Position and orientation estimation of fixed dipole emitters using an effective Hermite point spread function model

Sjoerd Stallinga* and Bernd Rieger

Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

*s.stallinga@tudelft.nl

Abstract: We introduce a method for determining the position and orientation of fixed dipole emitters based on a combination of polarimetry and spot shape detection. A key element is an effective Point Spread Function model based on Hermite functions. The model offers a good description of the shape variations with dipole orientation and polarization detection channel, and provides computational advantages over the exact vectorial description of dipole image formation. The realized localization uncertainty is comparable to the free dipole case in which spots are rotationally symmetric and can be well modeled with a Gaussian. This result holds for all dipole orientations, for all practical signal levels, and for defocus values within the depth of focus, implying that the massive localization bias for defocused emitters with tilted dipole axis found with Gaussian spot fitting is eliminated.

References and links

1. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution," Science **313**, 1643–1645 (2006).
2. K. A. Lidke, B. Rieger, T. M. Jovin, and R. Heintzmann, "Superresolution by localization of quantum dots using blinking statistics," Opt. Express **13**, 7052–7062 (2005).
3. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," Nat. Methods **3**, 793–795 (2006).
4. J. Fölling, M. Bossi, H. Bock, R. Medda, C. A. Wurm, B. Hein, S. Jakobs, C. Eggeling, and S. W Hell, "Fluorescence nanoscopy by ground-state depletion and single-molecule return," Nat. Methods **5**, 943–945 (2008).
5. M. Heilemann, S. van de Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer, "Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes," Angew. Chem., Int. Ed. Engl. **47** 6172–6176 (2008).
6. S. Stallinga and B. Rieger, "Accuracy of the Gaussian Point Spread Function model in 2D localization microscopy," Opt. Express **18**, 24461-24476 (2010).
7. J. Engelhardt, J. Keller, P. Hoyer, M. Reuss, T. Staudt, and S. W. Hell, "Molecular orientation affects localization accuracy in superresolution far-field fluorescence microscopy," Nano Lett. **11**, 209–213 (2010).
8. J. Enderlein, E. Toprak, and P. R. Selvin, "Polarization effect on position accuracy of fluorophore localization," Opt. Express **14**, 8111 (2006).
9. T. J. Gould, M. S. Gunewardene, M. V. Gudheti, V. V. Verkhusha, S.-R. Yin, J. A. Gosse, and S. T. Hess, "Nanoscale imaging of positions and anisotropies," Nat. Methods **5**, 1027–1031, 2008.
10. S. R. P. Pavani, J. G. DeLuca, and R. Piestun, "Polarization sensitive, three-dimensional, single-molecule imaging of cells with a double-helix system," Opt. Express **17**, 19644-19655 (2009).

11. M. R. Foreman, C. M. Romero, and P. Török, "Determination of the three-dimensional orientation of single molecules," Opt. Lett. **33**, 1020–1022 (2008).
12. M. R. Foreman and P. Török, "Fundamental limits in single-molecule orientation measurements," New J. Phys. **13**, 093013 (2011).
13. R. M. A. Azzam, "Division-of-amplitude Photopolarimeter (DOAP) for the Simultaneous Measurement of All Four Stokes Parameters of Light," Opt. Acta **29**, 685–689 (1982).
14. A. P. Bartko and R. M. Dickson, "Imaging three-dimensional single molecule orientations," J. Phys. Chem. B **103**, 11237–11241 (1999).
15. P. Dedecker, B. Muls, J. Hofkens, J. Enderlein, and J. Hotta, "Orientational effects in the excitation and de-excitation of single molecules interacting with donut-mode laser beams," Opt. Express **15**, 3372–3383 (2007).
16. F. Aguet, A. Geissbühler, I. Märki, T. Lasser, and M. Unser, "Super-resolution orientation estimation and localization of fluorescent dipoles using 3-D steerable filters," Opt. Express **17**, 6829–6848 (2009).
17. K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, "Optimized localization analysis for single-molecule tracking and super-resolution microscopy," Nat. Methods **7**, 377–381 (2010).
18. T. Wilson, R. Juskaitis, and P. D. Higdon, "The imaging of dielectric point scatterers in conventional and confocal polarisation microscopes," Opt. Commun. **141**, 298–313 (1997).
19. P. Török, P. D. Higdon, and T. Wilson, "Theory for confocal and conventional microscopes imaging small dielectric scatterers," J. Mod. Opt. **45**, 1681–1698 (1998).
20. C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, "Fast, single-molecule localization that achieves theoretically minimum uncertainty," Nat. Methods **7**, 373–375 (2010).
21. E. Zauderer, "Complex argument Hermite-Gaussian and Laguerre-Gaussian beams," J. Opt. Soc. Am. A **3**, 465–469 (1986).
22. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in Fortran 77," 2nd ed. (Cambridge Univeristy Press, 1992).
23. E. Toprak, H. Balci, B. H. Blehm, and P. R. Selvin, "Three-dimensional particle tracking via bifocal imaging," Nano Lett. **7**, 2043–2045 (2007).
24. M. F. Juette, T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf, "Three-dimensional sub-100nm resolution fluorescence microscopy of thick samples," Nat. Methods **5**, 527–530 (2008).
25. L. Holtzer, T. Meckel, and T. Schmidt, "Nanometric three-dimensional tracking of individual quantum dots in cells," Appl. Phys. Lett. **90**, (053902), 1–3 (2007).
26. B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy," Science **319**, 810–813 (2008).
27. S. R. P. Pavani and R. Piestun, "Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system," Opt. Express **16**, 22048–22057 (2008).

