mathcal{N}(\beta) \sin \beta d\beta . \quad (120)$$

Collecting the bits and pieces, Eq. (112) becomes

$$\begin{aligned} p_{\text{pol}}(x', y') &= \frac{\int_{4\pi} d\Omega \langle |\vec{E} \cdot \vec{\mu}|^2 \rangle_T p(x', y' | \alpha, \beta)}{\int_{4\pi} d\Omega \langle |\vec{E} \cdot \vec{\mu}|^2 \rangle_T} \\ &= \frac{3}{4} (|E_x|^2 + |E_y|^2 + |E_z|^2)^{-1} \\ &\quad \left[I_{\parallel} p_{\parallel} \left(I_{\beta}^{(0,4)} (|E_x|^2 + |E_y|^2) + 2I_{\beta}^{(2,2)} |E_z|^2 \right) \right. \\ &\quad + I_{\perp} p_{\perp} \left(I_{\beta}^{(2,2)} |E_x|^2 + I_{\beta}^{(2,2)} |E_y|^2 + 2I_{\beta}^{(4,0)} |E_z|^2 \right) \\ &\quad + \frac{1}{2} I_{\parallel} \Delta p_{\parallel} I_{\beta}^{(0,4)} (\cos(2\phi') (|E_x|^2 - |E_y|^2) + \sin(2\phi') (E_x E_y^* + E_x^* E_y)) \\ &\quad \left. + I_{\times} p_{\times} I_{\beta}^{(2,2)} \sin(\phi') (E_y E_z^* + E_y^* E_z) \right] , \end{aligned} \quad (121)$$

which along with Eqs. (106–111) defines the PSF for a fluorescent bead. **Figure 3** in the paper shows a fit of this PSF using MLE to experimental data.

3.6 Comparison with 2D Gaussian

The 2D Gaussian interprets the shoulders of the distribution as a constant background (**Supplementary Fig. 1c,f,i**) when only the pixels in **Supplementary Fig. 1a** are included in the fit. The Gaussian therefore interprets a large part of the distribution as background, where it in fact is due to the slow power law decay of the experimental PSF, hence misinterprets $\sim 40\%$ of the source photons as background. For localization purposes, the steep parts of the distribution matters most, and this part of the distribution is well captured by the 2D Gaussian. Therefore, a limited amount of localization precision is lost by approximating the experimental PSF by a 2D Gaussian.

3.7 Analytical approximation

We write the various Bessel functions occurring in Eqs. (60)–(62) in their integral representation

$$J_{\nu}(z) = \frac{i^{-\nu}}{2\pi} \int_{-\pi}^{\pi} d\varphi \exp(i\nu\varphi + iz \cos \varphi) . \quad (123)$$

Thus the Bessel function of zeroth order, e.g., reads

$$J_0(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\varphi \exp(iz \cos \varphi) . \quad (124)$$

With the Bessel function written in this manner, each of the three expressions $\mathcal{I}[A]$, $\mathcal{I}[B]$, and $\mathcal{I}[C]$ is an integral over η' of an integral over φ . We interchange the order of integration and, for given value of ϕ , we approximate the integral over η' . To this end, we introduce the shorthand notation

$$\langle \dots \rangle^{(0)} \equiv \frac{\mathcal{I}[(E_p^{\parallel} - E_s^{\parallel}) \dots]}{\mathcal{I}[(E_p^{\parallel} - E_s^{\parallel})]} , \quad (125)$$

$$\langle \dots \rangle^{(1)} \equiv \frac{\mathcal{I}[E_p^{\perp} \dots]}{\mathcal{I}[E_p^{\perp}]} , \quad (126)$$

$$\langle \dots \rangle^{(2)} \equiv \frac{\mathcal{I}[(E_p^{\parallel} + E_s^{\parallel}) \dots]}{\mathcal{I}[(E_p^{\parallel} + E_s^{\parallel})]} , \quad (127)$$

