Supplemental Material for "Optimal Point Spread Function Design for 3D Imaging"

Yoav Shechtman[1], Steffen J. Sahl[1], Adam S. Backer[1,2], W. E. Moerner[1]
[1]Department of Chemistry, Stanford University, 375 North-South Mall, Stanford, California 94305, USA
[2]Institute of Computational and Mathematical Engineering, 475 Via Ortega, Stanford, California 94305, USA
E-mail address: yoavsh@stanford.edu

PSF optimization process

In both examples considered in this work, the objective function being minimized is the mean $\sqrt{\text{CRLB}}$ in *x*, *y* and *z*, over a predetermined *z* range Z (3 μm as in the text, and 6 μm described below). In practice, only a subset of equally spaced points within this range is selected, at 250 nm increments and not finer, because diffraction prevents the function from changing too rapidly anyway. The optimization is performed over a set of Zernike polynomials [1] (the first 55 in our case), so that the sought solution is a coefficient vector $c \in R^N$ with $N = 55$. The mathematical optimization problem, solved over 64x64 pixels using Matlab's *fmincon* function ('interior point' method), is therefore:

$$\min_{\mathbf{c}} \sum_{j=\hat{x},\hat{y},\hat{z}} \sum_{z \in Z} \sqrt{\frac{1}{I_{jj}(\mathbf{c},z)}} \tag{s1}$$

where, assuming additive Poisson noise and a constant background of β, the Fisher information matrix for a point source along the optical axis is given by [2]:

$$I(\mathbf{c}; 0,0,z, N_{ph}, \beta) = \sum_{k=1}^{N_p} \frac{1}{\mu_{\mathbf{c},z}(k)+\beta} \left(\frac{\partial \mu_{\mathbf{c},z}(k)}{\partial \boldsymbol{\theta}}\right)^T \left(\frac{\partial \mu_{\mathbf{c},z}(k)}{\partial \boldsymbol{\theta}}\right). \tag{s2}$$

Here, $\boldsymbol{\theta} = (x, y, z)$ is the 3D position of the emitter, namely, in the optimization routine we practically only consider the 3×3 Fisher submatrix corresponding to (x,y,z), since the result is insensitive to this fact (Physically, this corresponds to the fact that the estimated position is not very sensitive to a change in the number of signal or background photons). We denote by $\mu_{\mathbf{c},z}$ the model of the detected PSF (detailed below) for an emitter at z, including the total number of signal photons per frame, magnification and pixelation, for a PSF produced by a Fourier-plane mask $P(x', y')$ defined in the pupil plane by:

$$P(x', y') = \text{circ}(r/R) \cdot \exp(iD_{zer} \cdot \mathbf{c}) \tag{s3}$$

where $r = \sqrt{x'^2 + y'^2}$, R is the radius of the pupil plane, $\text{circ}(\eta) = \begin{cases} 1, & \eta < 1 \\ 0, & \eta \geq 1 \end{cases}$, and D_{zer} is the linear operator transforming the vector of Zernike coefficients to the 2D phase pattern to be projected on the SLM. The SLM is discretized to a 256×256 grid, so that in our implementation $D_{zer} \in R^{256^2 \times 55}$, where each column is a vector stacked 2D image of the corresponding Zernike polynomial.

The image formation model, using u,v for the real space coordinates at the detector plane, is derived from the Fresnel approximation [3], and the intensity at the detector plane I_p is given by:

$$I_p(u, v) \propto \left| \Im \left\{ P(x', y') \cdot \exp\left(-i\frac{2\pi}{2\lambda f}\left(\frac{\alpha M^2 z}{f}\right)(x'^2 + y'^2) \right) \cdot \exp\left(i\frac{2\pi M}{2\lambda f}(x + y) \right) \right\} \right|^2 \tag{s4}$$