1. Introduction

The topic of this paper is the role of emitter dipole orientation in localization microscopy techniques such as PALM, STORM, or GSDIM [1–5]. In these techniques, emitters can switch between off and on-states such that only a sparse subset of all emitters is in the on-state for each image frame of a time lapse series. The location of all emitters can then be determined from the sequence of all image frames which each have a randomly different sparse subset of emitters that is active. The localization uncertainty can be well below the diffraction limit, mainly dependent on the photon count, and typically amounting to several tens of nanometers. In many cases the emitter can either rotate or change its conformation freely during the lifetime of the excited state of the emitter, typically a few ns. For the many excitation-emission cycles comprising an image frame of typically 10 ms, an average over randomly distributed emission dipole orientations will be observed. In other cases the single emitter cannot rotate freely, such as in solid-state or cryofrozen samples or for emitters immobilized at a surface or in a matrix, or rigidly bound to a macromolecuar complex. In those cases the fixed dipole orientation needs to be taken into account in the analysis, as it affects the observed spot shape.

In a previous paper written by us [6] we have investigated the role of dipole orientation and classical optical aberrations on localization accuracy of spot fitting with a Gaussian Point Spread Function (PSF) model. One of the key findings was that the combination of tilted dipoles and symmetric aberrations such as defocus or spherical aberration, gives rise to a very large localization bias of tens of nanometers. Similar conclusions have been obtained by Engelhardt
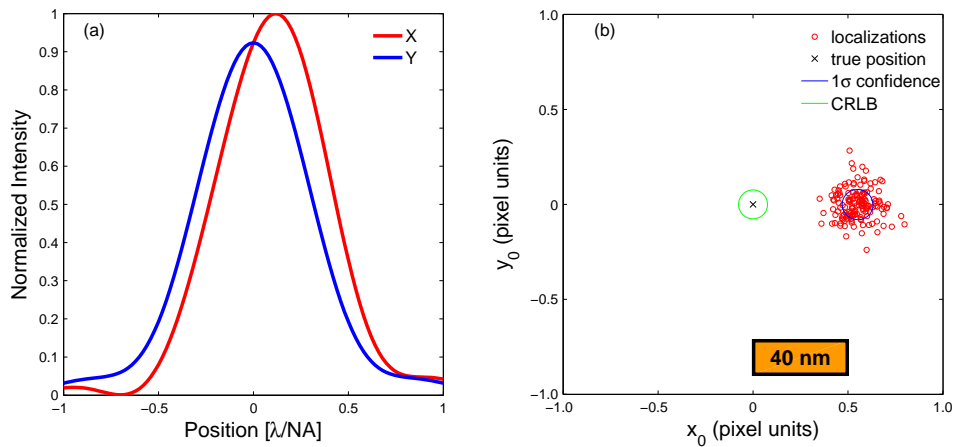
Fig. 1. Cross-sections of the PSF for a tilted dipole with defocus equal to the diffraction limit (left, (a)), and scatter plot of simulated localizations for a fixed dipole emitter with tilted dipole axis with defocus equal to the diffraction limit (boiling down to an axial object displacement of about $0.15\,\mu$m for the parameters assumed, $NA_{ob} = 1.25$, and $\lambda = 500$ nm) using a Gaussian PSF (right, (b)). The found positions, the $1 \times \sigma$ confidence level and the Gaussian Cramer-Rao Lower Bound (CRLB) are plotted. The localization was done using Maximum Likelihood Estimation (MLE) with a Gaussian PSF, taking 500 detected signal photons, and 25 background photons, the photons distributed over an 11×11 pixel large Region Of Interest (ROI) according to Poisson statistics.

and co-workers [7], and previously, in the context of Total Internal Reflection Fluorescence (TIRF) imaging, by Enderlein and co-workers [8]. The results of simulated localizations of a tilted dipole emitter with defocus are shown in Fig. 1 in order to illustrate the problem. Clearly, the variance of the fit is small and about equal to the Cramer-Rao Lower Bound (CRLB). There is, however, a bias of about 50 nm, which is unacceptably high.