hence have

$$\mathcal{I}[A] = \mathcal{I}[(E_p^{\parallel} - E_s^{\parallel}) J_0] = \mathcal{I}[(E_p^{\parallel} - E_s^{\parallel})] \langle J_0 \rangle^{(0)} \quad (128)$$

$$\mathcal{I}[B] = \mathcal{I}[(E_p^{\parallel} + E_s^{\parallel}) J_2] = \mathcal{I}[(E_p^{\parallel} + E_s^{\parallel})] \langle J_2 \rangle^{(2)} \quad (129)$$

$$\mathcal{I}[C] = \mathcal{I}[E_p^{\perp} J_1] = \mathcal{I}[E_p^{\perp}] \langle J_1 \rangle^{(1)} . \quad (130)$$

We use the *cumulant* expansion

$$\ln \langle e^X \rangle = \langle e^X - 1 \rangle_c = \sum_{n=1}^{\infty} \langle X^n \rangle_c / n! , \quad (131)$$

where $\langle 1 \rangle_c = 1$, $\langle \eta' \rangle_c = \langle \eta' \rangle = \mathcal{O}(1/M)$, $\langle \eta'^2 \rangle_c = \langle \eta'^2 \rangle - \langle \eta' \rangle^2 = \mathcal{O}(1/M^2)$, etc. [8, Section 3.12]. With this notation and $a \equiv k' \rho'$, the integral over η' is approximated as follows:

$$\begin{aligned} \langle J_0 \rangle^{(0)} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} d\varphi \langle \exp(i a \eta' \cos \varphi) \rangle^{(0)} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} d\varphi \exp(\langle \exp(i a \eta' \cos \varphi) \rangle_c^{(0)} - 1) \\ &\simeq \frac{1}{2\pi} \int_{-\pi}^{\pi} d\varphi \exp \left(i a \cos \varphi \langle \eta' \rangle^{(0)} - \frac{a^2}{2} \cos^2 \varphi \langle \eta'^2 \rangle_c^{(0)} \right) \\ &\simeq \frac{1}{2\pi} \int_{-\pi}^{\pi} d\varphi \exp(i a \cos \varphi \langle \eta' \rangle^{(0)}) \left(1 - \frac{a^2}{2} \cos^2 \varphi \langle \eta'^2 \rangle_c^{(0)} \right) \\ &= J_0(a \langle \eta' \rangle^{(0)}) - \langle \eta'^2 \rangle_c^{(0)} \frac{a^2}{4\pi} \int_{-\pi}^{\pi} d\varphi \cos^2 \varphi \exp(i a \cos \varphi \langle \eta' \rangle^{(0)}) \\ &= J_0(a \langle \eta' \rangle^{(0)}) + \frac{a^2}{2} \langle \eta'^2 \rangle_c^{(0)} \left(J_2(a \langle \eta' \rangle^{(0)}) - \frac{J_1(a \langle \eta' \rangle^{(0)})}{a \langle \eta' \rangle^{(0)}} \right) . \end{aligned} \quad (132)$$

Similarly for the integrals involving J_1 and J_2 ,

$$\langle J_1 \rangle^{(1)} = (a \langle \eta' \rangle^{(1)}) + \frac{a^2}{2} \langle \eta'^2 \rangle_c^{(1)} \left(-J_1(a \langle \eta' \rangle^{(1)}) - \frac{J_0(a \langle \eta' \rangle^{(1)})}{a \langle \eta' \rangle^{(1)}} + \frac{2J_1(a \langle \eta' \rangle^{(1)})}{(a \langle \eta' \rangle^{(1)})^2} \right) , \quad (133)$$

and

$$\langle J_2 \rangle^{(2)} = J_2(a\langle \eta' \rangle^{(2)}) + \frac{a^2}{2} \langle \eta'^2 \rangle_c^{(2)} \left(J_0(a\langle \eta' \rangle^{(2)}) - \frac{3J_1(a\langle \eta' \rangle^{(2)})}{a\langle \eta' \rangle^{(2)}} + \frac{6J_2(a\langle \eta' \rangle^{(2)})}{(a\langle \eta' \rangle^{(2)})^2} \right). \quad (134)$$

These expressions are used by evaluating $\langle \eta' \rangle^{(n)}$ and $\langle \eta'^2 \rangle_c^{(n)}$ once and for all for a given objective and wavelength of fluorophore emission^a, with the extra subtlety added that because each integral over η' approximated here is written naturally as a sum of an integral over sub-critical, respectively super-critical η -values, and the latter has both a real and an imaginary part, we apply the approximation just outlined to each of the three real integrals in this sum, e.g., in the program in **Supplementary Software 1**.