In the expression above, the first exponential term accounts for the quadratic phase induced in the pupil plane when an emitter is a distance z from the objective's focal plane. The second exponential term is a linear phase-ramp used to account for the lateral displacement of the emitter within the field of view. f = 15 cm denotes the focal length of the lenses used in the 4*f*-system, λ = 560 nm is the peak emission wavelength and M = 100 is the magnification. The parameter α = 0.81 is used to

augment the Fresnel propagation model to account for defocus in a high-index immersion medium (n = 1.518), and better approximate the true phase of light propagating through a non-paraxial optical system. We used a pupil plane radius of R = 2.15 mm. Note that the coordinate scaling is such that camera coordinates (u, v) correspond to Fourier plane frequencies $(u/\lambda f, v/\lambda f)$. Additionally, the actual measured image is pixelated, due to the finite pixel size of the detector. For a given pixel size of $\Delta \times \Delta$ ($\Delta = 16\mu m$ was used in our simulations), this gives:

$$\mu(k) = C_{ph} \cdot \int_{u_k-\Delta/2}^{u_k+\Delta/2} \int_{v_k+\Delta/2}^{v_k+\Delta/2} I_p(u, v) dv du \qquad (s5)$$

where μ_k is the value of the measured PSF in pixel k, centered around (u_k, v_k), and C_{ph} is a constant multiplier, normalizing for a total number of N_{ph} signal photons, when summing over all pixels. The parameters used in our simulation matched our imaging setup under typical (noisy) imaging conditions.

Designing a PSF for a large *z*-range (~6 µm)

We illustrate and validate our method by using it to optimize a PSF for a further challenge: 6 µm imaging depth. This is done by minimizing the 3D CRLB with the same parameters as for the 3 µm range (which yielded the saddle-point mask), but this time over a 6 µm *z*-range. The resulting mask, termed the Cat Mask (due to its feline appearance), the PSF, and experimental precision results are shown in Fig. S1. A nanosphere was scanned over a 6 µm range for calibration, and imaged for 500 frames at defined *z* positions, with mean detected signal photon count of ~3300, and $\beta \approx 43$. As before, the process was repeated for the DH-PSF mask, and repeated for four nanospheres.

The Cat Mask (Fig. S1a) focuses energy in the image (detector) plane into what can loosely be called five-lobe patterns, tracing out a "stick-figure" further from the objective ($z < 0$), concentrating the light in focus, and then extending outward again into five lobes arranged in a pentagon closer to the objective ($z > 0$) (Fig. S1b). Figure S1c shows experimental precision measurements. The DH-PSF, designed for a ~3 µm z-range, is clearly not applicable for this much larger range. The Cat PSF, on the other hand, exhibits relatively constant precision over a large ~5 µm range under high background conditions. The large applicable range does come at the cost of somewhat inferior precision compared to the DH PSF around the middle ~2 µm (Fig. S1c insets) and a larger pixel footprint (Fig. S1b).
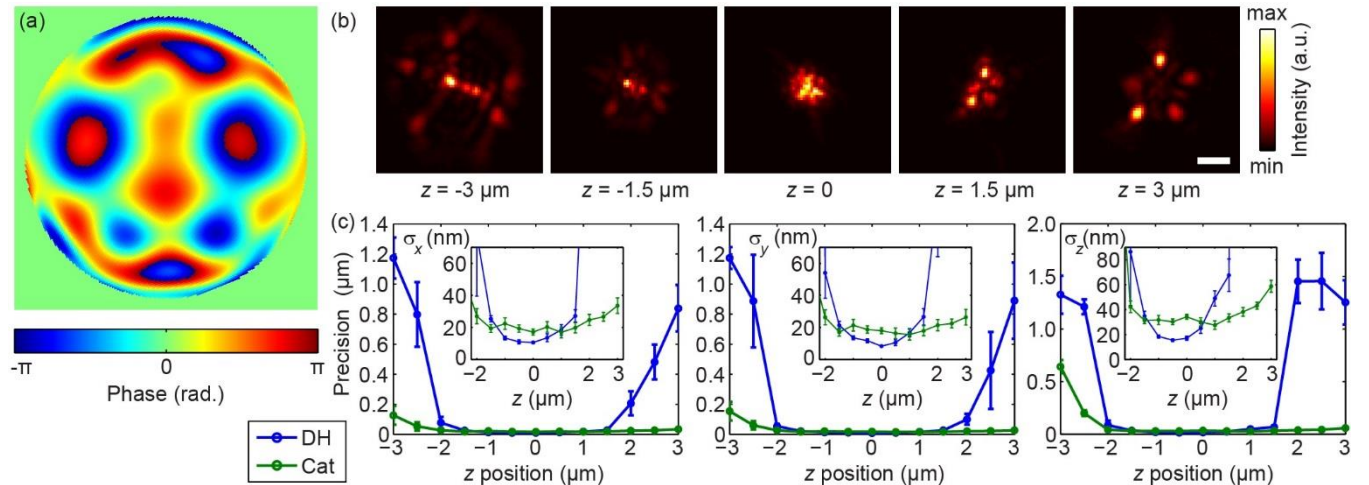