The first goal of the current paper is to solve this bias problem for fixed dipole emitters with a tilted dipole axis. To that end the orientation of the dipole emitter must be measured. There are basically two ways to deduce dipole orientation. The first method is based on the use of polarization optics in the detection light path. Gould et al. [9] split the image in two sub-images via a Polarizing Beam Splitter (PBS) for measuring the rotational anisotropy (via cross-polarized illumination-imaging). Pavani et al. [10] also use a PBS in the detection light path for measuring a polarization contrast imaging modality. It appears that an accurate measurement of the full polarization and in that way of both the polar and azimuthal dipole angle requires a more sophisticated polarimeter architecture, e.g. as proposed in [11], or by one of the architectures described in [12]. In particular the Azzam-polarimeter [13] in which the image is split into four sub-images seems a suitable option. The second method to determine dipole orientation is via spot shape information. The root cause of the tilted dipole bias problem is that the spot of a tilted dipole with a little defocus cannot be reliably distinguished from the spot of a horizontal dipole with a position offset. A differentiation in shape can be accomplished by defocused image acquisition. Several papers have appeared about the measurement of dipole orientation using spot shape varations of defocused images of single dipole emitters [14, 15]. Extensions of this idea to measure both orientation and position have appeared as well. Aguet and co-workers determine position and orientation from defocused images by including a steerable filter approach for solving the orientation variables into a Least Mean Squares (LMS) fit for the

position [16]. Another study of combined position-orientation measurement has been presented by Mortensen and co-workers [17]. They consider methods that apply to in-focus TIRF imaging. Both papers contain results for images with relatively low noise levels (corresponding to photon counts which we estimate to be of the order 3×10^3 [16] to 3×10^4 [17]). It is not clear how these methods perform for more practical noise levels (corresponding to say 200-1000 photons per emitter per acquired image). In addition, the extent to which these methods suffer form the tilted dipole bias problem is not fully clarified.

A common denominator of all of the above work is the use of realistic vectorial PSF models, taking all effects of high NA and polarization into account [18, 19]. A major drawback of the use of such models in an iterative procedure to fit position and orientation is their computational inefficiency. This stands in stark contrast to the popular Gaussian PSF model, which works fine for freely and rapidly rotating dipoles [6] and can be used for real-time position estimation when programmed on paralel processing platforms such as graphics cards [20]. An additional, albeit minor, drawback of the full-fledged PSF models is that asymmetric aberrations, such as astigmatism for the purpose of 3D-imaging, cannot be easily introduced in the model.

In this paper we propose a method for determining position and orientation with two novel ingredients. First, we combine polarization and spot shape information to determine orientation rather than tread along a single one of these routes. In particular, we propose the use of an Azzam-polarimeter in the detection light path. We will show that this gives orientational information from polarization as well as from hitherto unreported spot shape variations. Second, we propose an effective PSF model based on Hermite functions. These provide a straightforward generalization of the isotropic Gaussian PSF model for describing the essentials of spot shape variations with dipole orientation and with the polarization branch in the detection light path. We show that the proposed method (i) solves the tilted dipole bias problem, (ii) has a defocus tolerance which is at least equal to Maréchal's diffraction limit, and (iii) works well at moderate photon counts corresponding to usually encountered experimental conditions.

The paper is organized as follows. In Section 2 the PSFs for imaging through the Azzam-polarimeter are presented, and section 3 introduces the effective Hermite PSF model. The unknown parameters of the model are found by MLE fitting decribed in section 4. Results are presented in section 5, conclusions and an outlook in section 6. A number of technical details are described in three appendices.

2. The Point Spread Functions in polarization imaging

A possible experimental implementation of the Azzam polarimeter is shown in Fig. 2. The beam captured by the objective is split into two parts, the polarization of one part is rotated over $\pi/4$, then both beams are split in two parts by a PBS. The four images are captured on four quadrants of a single EMCCD. Standard fields of view (512×512 pixels) are big enough to accommodate four images. Such a setup is quite akin to commonly used 'Optosplit'-setups. Not drawn in the figure are optical path length compensation plates for making the defocus in all channels equal.