The result is an approximation to the PSF at $(x', y') = \rho'(\cos \phi', \sin \phi')$ given in Eq. (59), which is exact in its combination of elementary trigonometric functions of ϕ' and dipole angles α and β , while the ρ' -dependent coefficients to these trigonometric functions are exact at $\rho' = 0$, but for $\rho' > 0$ they are approximated by combinations of Bessel functions of ρ' , as given in Eqs. (128–130) and (132–134).

Note that $\langle \eta' \rangle^{(n)} k' \rho' = \mathcal{O}(1)$ where the PSF is small or vanishing at the edge of the diffraction limited spot, it describes. This means that our approximation is controlled by a small parameter which ceases to be small outside the spot. It is uncontrolled there, and the approximation breaks down gradually with increasing values of ρ' , as illustrated in **Fig. 1q–t**. They show that the approximation has more than 10% error in the corners of the pixel area that we analyze. However, even a substantial approximation error on the theoretical PSF does not matter when it occurs in the tail where the photon counts from the fluorescent probe drowns in the photon count from the background, which it does in our case. If there is little or no background, on the other hand, the Poisson statistics of the few photons that are counted per pixel in the tail, makes their number scatter with a standard deviation comparable to their expected value. This ensures that approximation errors in functions describing their expected value will not affect localization accuracy and precision.

This explicit, analytical approximate PSF of a fixed dipole is a very good approximation to the exact theoretical PSF, as illustrated in **Fig. 1**. We use the same approximations when evaluating the PSF for an isotropic distribution of dipoles, e.g. Eq. (122).

Computer code that evaluates this approximation and uses it for localization of a fixed dipole point source and determines the orientation of the dipole is given in **Supplementary Software 1**. Since this code evaluates $\langle \eta' \rangle^{(n)}$ and $\langle \eta'^2 \rangle_c^{(n)}$ once and for all, and after that only needs to call a combination of Bessel functions repeatedly in its fitting procedure for the PSF to data, this code runs quickly.

^aHere we have assumed that the refractive index of the medium of the fluorophore, the buffer, is identical to water's, hence enters the once-and-for-all evaluation once-and-for-all. If the refractive index of the buffer can vary, the once-and-for-all evaluation must be done anew for each new value of this index.

3.8 Covariance matrix for parameters of PSF

The same computer code gives error bars on the coordinates (x, y) and angles of orientation (α, β) of the dipole, and the covariance matrix for these variables. This covariance matrix is the one given in general form in Eq. (53). Its specific form for the present case is a 4×4 matrix, which it makes little sense to write down explicitly here, because it contains integrations over the image plane that do not simplify to a one-dimensional integral, as they did in the isotropic case because of the isotropy.

4 What we *really* see with an EMCCD

4.1 Expected input, stochastic output

For each photon detection event in a pixel, the EMCCD camera's on-chip electron-multiplication process outputs a random number of electrons S with exponential probability distribution $\alpha \exp(-\alpha S)$ [9, Supplementary Methods]. Here the rate parameter α , the inverse *gain*, is so small that we can treat S as a real number. Consequently, n independent photon detection events will output a random number of electrons, which follows the Gamma distribution with shape parameter n and rate parameter α , $\Gamma_n(S; \alpha) = \alpha^n S^{n-1} \exp(-\alpha S) / (n-1)!$, since this is what convolution of n identical exponential distributions give. It follows that the distribution q_i of signal values S_i generated in the i 'th pixel by Poisson distributed photons with expectation value E_i , is the sum over photon numbers,