FIG S1: (a) Optimized Cat Mask for high background, low signal 3D precision over a 6 µm depth range (rotated). (b) Experimental realizations of the Cat Mask PSF for various *z* positions (stated). Scale bar: 2 µm (in sample space). Images were re-scaled to min/max intensity on an individual basis. (c) Experimental measurements of statistical localization precisions as the standard deviations σ_x, σ_y, σ_z of localization outcomes from 500 camera frames. Bars show standard deviations derived from $n = 4$ independent experiments. Insets show magnified views of precision data over a 5 µm range.

Experimental demonstration – tracking a moving emitter

In order to demonstrate the applicability of the engineered PSF to a large *z*-range (~5 µm) tracking application, we demonstrate tracking of a 100 nm fluorescent bead undergoing random Brownian motion in a drop of water with ~1% polyvinyl alcohol for the duration of 10.5 seconds, using the cat-mask. In each 15ms frame of the 10.5 sec movie (**Supplemental Movie 1**), the emitter was first localized coarsely by finding the center of mass (COM) of pixel intensities in the frame, and then a 64x64 pixel region around the COM was used as the input to the localization procedure. Figure S2 shows the measured three-dimensional trajectory of the bead during the 700 analyzed frames constituting the 10.5 second movie.
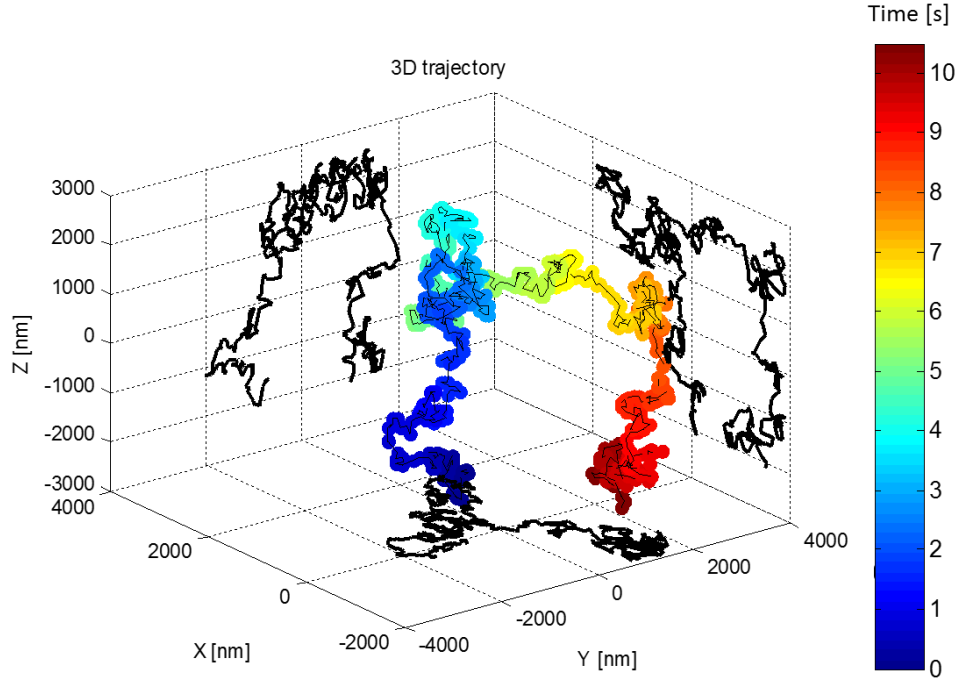

FIG S2: Three-dimensional trajectory of a 100 nm fluorescent bead in water (see **supplemental movie 1**). The trajectory is color coded by time, and projections onto the XY, XZ and YZ planes are plotted in black.