Polarization rotation over $\pi/4$ can be achieved in a number of ways, for example with a quartz waveplate with fast axis along the optical axis (thus using only the optical activity, not the birefringence), or with a twisted liquid crystal waveplate operating in the Mauguin-regime, or with a set of two $\lambda/2$-plates, one with fast axis oriented along the x-axis, and the other with fast axis oriented at $\pi/8$ with the x-axis. An even different arrangement for producing the four polarization images can be made using a switchable polarization rotator. For example, having a first fixed $\lambda/2$-plate and a second fast switchable $\lambda/2$-plate, such as a Ferroelectric Liquid Crystal based shutter, in front of the PBS allows to record the $\pi/4$ rotated polarization pairs of images consecutively, rather than in parallel. Presumably, other polarization optical schemes
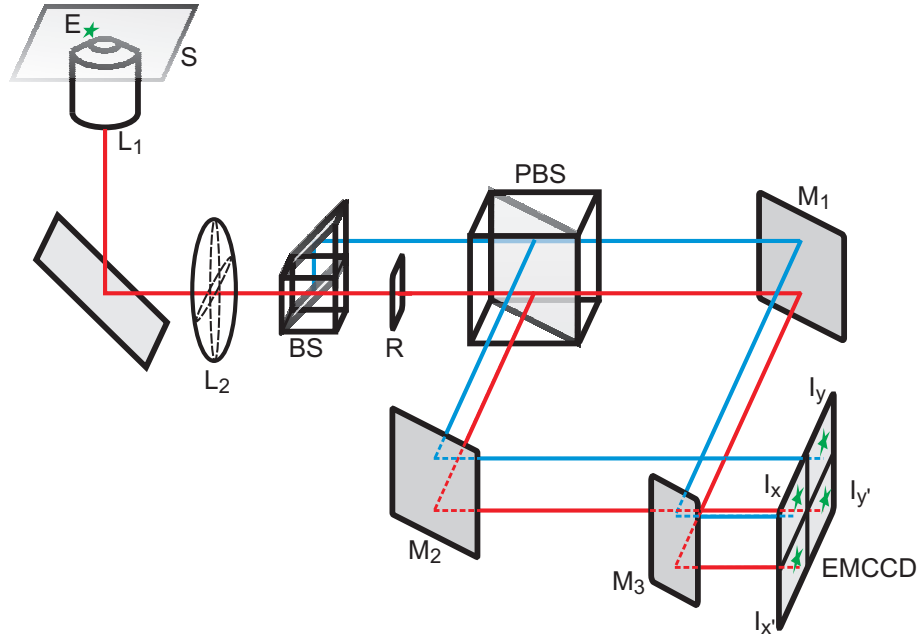
Fig. 2. Schematic view of a light path for simultaneous acquisition of four polarization channels on a single camera based on the Azzam polarimeter architecture. Light from emitter E is captured by lenses L_1 and imaged by lens L_2 onto the camera. In the imaging branch the beam is split into two equal parts with a normal beam splitter BS. The two branches are color coded for the sake of clarity, it does not indicate wavelength. The 'bottom' (red) branch passes through a polarization rotating (over $\pi/4$) component R. Both the 'top' (blue) and 'bottom' (red) branches pass a polarizing beam splitter (PBS) splitting each branch into two orthogonally polarized parts. The horizontal (x) and vertical (y) polarized beams of both branches are directed towards the camera ($EMCCD$) via three mirrors M_1, M_2, and M_3. The four channels I_x, I_y, $I_{x'}$, and $I_{y'}$ are projected onto the four quadrants of the camera, thus generating four images of the emitter E.

for reaching the same goal can be invented too.

The four PSFs corresponding to the four sub-images have a distinct shape depending on the emitter dipole orientation. These shapes are analyzed with the following formal developments. These developments closely follow the notations and conventions of our previous paper [6]. For the sake of completeness, a summary of all definitions and conventions on coordinate scalling, and of the basic equations of dipole image formation are given in appendix A. First we consider the pair of polarization images without $\pi/4$ polarization rotations. The electric field in the image plane (on the detector) at scaled position $\vec{u}$ originating from a dipole at scaled position $\vec{u}_d$ with dipole oriented along the unit vector $\vec{d} = (d_x, d_y, d_z)$ is given by:

$$E_{\text{im},j}(\vec{u}, \vec{u}_d) = E_0 \sum_{k=x,y,z} w_{jk}(\vec{u} - \vec{u}_d) d_k, \tag{1}$$

for $j = x, y$, where E_0 is the amplitude, and where the functions $w_{jk}(\vec{u})$ are the Fourier transforms of the pupil function times a matrix of angular functions. If there are no aberrations or only azimuthally symmetric aberrations (defocus, spherical aberration) the electric field can be expressed in terms of the three Richards and Wolf type of functions F_k ($k = 0, 1, 2$) defined in appendix A. It appears that the contribution from F_2 is generally much smaller than the contri-

bution from the other two functions. It is therefore a reasonable approximation to neglect this term. The electric field then turns out to be:

$$E_{\text{im},x}(u, \psi) = E_0 \left[F_0(u) d_x + iF_1(u) \cos \psi d_z \right], \tag{2}$$
$$E_{\text{im},y}(u, \psi) = E_0 \left[F_0(u) d_y + iF_1(u) \sin \psi d_z \right], \tag{3}$$