$$\begin{aligned} q(S_i|E_i) &= P_0(E_i) \delta(S_i) + \sum_{n=1}^{\infty} P_n(E_i) \Gamma_n(S; \alpha) \\ &= \exp(-E_i) \delta(S_i) + \sum_{n=1}^{\infty} \alpha^n \frac{S_i^{n-1}}{(n-1)!} \exp(-\alpha S_i) \exp(-E_i) \frac{E_i^n}{n!} \\ &= \exp(-E_i) \delta(S_i) + \sqrt{\frac{\alpha E_i}{S_i}} \exp(-\alpha S_i - E_i) I_1(2\sqrt{\alpha E_i S_i}), \end{aligned} \quad (135)$$

where I_1 is the modified Bessel function of the first kind of order one. Further convolution with a Gaussian *readout-noise* distribution of width σ_{noise} gives the true distribution of the signal. This signal can take negative values because of the readout noise, hence is offset by a factory-set positive constant $S_{\text{offset}} \approx 1000$. By offsetting our theory for the signal by the same amount, we finally have a theory $\tilde{q}(S_i|E_i)$ for what one “sees” with an EMCCD, i.e., for the stochastic electronic output in each pixel for given expected number E_i of recorded photons. This is the theory that we fit to EMCCD images, using either the theoretical PSF in E_i and MLE for fitting (MLEwT), *or* using a 2D Gaussian-plus-constant in E_i and either unweighted least-squares fitting (GME), or weighted least-squares fitting (WLS), or MLE (MLEwG).

First, however, we determined the inverse gain α of the electron multiplication process and the readout noise distribution in a calibration of the camera [9, Sup-

plementary Methods]. We found $\alpha = 0.025$, $S_{\text{offset}} = 985$, and $\sigma_{\text{noise}} = 3.0$ for one camera and for another camera, respectively, 0.09, 400, and 2.5. Since σ_{noise} is much smaller than the average output per photon, $1/\alpha = 40$, we can, in theoretical calculations that are not sensitive to details right at $S = 0$, ignore the readout noise. Doing that, subtraction of S_{offset} from all recorded values S_i results in signal values which are distributed according to q above. Hereafter, we use S_i to denote these offset-subtracted signal values, as already done in Eq. (135).

4.2 Mistakes to avoid

Subtraction of the constant offset S_{offset} should not be confused with subtraction of a constant photonic background. A constant photonic background is only constant in its expectation value, while the actual photon count in any pixel is Poisson distributed, hence fluctuating, hence contributing to the localization error, hence not something one can handle correctly by a mere subtraction. We model constant photonic background with the parameter b^2 .

Conversely, if one *forgets* to treat the constant offset as truly constant, it will add to the value of b^2 found in a fit. Since all procedures considered here interpret b^2 as the expectation value of a Poisson-distributed photonic background, inclusion of S_{offset} in it will assign an equally larger variance to the signal in each pixel. This matters most for pixels with lower counts, so they are given relatively less weight than they should have.

If done, these mistakes will not affect fits with GME, since this method ignores weights. They do affect the precision estimate for the fit, since it is our estimate \hat{b}^2 for the photonic background that should be substituted for b^{*2} in Eq. (30) to obtain a precision estimate.

If one forgets to treat the constant offset as truly constant, it will add to the value for b^2 found in a fit and thus mistakenly increase the estimated error of MLEwG and MLEwT. It also causes reduced precision in WLS where WLS is stable. Where WLS is unstable, it stabilizes it, but with reduced precision, as described above.