Localization method

Emitter localization in this work was performed using maximum-likelihood estimation (MLE). Given an imaging model and a noise model, it is possible to calculate the *likelihood* of a measurement given a set of underlying parameters [4,5]. The set of parameters we seek in each frame is the five element vector (x, y, z, N_{ph}, β), comprising the 3D position of the emitter, the total signal photons, and the mean background level, respectively. The estimated parameters are found by seeking the maximum of the likelihood function. In practice, the negative log-likelihood (LL) is minimized, assuming each pixel of the image is corrupted by Poisson noise. The appropriate objective function is thus given by [5]:

$$LL(\boldsymbol{\mu}; \boldsymbol{s}) = \sum_{k=1}^{N_p} \mu(k) - s(k) \ln[\mu(k)] + C \qquad (s6)$$

where $s(k)$ is the number of photons detected in pixel k of the measured image, $\mu(k)$ is the expected number of photons in pixel k, computed using the parameters we wish to optimize, and C is an additive term which does not affect the minimization. The minimization of the LL is performed in this work by Matlab's *fmincon* function.

Image formation model based on local phase retrieval

The MLE requires a continuous image formation model, namely, a method to calculate the PSF for any given emitter position (x,y,z) within the relevant range. In principle, one can rely on computational propagation, by plugging in the phase-mask that is put on the SLM into the imaging model (Eq. s5). However, in order to create a more realistic imaging model, in which the system's optical aberrations are accounted for, it is better to use the calibration measurements of the actual experimental PSF for this purpose. Thus, we create such an image formation model using Eq. s5, where the phase mask $P(x', y')$ is produced from the measured 50 nm dictionary, using a local phase-retrieval process similar to the one detailed in [6].

The local phase retrieval process calculates an effective phase mask $P_z(x', y')$ for any given emitter position z as follows: A set of 7 consecutive dictionary images in the range $z' = \pm150$ nm is used in order to find *a single* phase mask that would produce these dictionary images. This is done using a multi-plane Fienup style phase-retrieval method [6--8], which iterates between the measured real space defocused images and the calculated effective phase mask.

The phase retrieval procedure is summarized by the following algorithmic steps:

Given a set of *n* equally spaced dictionary images, $I_i(u, v)$ $i = 1 \dots n$ around an estimated position *z'*:

0. Initialization: Guess an initial pupil-plane pattern (intensity & phase) $P_0(x', y')$.
1. Given current pupil-plane pattern, $P_k(x', y')$, use the image formation model (Eq. s4) to calculate the resulting complex field distributions on the EMCCD plane corresponding to the *n* equally spaced positions in the range z = ±150 nm: $f_i(u, v)$ $i = 1 \dots n$.
2. Replace the magnitudes of the calculated field distributions by the square-root of the measured dictionary images at the corresponding z positions: $f'_i(u, v) = \sqrt{I_i(u, v)} \cdot \exp(\arg\{f_i(u, v)\})$
3. For each z position, calculate a corresponding current mask estimate by inverse Fourier-transforming and dividing by the quadratic phase factor (see Eq. s4) that corresponds to this z position:

$$P_k^i(x', y') = \Im^{-1}\{f'_i(u, v)\}/\exp\left(-i\frac{2\pi}{2\lambda f}\left(\frac{\alpha M^2 z}{f}\right)(x'^2 + y'^2)\right)$$

4. Calculate current pupil plane estimate by averaging the results of step 3: $P_{k+1}(x', y') = \frac{1}{n}\sum_{i=1}^{n} P_k^i(x', y')$
5. Go to 1. Stop after *MaxIt* iterations, or if the averaged l_2 error between calculated intensities and measured intensities is below a specified threshold *thresh*.

In our implementation we used $n = 7$, *MaxIt* = 1000 and *thresh* = 0.04.