where polar coordinates (u, ψ) are used in the image plane. Polarization rotation over an angle χ changes this to:

$$E_{\text{im},x}(u, \psi) = E_0 \left[F_0(u) d_{x'} + iF_1(u) \cos(\psi - \chi) d_z \right], \tag{4}$$
$$E_{\text{im},y}(u, \psi) = E_0 \left[F_0(u) d_{y'} + iF_1(u) \sin(\psi - \chi) d_z \right], \tag{5}$$

where the rotated dipole vector components are:

$$d_{x'} = \cos \chi d_x + \sin \chi d_y, \tag{6}$$
$$d_{y'} = -\sin \chi d_x + \cos \chi d_y. \tag{7}$$

The resulting intensities after passing through the PBS follow as:

$$PSF_x(u, \psi) = E_0^2 \left[|F_0(u)|^2 d_{x'}^2 + 2\text{Im}\{F_0(u) F_1(u)^*\} \cos(\psi - \chi) d_{x'} d_z \right.$$
$$\left. + |F_1(u)|^2 \cos^2(\psi - \chi) d_z^2 \right], \tag{8}$$

$$PSF_y(u, \psi) = E_0^2 \left[|F_0(u)|^2 d_{y'}^2 + 2\text{Im}\{F_0(u) F_1(u)^*\} \sin(\psi - \chi) d_{y'} d_z \right.$$
$$\left. + |F_1(u)|^2 \sin^2(\psi - \chi) d_z^2 \right], \tag{9}$$

where $\chi = 0, \pi/4$ is the rotation angle. Clearly, the two PSFs with $\chi = \pi/4$ (from now on referred to as the x' and y' channels) give rise to the same spot shapes as for the pair with $\chi = 0$ (from now on referred to as the x and y channels), but now rotated over an angle $\pi/4$. The relative intensity of the two spots is however different, as it depends on the azimuthal angle of the dipole.

These expressions for the polarization PSFs can describe the gross features of spot shape as a function of dipole orientation. Figure 3 shows the numerically calculated PSFs for the four polarization channels for horizontal, tilted and vertical dipoles. For (near) horizontal dipoles the first term on the r.h.s of equations 8 and 9 is dominant, giving rise to conventional peaked spots with peak intensities d_x^2 and d_y^2 for the x and y-channels, and $d_{x'}^2$ and $d_{y'}^2$ for the x' and y'-channels. Clearly, these intensities can be very different depending on the azimuthal dipole angle. For (near) vertical dipoles the third term on the r.h.s of equations 8 and 9 is dominant, leading to double spots in each of the two channels, such that the line through the two sub-spots of each channel is oriented along the polarization axis of that channel. The sum of the x and y polarization spots, as well as the sum of the x' and y' channels, is an azimuthally symmetric doughnut spot. For tilted dipoles all terms matter, and we see a cross-over between the two different spot shapes. On top of that, the second term on the r.h.s of equations 8 and 9 is now of relevance and gives rise to offsets in the position of the peaks. The offset is directed along the axis of polarization of each of the polarizations channels, i.e. the x-polarized spot has an offset in the x-direction, whereas the y-polarized spot has an offset in the y-direction. The offsets in the x' and y' channels are along the diagonals of the image. This behaviour enables the disentanglement of dipole tilt and defocus from a mere displacement of the emitter.
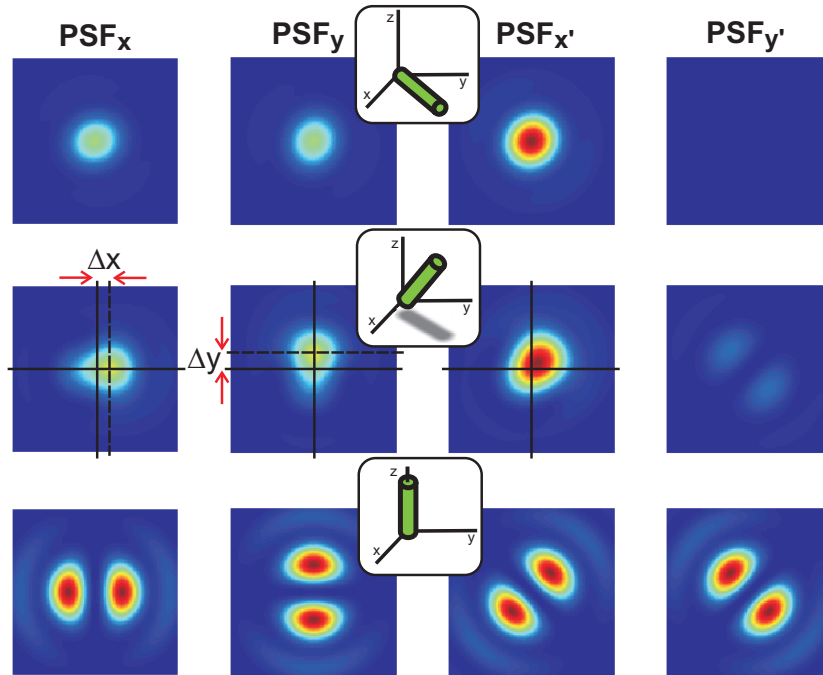
Fig. 3. The polarization PSFs in the four detection channels of the Azzam polarimeter architecture for horizontal dipoles with azimuthal angle $\pi/4$ (top row), tilted dipoles with azimuthal angle $\pi/4$ (middle row), and vertical dipoles (bottom row) dipoles for a field of view of size $2.2\times\lambda/\mathrm{NA}$, a water objective (NA = 1.25 in a medium with $n_{\mathrm{med}} = 1.33$), and diffraction limited defocus (72 mλ rms).