4.3 Expected localization errors in the real world of electron statistics

For a given image $(S_i)_i$, estimation was done with q essentially as given in Eq. (135), except q was convoluted with a narrow Gaussian describing read-out noise before estimation was done. The errors on estimates are consequently found using q . They are found by a standard calculation in Information Theory sketched below. However, since $q(S|E)$ can be written as the sum in Eq. (135), its moments are simply related to those of the Poisson distribution in that sum, e.g.,

$$\langle S \rangle = E/\alpha \quad ; \quad \text{Var}(S) = 2 \text{Var}(E)/\alpha^2 . \quad (136)$$

Note here that the expectation values are proportional with the gain α^{-1} as constant of proportionality, but the variances are proportional *not* by a factor α^{-2} , but the

factor $2\alpha^{-2}$. This doubling of $\text{Var}(S)/\langle S \rangle^2$ relatively to $\text{Var}(E)/\langle E \rangle^2$ is caused by the extra stochastic element introduced by the photon multiplication process, referred to as *excess noise*.

The variances of estimators like $\hat{\mu}_x$, \hat{b}^2 and \hat{N} are more complicated to find, because we cannot use that the covariance matrix is the inverse information matrix, as proven in MLE in the limit of large data sets. We do have large expected photon counts $N \sim 10^4$, but not large enough to ensure large expected photon counts in every pixel in the tail of the PSF. For this reason—and because we have seen how low expected photon counts make WLS unstable—we calculate the covariance matrix explicitly below.

In the end it turns out that excess noise just doubles the variances compared to those calculated with photon statistics above. Which is the result we also would have arrived at by using the inverse information matrix for covariance matrix. Our path to this result is more tricky, but boils down to the fact that low photon count occur only in the tail of the PSF, where the counts are due to constant background, if there, plus the flat shoulders of the experimental PSF, which are modeled in the same manner as background. Consequently, their statistics do not contribute to the variance of $\hat{\mu}_x$ and $\hat{\mu}_y$. Below, we sketch the proof of this very useful result for the case of $\hat{\mu}_x$.

Fisher's Likelihood function now reads

$$\mathcal{L}(\theta|(S_i)_i) = \prod_i q(S_i|E_i(\theta)) \quad (137)$$

and its logarithm

$$\ell = \ln \mathcal{L} = \sum_i \ln q(S_i|E_i(\theta)) \quad (138)$$

For $S > 0$ we have

$$\ln q(S|E) = \frac{1}{2} \ln \alpha + \frac{1}{2} \ln E - \frac{1}{2} \ln S - \alpha S - E + \ln I_1(2\sqrt{\alpha ES}) \quad (139)$$

where I_1 is the modified Bessel function of the first kind and order one. Below, the same type of function of order zero occurs as well.

The MLE $\hat{\theta}((S_i)_i)$ is the value for θ that maximizes ℓ for given data $(S_i)_i$. It solves the stationarity equations

$$\sum_i E_{i,a} \frac{\partial \ln q}{\partial E}(S_i|E_i) = 0 \quad \text{for all } a, \quad (140)$$

where the θ -dependence of E_i and $E_{i,a}$ is understood, and

$$\frac{\partial \ln q}{\partial E}(S|E) = \sqrt{\frac{\alpha S}{E}} \frac{I_0(2\sqrt{\alpha ES})}{I_1(2\sqrt{\alpha ES})} - 1 \quad (141)$$

As above, we find the variance of the estimated parameter values by linearizing the stationarity equations w.r.t. $\Delta\theta$ at the true parameter values θ^* . As above, S_i

in the term containing $\Delta\theta_a$ is replaced by its expected value $S_i^* = \langle S_i \rangle = E_i^*/\alpha$, because fluctuations in S_i about its expected value, though not necessarily small compared to the expected value, occur only in sums that are small compared to other sums, which they add to. This results in

$$\sum_b I_{a,b} \Delta\theta_b = \sum_i E_{i,a}^* \frac{\partial \ln q}{\partial E}(S_i|E_i^*) , \quad (142)$$

where $I_{a,b}$ in general is called *Fisher's Information Matrix*, and for our specific case has the form