PSFs used in numerical calculation

Three PSFs were used in numerical CRLB calculations:

1. The saddle-point PSF (SP-PSF): Obtained by our optimization procedure, as described in the main text, and displayed in Fig. 2.
2. The Double-helix PSF (DH-PSF): Taken from [9].
3. Astigmatic PSF: An astigmatic PSF was generated by creating a pupil-plane phase mask consisting of a single Zernike astigmatism mode. The magnitude of the Zernike coefficient was selected such that using our simulated system parameters, the mean $\sqrt{\text{CRLB}}$ over the tested 3μm range was minimal. This yields a separation of 560nm between the two planes of maximum focusing in the two orthogonal directions, which is a reasonable experimental value [10].

Experimental details

All calibration and precision measurements were performed using fluorescent nanospheres on an inverted microscope system (Olympus IX-71, objective lens 100×/NA 1.4 oil immersion UPlanSApo, Olympus) with custom widefield laser excitation and equipped with an EMCCD image sensor (iXon+, DU 897-E, Andor).

Phase masks were loaded onto an SLM (XY Phase Series, Boulder Nonlinear) placed in the Fourier plane as described in [11]. A sample of sparse immobilized emitters (~1 per 50×50 µm2) was created by spincoating 100 nm diameter fluorescent spheres (FluoSpheres Orange (540/560), Molecular Probes) diluted in ~2% polyvinyl alcohol (Carl Roth Chemicals) in water onto a pre-cleaned standard microscopy cover glass (no. 1, Fisherfinest Premium).

The sample was index-matched at the glass-nanosphere interface by coating the top of the microscope coverslip with a droplet of immersion oil. Sample fluorescence was excited using the 514-nm line of an argon ion laser (Innova 90, Coherent) set at ~1 W/cm^2 (at the sample plane, ~circularly polarized), filtered by a dichroic (Z514RDC, Chroma) and a band pass filter (578/105, Semrock).

A library of experimental PSF samples was created by stepping with an axial objective positioner (C-Focus, Mad City Labs) in Δz = 50 nm steps. In order to take localization precision measurements at various z, 500 frames were recorded at each z position using frame times of 15 ms and an EM gain setting of 200. A controllable background level of β at the sample was produced by illuminating the sample with white light. A sheet of soft lens paper placed between the objective and the light source was used as a diffuser. Since our SLM was only capable of modulating one polarization component of incident light, detected fluorescence was filtered using an appropriately aligned polarizing beamsplitter placed in the imaging pathway. Schemes are already available for recovering the loss that is introduced here, for example a pyramid design to capture and modulate both polarizations [11], or fixed transmissive phase masks. For the present study, this loss is not essential to the conclusions and the reported detected photons are for one polarization.

To correct for aberrations present in our optical setup, we numerically add an aberration correction phase pattern to each mask. The aberration correction pattern is calculated using the same multi-plane phase retrieval procedure described above, using a clear aperture (i.e. constant phase pattern on the SLM) and measurements of a single fluorescent bead obtained over a z-range of 3 µm, in 50 nm increments.

Experimental calibration measurements of the cat PSF compared with theoretical calculation over the z=[-3,3] µm range are shown in Fig. S3.
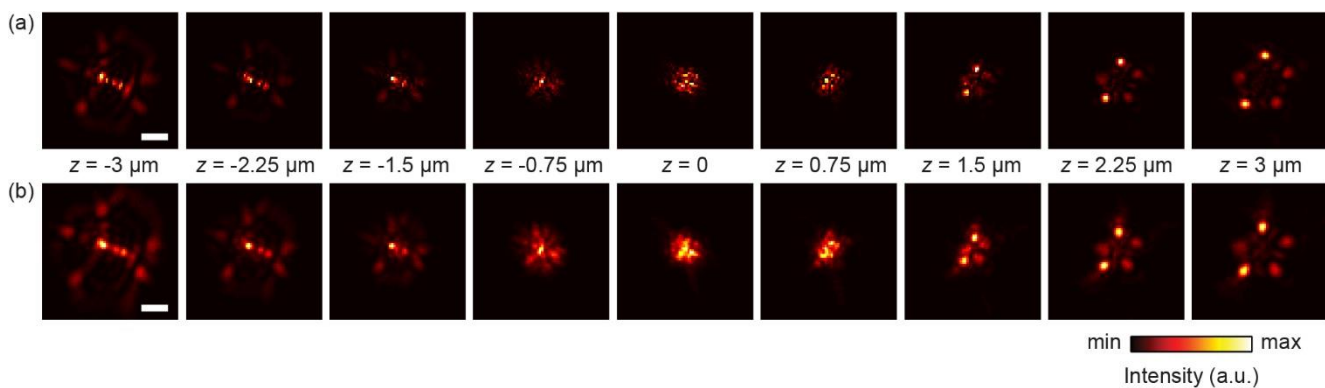