3. Effective Point Spread Function model based on Hermite functions

Localization algorithms ususally require the numerical optimization of some merit function. In each iteration step the PSF model must be compared to actual image data. Clearly, this requires evaluating the full PSF in each iteration step, which is numerically challenging. Moreover, it is questionable whether such an exact PSF model is needed, as much of the finer detail of the spot shape is presumably washed away in the noise and background. For these two reasons it makes sense to develop a PSF model depending on functions that are numerically easy to calculate and nonetheless are able to describe the wide variety of spot shapes that occur for different dipole orientations. These functions must execute these desirable tasks with reasonable accuracy for practical signal-to-noise and signal-to-background levels. The simplicity and convenience of the popular Gaussian suggests that we try a modification of the Gaussian function. Such a modification is for example the product of a polynomial with a Gaussian. A function like $x^2 \exp\left(-x^2/2\sigma^2\right)$ can be tried for the doughnut spots that occur for vertical dipoles, a function such as $x\exp\left(-x^2/2\sigma^2\right)$ can describe the spot asymmetry for tilted dipoles, and a Gaussian $\exp\left(-x^2/2\sigma^2\right)$ reasonably fits the spot shape for horizontal dipoles. An effective PSF model can be found by a linear combination of these functions, weighted with factors quadratic in the dipole vector components. It appears to be advantageous to write these products of polynomials

and a Gaussian function in terms of Hermite functions. The Hermite functions are defined as:

$$\psi_n(u) = H_n(u)\exp(-u^2),\tag{10}$$

with $H_n(s)$ the nth order Hermite polynomial ($H_0(u) = 1$, $H_1(u) = 2u$, $H_2(u) = 4u^2 - 2$, ...). Hermite polynomials have only a few applications in physics and engineering. Most famously, of course, is in the eigenfunctions of the Schrödinger equation for the quantum mechanical harmonic oscillator. Another applications is in Hermite-Gaussian beams which solve the paraxial wave equation [21].

The aforementioned linear combination of Hermite functions up to second order is:

$$
\begin{aligned}
PSF_x(x,y) &= \frac{A}{2\pi\sigma^2}\left(d_{x'}^2 + \frac{1}{2}\beta d_z^2\right)\Psi_{00}(x,y) + b/a^2 \\
&+ \frac{A\alpha\sqrt{\beta}}{2\pi\sigma^2}d_{x'}d_z\left(\cos\chi\Psi_{10}(x,y) + \sin\chi\Psi_{01}(x,y)\right) \\
&+ \frac{A\beta}{8\pi\sigma^2}d_z^2\left(\cos^2\chi\Psi_{20}(x,y) + 2\sin\chi\cos\chi\Psi_{11}(x,y) + \sin^2\chi\Psi_{02}(x,y)\right),\tag{11} \\
PSF_y(x,y) &= \frac{A}{2\pi\sigma^2}\left(d_{y'}^2 + \frac{1}{2}\beta d_z^2\right)\Psi_{00}(x,y) + b/a^2 \\
&+ \frac{A\alpha\sqrt{\beta}}{2\pi\sigma^2}d_{y'}d_z\left(-\sin\chi\Psi_{10}(x,y) + \cos\chi\Psi_{01}(x,y)\right) \\
&+ \frac{A\beta}{8\pi\sigma^2}d_z^2\left(\sin^2\chi\Psi_{20}(x,y) - 2\sin\chi\cos\chi\Psi_{11}(x,y) + \cos^2\chi\Psi_{02}(x,y)\right),\tag{12}
\end{aligned}
$$

with the Hermite basis functions:

$$\Psi_{lm}(x,y) = \psi_l\left(\frac{x-x_0}{\sqrt{2}\sigma}\right)\psi_m\left(\frac{y-y_0}{\sqrt{2}\sigma}\right),\tag{13}$$

for integer l and m. Here $x = u\cos\psi$ and $y = u\sin\psi$ are the image plane coordinates, and χ is the polarization rotation angle which takes values equal to zero for the first image pair and equal to $\pi/4$ for the second image pair. The emitter coordinates are x_0 and y_0, A is a measure for the signal photon count, σ is the width of the PSF, the number of background photons in each polarization channel is b per pixel area $a \times a$ (the background is assumed to be unpolarized), α is an effective asymmetry parameter, and β is a measure of the relative magnitude of the doughnut contribution to the total spot. It can be shown that the PSFs are positive definite provided that the asymmetry parameter satisfies $|\alpha| \leq 1$ and the background parameter $b \geq 0$.