$$\begin{aligned} I_{a,b} &= - \sum_i E_{i,a}^* E_{i,b}^* \frac{\partial^2 \ln q}{\partial E^2}(S_i^*, E_i^*) - \sum_i E_{i,a,b}^* \frac{\partial \ln q}{\partial E}(S_i^*, E_i^*) \\ &= \sum_i E_{i,a}^* E_{i,b}^* \left(\frac{I_0^2(2E_i^*)}{I_1^2(2E_i^*)} - 1 \right) - \sum_i E_{i,a,b}^* \left(\frac{I_0(2E_i^*)}{I_1(2E_i^*)} - 1 \right) \\ &= \sum_i E_{i,a}^* E_{i,b}^* \left(\frac{I_0^2(2E_i^*)}{I_1^2(2E_i^*)} - 1 \right) - \mathcal{O}(1) , \end{aligned} \quad (143)$$

where we have used that $S_i^* = E_i^*/\alpha$, and in the last identity that the explicitly written term is of order N , which typically is 10^3 – 10^4 , hence renders the term of order 1 negligible.

Since Eq. (140) is solved by $(S_i, E_i)_i = (S_i^*, E_i^*)_i$, we can write

$$\begin{aligned} \sum_b I_{a,b} \Delta\theta_b &= \sum_i E_{i,a}^* \left(\frac{\partial \ln q}{\partial E}(S_i|E_i^*) - \frac{\partial \ln q}{\partial E}(S_i^*|E_i^*) \right) \\ &= \sum_i E_{i,a}^* \left(\sqrt{\frac{\alpha S_i}{E_i^*}} \frac{I_0(2\sqrt{\alpha E_i^* S_i})}{I_1(2\sqrt{\alpha E_i^* S_i})} - \frac{I_0(2E_i^*)}{I_1(2E_i^*)} \right) . \end{aligned} \quad (144)$$

Thus we have that the covariance matrix of the estimator $\hat{\theta}$ is

$$\begin{aligned} \langle \Delta\theta_a \Delta\theta_b \rangle &= \\ \sum_{a',b'} (I^{-1})_{a,a'} \sum_i E_{i,a'}^* E_{i,b'}^* \left\langle \left(\frac{\partial \ln q}{\partial E}(S_i|E_i^*) - \frac{\partial \ln q}{\partial E}(S_i^*|E_i^*) \right)^2 \right\rangle (I^{-1})_{b',b} . \end{aligned} \quad (145)$$

In MLE one proves that in the limit of large data sets this covariance matrix is just the inverse information matrix. We cannot just use this result because we may have low photon counts in some pixels, and must understand if this could possibly destabilize the estimator, as it does for WLS. So we proceed, and the next step consists in using the approximation

$$\frac{\partial \ln q}{\partial E}(S_i|E_i^*) - \frac{\partial \ln q}{\partial E}(S_i^*|E_i^*) = \frac{\partial^2 \ln q}{\partial E \partial S}(S_i^*|E_i^*) \Delta S_i \quad (146)$$

in the last result, with the definition $\Delta S_i = S_i - S_i^*$. Here

$$\frac{\partial^2 \ln q}{\partial E \partial S}(E/\alpha|E) = \frac{\alpha}{E} \left(\frac{I_0(2E)}{I_1(2E)} - E \left(\frac{I_0^2(2E)}{I_1^2(2E)} - 1 \right) \right) . \quad (147)$$

This approximation replaces the expectation value in Eq. (145) with a larger value for $E > 4.5$ and is better the larger E_i^* is. In the tail of the PSF the approximation is not good, but there we are saved by the fact that the terms approximated barely contribute to the sum over i in our case of interest, $\theta_a = \mu_x$, because the factor $E_{i,a'}^* E_{i,b'}^*$ is proportional to the PSF squared, hence vanishes there. Thus

$$\begin{aligned} \langle \Delta \theta_a \Delta \theta_b \rangle &= \\ &= \sum_{a',b'} (I^{-1})_{a,a'} \sum_i E_{i,a'}^* E_{i,b'}^* \left(\frac{\partial^2 \ln q}{\partial E \partial S}(S_i^*|E_i^*) \right)^2 \text{Var}(S_i) (I^{-1})_{b',b} \\ &= \sum_{a',b'} (I^{-1})_{a,a'} \sum_i \frac{E_{i,a'}^* E_{i,b'}^*}{2E_i^*} \left(2 \frac{I_0(2E_i^*)}{I_1(2E_i^*)} - 2E_i^* \left(\frac{I_0^2(2E_i^*)}{I_1^2(2E_i^*)} - 1 \right) \right)^2 (I^{-1})_{b',b} . \end{aligned} \quad (148)$$