FIG S3: (a) Calculated vs. (b) experimentally realized Cat PSF. Scale bar: 2µm (in sample space).

Experimental precision vs. calculated CRLB

While the experimental behavior of the tested phase masks is qualitatively similar to the theoretical CRLB calculations, there are some small discrepancies between theory and experiment. First, the noise model assumed in the CRLB is purely Poisson, while in practice there is additional noise due to the gain of the EMCCD and detector readout. Second, the model assumes a constant background level in all pixels, as well as a constant number of signal photons per frame, assumptions which are inevitably violated to some extent in an actual experiment.

However, the most crucial factor contributing to the discrepancy between theoretical CRLB and experimental precision is image-model mismatch. In other words, the experimentally produced PSF differs from the engineered one. This is due to additional aberrations in the optical system that are unaccounted for in the imaging model. For example, the scalar model used for electromagnetic wave propagation (Eq. s4) does not account for the true vector behavior of light, as well as for the fact that, using an SLM implementation in the $4f$ system, we only measure one polarization. In addition, fluorescence emission is not monochromatic but spans a finite window of wavelengths.

The imaging model mismatch is clearly visible both for the SP PSF, when comparing Fig. 2b with Fig. 3a, and in Fig. S3, comparing the calculated cat PSF with experimental measurements. However, it is possible to calculate the CRLB for the *measured* PSF. This is done simply by using the locally phase-retrieved phase-masks in different z positions in the CRLB calculation. The result of this calculation vs. experimental precision measurements for both SP and DH PSFs is shown in Fig S4 (right). For comparison, Fig. S4 (left) shows the CRLB calculation from the calculated model vs. the same experimental measurements for both masks. The excellent agreement between calculated CRLB based on the phase-retrieved experimental PSF and measured precision in Fig. S4 (right), strengthens the following important conclusion, which supports similar findings [5,12]: G*iven an accurate imaging model, **the MLE can approach the CRLB**.

This fact validates the approach of PSF design by CRLB minimization, in that the CRLB is actually indicative of physically attainable precision.
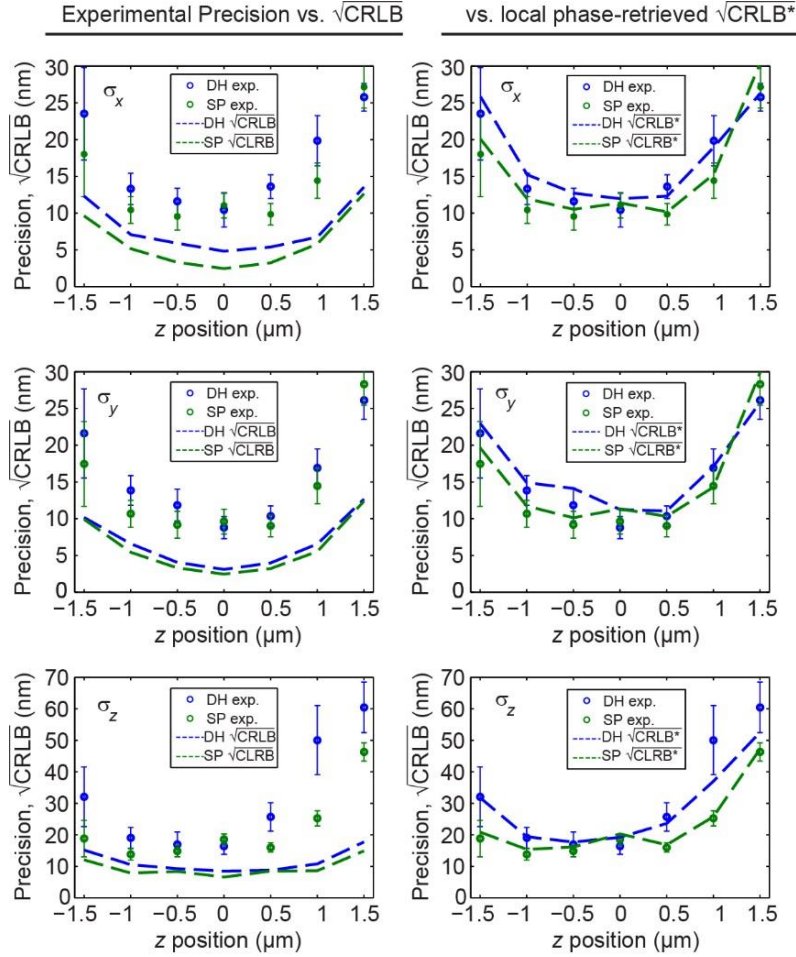
FIG S4: Experimental localization precisions vs. CRLB and phase-retrieved local CRLB* (saddle-point (SP) vs. double-helix (DH) mask). The left plots show the experimentally obtained precisions as well as the theoretically calculated CRLB, using the ideal masks (DH and SP), vs. z position. The right plots show the same experimental precisions, but this time compared to the local phase retrieved masks, obtained from the experimentally measured PSF. The excellent agreement on the right plots suggests that given the correct imaging model, CRLB is a very good predictor for attainable precision using MLE.