The distribution of the signal photons over the two channels is found from the integral of the PSFs over the image plane:

$$N_x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dxdy\left(PSF_x(x,y) - b/a^2\right) = A\left(d_{x'}^2 + \frac{1}{2}\beta d_z^2\right),\tag{14}$$

$$N_y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dxdy\left(PSF_y(x,y) - b/a^2\right) = A\left(d_{y'}^2 + \frac{1}{2}\beta d_z^2\right),\tag{15}$$

which gives a total signal photon count $N = N_x + N_y = A\left(d_{x'}^2 + d_{y'}^2 + \beta d_z^2\right) = A\left(d_x^2 + d_y^2 + \beta d_z^2\right)$.

It appears to be convenient to absorb the signal photon count variable A, the unit-vector along the dipole axis $\vec{d}$ and the parameter β into a single vector $\vec{D}$ of three independent variables defined by:

$$(D_x, D_y, D_z) = \sqrt{A}\left(d_x, d_y, \sqrt{\beta}d_z\right).\tag{16}$$

With this substitution the model has eight free parameters: $(x_0, y_0, \sigma, D_x, D_y, D_z, \alpha, b)$. The signal photon count and the dipole unit vector can be found from the $\vec{D}$-vector via the inverse transformation:

$$N = D_x^2 + D_y^2 + D_z^2, \tag{17}$$

$$(d_x, d_y, d_z) = \frac{1}{\sqrt{D_x^2 + D_y^2 + D_z^2/\beta}} (D_x, D_y, D_z/\beta). \tag{18}$$

This necessitates prior knowledge of the parameter β, which thus is the only 'magic' parameter of our model. It turns out that β only depends on fixed parameters such as NA and refractive indices and on the value of the asymmetry parameter α. Results in support of this connection are presented in section 5.

Deviations of the found orientation from a ground truth dipole direction $\vec{d}_0 = (\sin\theta_0\cos\phi_0, \sin\theta_0\sin\phi_0, \cos\theta_0)$ can be quantified with two orientational fluctuation variables:

$$Q_p = \vec{p}_0 \cdot \vec{d}, \tag{19}$$

$$Q_s = \vec{s}_0 \cdot \vec{d}, \tag{20}$$

where $\vec{p}_0 = (\cos\theta_0\cos\phi_0, \cos\theta_0\sin\phi_0, -\sin\theta_0)$ and $\vec{s}_0 = (-\sin\phi_0, \cos\phi_0, 0)$ are the two unit vectors in the polar and azimuthal direction.

The power captured by the pixel centered at (x_k, y_k) follows from integration of the intensity over the $a \times a$ pixel area:

$$\mu_k^l = \int_{x_k-a/2}^{x_k+a/2} \int_{y_k-a/2}^{y_k+a/2} dx\, dy\, PSF_l(x, y). \tag{21}$$

This integration can be done analytically using the different (recursion) relations satisfied by the Hermite functions:

$$\frac{d\psi_m(u)}{du} = \left(\frac{dH_m(u)}{du} - 2uH_m(u) \right) e^{-u^2} = (2mH_{m-1}(u) - 2uH_m(u)) e^{-u^2} = -H_{m+1}(u)e^{-u^2}$$

$$= -\psi_{m+1}(u). \tag{22}$$

It now follows that:

$$\mu_k^x = \frac{1}{\pi} \left[\left(D_{x'}^2 + \frac{1}{2}D_z^2 \right) \Sigma_{00}(x_k, y_k) + \alpha D_{x'}D_z \left(\cos\chi\Sigma_{10}(x_k, y_k) + \sin\chi\Sigma_{01}(x_k, y_k) \right) \right.$$
$$\left. + \frac{1}{4}D_z^2 \left(\cos^2\chi\Sigma_{20}(x_k, y_k) + 2\sin\chi\cos\chi\Sigma_{11}(x_k, y_k) + \sin^2\chi\Sigma_{02}(x_k, y_k) \right) \right] + b, \tag{23}$$