Since

$$\left(2 \frac{I_0(2E)}{I_1(2E)} - 2E \left(\frac{I_0^2(2E)}{I_1^2(2E)} - 1 \right) \right)^2 < 1 < 2E \left(\frac{I_0^2(2E)}{I_1^2(2E)} - 1 \right) \quad (149)$$

for all values of E , we obtain a larger right-hand-side in Eq. (148) by substituting 1 for $\left(2 \frac{I_0(2E)}{I_1(2E)} - 2E \left(\frac{I_0^2(2E)}{I_1^2(2E)} - 1 \right) \right)^2$ in the numerator there, and by substituting 1 for $2E \left(\frac{I_0^2(2E)}{I_1^2(2E)} - 1 \right)$ in $I_{a,b}$ in the denominator there. In this manner we arrive at

$$\text{Var}(\theta_a) \leq \left(\sum_i \frac{E_{i,a}^{*2}}{2E_i^*} \right)^{-1} , \quad (150)$$

where the equality holds to a very good approximation. This result is twice the result obtained with photon statistics—compare with Eqs. (52) and (53)—the factor two being due to the extra stochastic element of the electron multiplication process, known as “excess noise.” For E_i as in Eq. (12) and p as in Eq. (3), this simplifies to *twice* the result from photon statistics given in Eq. (54).

It is with this result—i.e., *twice* the result for photon statistics given in Eq. (54)—that we estimate the localization error for MLEwG. **Supplementary Fig. 2** shows that this is in full agreement with the localization error found “experimentally” in computer experiments with known truth. So the approximations done in its derivation introduced only indiscernible errors. **Figure 4** shows that the formula agrees fully with the localization error found experimentally in repeated measurements of the same distance. Both figures show that this error is distinctly smaller than localization errors of other methods. So use of MLE increases precision at no extra cost, and its precision is reliably captured by twice Eq. (54).

5 Effects on precision from background and finite pixel width

Formulas dealing exactly with a constant background and finite pixel sizes have already been given for the estimators, MLEwT, MLEwG, GME, and WLS. Using photon statistics, we here test the formulas on simulated data from the theoretical PSF for various background levels and different pixel sizes. This simulation is analogous to the one done in [1], except we simulate in 2D and our simulated photon counts are distributed not as a Gaussian plus background, but according to the theoretical PSF plus background.

For each pixel, we simulate a Poisson random variable n_i with expectation value $\langle n_i \rangle = Np_i + b^2$. We generate a number of data sets for each set of parameters, then we analyze each data set with each of the estimators MLEwT, MLEwG, and GME. (WLS was not considered because its precision depends on how we choose to patch up WLS to avoid large loss of precision due to arbitrarily large experimental weights of some pixels and because it is suboptimal no matter how we do that.) From each set of center position estimates, the variance of each scheme is estimated and compared to the predictions of theory.

Supplementary Fig. 2a shows the results from a simulation of 3,000 images for different background levels b^{2b} , where each image consists of a 32×32 pixel array of $(25 \text{ nm})^2$ pixels. By binning of the pixels, the same data set was used to imitate a simulation of an 8×8 pixel array of $(100 \text{ nm})^2$ pixels, producing the results shown in **Supplementary Fig. 2c**.

The theoretical formulas for the variances of GME, MLEwG, and MLEwT agree nicely with the computer-experimental variances over the full range of backgrounds for $(25 \text{ nm})^2$ pixels, while the formula in [1, Eq. (17)] falls below the information limit shown as a fully drawn black curve—i.e., the precision it states that one has achieved, is actually unattainable.