Beyond CRLB

As noted in the previous section, when using an experimentally derived imaging model, the experimental precision follows the CRLB quite well over a significant part of the tested z-range. However, this behavior breaks down at the extremes of the z-range. This happens when the signal-to-noise ratio (SNR) is very low, due to diffraction of the signal pattern far away from the focal plane, as can be clearly seen in Fig. S5.

In such a very low SNR regime, the MLE becomes increasingly susceptible to degeneracies in the PSF (i.e. very different parameter values possess similar likelihood). This occurs when the shape of the PSF in one z position is similar to the shape in another z position, e.g. when the double helix PSF completes a full rotation, after ~3 μm. Such a degeneracy will cause the estimator to be biased, and a sequence of localization measurements will yield a highly non-Gaussian distribution. Note that the CRLB is completely blind to such degeneracies, because the CRLB is computed *locally*, around a point in parameter space. Therefore, in such cases, the CRLB is not a good estimate for precision. This suggests that as a future research direction, for very low SNR, one might modify the optimization objective function from relying purely on CLRB to include additional non-local terms.
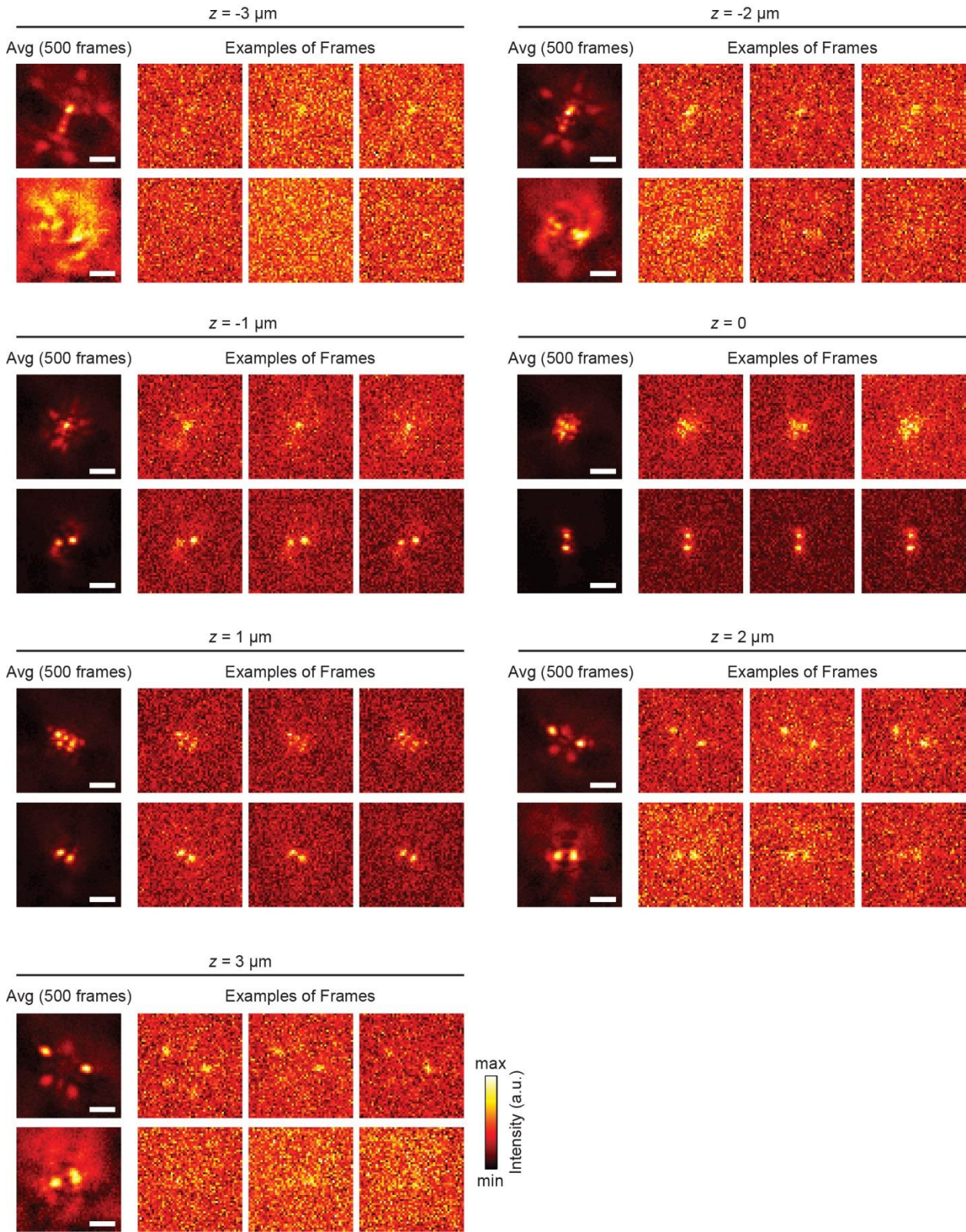