The precision stated by the formula for estimation with MLEwG is seen to fall slightly above the information limit—i.e., negligible information is lost by using MLEwG instead of MLEwT. This is born out by variances found in the computer-experiment falling practically on top of each other for MLEwG and MLEwT.

For vanishing true background—i.e., for $b^2/a^2 = 0$ in **Supplementary Fig. 2**, the fit of a 2D Gaussian+constant still returns a positive constant, because it treats the shoulders of the PSF as background. This is the reason why the formula in Eq. (57) and the variance of MLEwG differ for $b^2/a^2 = 0$ in **Supplementary Fig. 2**.

For large background levels, the variance of GME is seen to approach the variance

^bThis b^2 parameterizes real photonic background to which the theoretical PSF adds, and is not the same as the constant estimated by fitting a 2D Gaussian-plus-constant to data, since such a fit treats the counts in the shoulders of the PSF as additional background, and estimates *the sum* of this false background and the true background. We denote the latter background by b_G^2 with G for Gauss, when we wish to distinguish the two in our notation.

of MLEwT asymptotically, so the relative advantage of MLE over GME decreases in this regime. The interpolation formula given in Ref. [1, Eq. (17)] keeps exaggerating the precision also in this regime, though relatively less. This is because it was conceived as an interpolation formula, correct only at the extremes of no background and infinitely large background, as made perfectly clear by the authors in their paper.

Supplementary Fig. 2c tells essentially the same story, except we see that the theoretical formulas for the variances of GME and MLEwG fall slightly under the simulation results, and since the simulation results for MLEwG and MLEwT are nearly identical, the formula for MLEwG falls slightly *below* the information limit. These slight underestimates are due to the large pixel size, $(100\text{ nm})^2$. Our formulas are valid only to order $a^2/(12\sigma^2)$ and higher-order terms would have to be included in order to correct this $\sim 1\%$ error on our error estimates.

Comparing **Supplementary Fig. 2a** and **Supplementary Fig. 2c**, one notes that one loses little precision by the large increase in pixel size.

Supplementary Fig. 2b,d compares the total number of photons N_G obtained from MLEwG, respectively GME, in the simulations described above to the expected number of photons N . We observe that $N_G/N \sim 0.6$ largely independent of background level and pixel size. Both MLEwG and GME therefore underestimate the number of photons in an image by interpreting the shoulders of the theoretical PSF as a constant background. To count the numbers of photons in an image, one should use MLEwT.

References

- [1] Russell E. Thompson, Daniel R. Larson, and Watt W. Webb. Precise nanometer localization analysis for individual fluorescent probes. *Biophys J*, 82(5):2775–2783, May 2002.
- [2] C. R. Rao. *Linear Statistical Inference and Its Applications*. John Wiley & Sons, Inc., New York, 1965.
- [3] Rainer Heintzmann. *Resolution Enhancement of Biological Light Microscopy Data*. PhD thesis, University of Heidelberg, 1999. http://www.kip.uni-heidelberg.de/AG_Cremer/rheintz/Dissertation.
- [4] Jörg Enderlein, Erdal Toprak, and Paul R. Selvin. Polarization effect on position accuracy of fluorophore localization. *Optics Express*, 14:8111–8120, 2006.
- [5] M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, New York, 1999.
- [6] L. Stirling Churchman, Henrik Flyvbjerg, and James A. Spudich. A non-gaussian distribution quantifies distances measured with fluorescence localization techniques. *Biophys J*, 90(2):668–671, Jan 2006.
- [7] Daniel Axelrod, Thomas P. Burghardt, and Nancy L. Thompson. Total internal reflection fluorescence. *Ann. Rev. Biophys. Bioeng.*, 13:247–268, 1984.
- [8] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 1. Griffin, London, 3rd edition, 1969.
- [9] Maximilian H. Ulbrich and Ehud Y. Isacoff. Subunit counting in membrane-bound proteins. *Nat Methods*, 4(4):319–321, Apr 2007.