FIG S5: Localization in presence of background. Experimental cat PSF (upper row) vs. double-helix (DH) PSF (lower row) images at indicated z positions. Shown are the averages of all 500 frames (denoised) and three individual frames (frames 240, 250, 260 from the image sequence). Scale bar: 2μm (in sample space).

[1] M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, Cambridge, 1999).

[2] R. J. Ober, S. Ram, and E. S. Ward, Biophys. J. **86**, 1185 (2004).

[3] J. W. Goodman, *Introduction to Fourier Optics* (Roberts & Company Publishers, Greenwood Village, CO, 2005).

[4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall PTR, Englewood Cliffs, NJ, 1993).

[5] A. V. Abraham, S. Ram, J. Chao, E. S. Ward, and R. J. Ober, Opt. Express **17**, 23352 (2009).

[6] S. Quirin, S. R. P. Pavani, and R. Piestun, Proc. Natl. Acad. Sci. U. S. A. **109**, 675 (2012).

[7] J. R. Fienup, Opt. Lett. **3**, 27 (1978).

[8] B. Hanser, M. Gustafsson, and D. Agard, J. Microsc. **216**, 32 (2004).

[9] S. R. P. Pavani, M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. E. Moerner, Proc. Natl. Acad. Sci. U. S. A. **106**, 2995 (2009).

[10] M. J. Mlodzianoski, M. F. Juette, G. L. Beane, and J. Bewersdorf, Opt. Express **17**, 8264 (2009).

[11] M. P. Backlund, M. D. Lew, A. S. Backer, S. J. Sahl, G. Grover, A. Agrawal, R. Piestun, and W. E. Moerner, Proc. Natl. Acad. Sci. U. S. A. **109**, 19087 (2012).

[12] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, Nat. Methods **7**, 373 (2010).