$$\mu_k^y = \frac{1}{\pi} \left[\left(D_{y'}^2 + \frac{1}{2}D_z^2 \right) \Sigma_{00}(x_k, y_k) + \alpha D_{y'}D_z \left(-\sin\chi\Sigma_{10}(x_k, y_k) + \cos\chi\Sigma_{01}(x_k, y_k) \right) \right.$$
$$\left. + \frac{1}{4}D_z^2 \left(\sin^2\chi\Sigma_{20}(x_k, y_k) - 2\sin\chi\cos\chi\Sigma_{11}(x_k, y_k) + \cos^2\chi\Sigma_{02}(x_k, y_k) \right) \right] + b, \tag{24}$$

where:

$$\Sigma_{lm}(x_k, y_k) = \Delta\psi_{l-1}\left(\frac{x_k - x_0}{\sqrt{2}\sigma} \right) \Delta\psi_{m-1}\left(\frac{y_k - y_0}{\sqrt{2}\sigma} \right), \tag{25}$$

with the shorthand defined for functions $f(u)$:

$$\Delta f(u) = f\left(u + \frac{a}{2\sqrt{2}\sigma} \right) - f\left(u - \frac{a}{2\sqrt{2}\sigma} \right), \tag{26}$$

and where the definition of the Hermite functions is extended by :

$$\psi_{-1}(u) = -\frac{\sqrt{\pi}}{2}\text{erf}(u).$$ (27)

The components of the vector $\vec{D}$ in the rotated frame are $D_{x'} = \cos\chi D_x + \sin\chi D_y$ and $D_{y'} = -\sin\chi D_x + \cos\chi D_y$.

The recursion relations provide another numerical advantage, apart from the exact integration over the pixel area. Namely, it is possible to obtain analytical expressions for derivatives of the pixel averaged PSF μ_k^l w.r.t. the fit parameters, which can then be evaluated numerically with great efficiency. Such expressions are needed in the MLE fitting routine.

Figure 4 shows cross-sections along the x and y-axes of realistic vectorial PSFs and Hermite PSF model fits for the four polarization channels, and for horizontal, tilted, and vertical dipoles. Apparently, the Hermite PSF model gives a good quantitative description of the different spot shapes, small differences arise with the fringe structures surrounding the main spot.

4. Levenberg-Marquardt based Maximum Likelihood Estimation

The observed parameters are the set of photon counts in the four channels $N = \{n_k^l\}$ for $l = x, y, x', y'$, the unknown parameters are $\theta = \{x_0, y_0, \sigma, D_x, D_y, D_z, \alpha, b\}$. In Maximum Likelihood Estimation (MLE) the likelihood for observing N given the unknown parameters θ is maximized. It is customary to work with the logarithm of the likelihood function, as that appears to be a more amenable functional than the likelihood itself. The log-likelihood function is derived from the properties of the statistical process leading to the observed parameters N. In case the photon count per pixel follows a Poisson distribution with expectation value $\mu_k^l(\theta)$, we obtain the log-likelihood:

$$\log L = \sum_l \sum_k \left[n_k^l \log \mu_k^l - \mu_k^l - \log\left(n_k^l!\right) \right].$$ (28)

The log-likehood function is maximum if the gradient of the functional w.r.t. the parameters θ is zero. The first order derivatives (gradient) and second order derivatives (Hessian) of the log-likelihood are given by:

$$G_i = \frac{\partial \log L}{\partial \theta_i} = \sum_l \sum_k \left(\frac{n_k^l}{\mu_k^l} - 1 \right) \frac{\partial \mu_k^l}{\partial \theta_i},$$ (29)

$$H_{ij} = \frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} = -\sum_l \sum_k \frac{n_k^l}{\mu_k^{l2}} \frac{\partial \mu_k^l}{\partial \theta_i} \frac{\partial \mu_k^l}{\partial \theta_j} + \sum_l \sum_k \left(\frac{n_k^l}{\mu_k^l} - 1 \right) \frac{\partial^2 \mu_k^l}{\partial \theta_i \partial \theta_j}.$$ (30)

Expressions for the first order derivatives of the pixel Poisson-rates μ_k^l, derived by applying the recursion relations satisfied by the Hermite functions, are given in appendix B. The terms in Eq. (30) involving the second order derivatives of the pixel Poisson-rates are usually neglected in the optimization procedure [22], giving rise to a simplified Hessian:

$$H_{ij}^s = -\sum_l \sum_k \frac{n_k^l}{\mu_k^{l2}} \frac{\partial \mu_k^l}{\partial \theta_i} \frac{\partial \mu_k^l}{\partial \theta_j}.$$ (31)

The reason is that these terms are linear in the difference $\mu_k^l - n_k^l$, and are therefore expected to be small close to the optimum, provided the model reasonably fits the observations, and provided the initial values for the parameters are not too far off. An additional justification is that
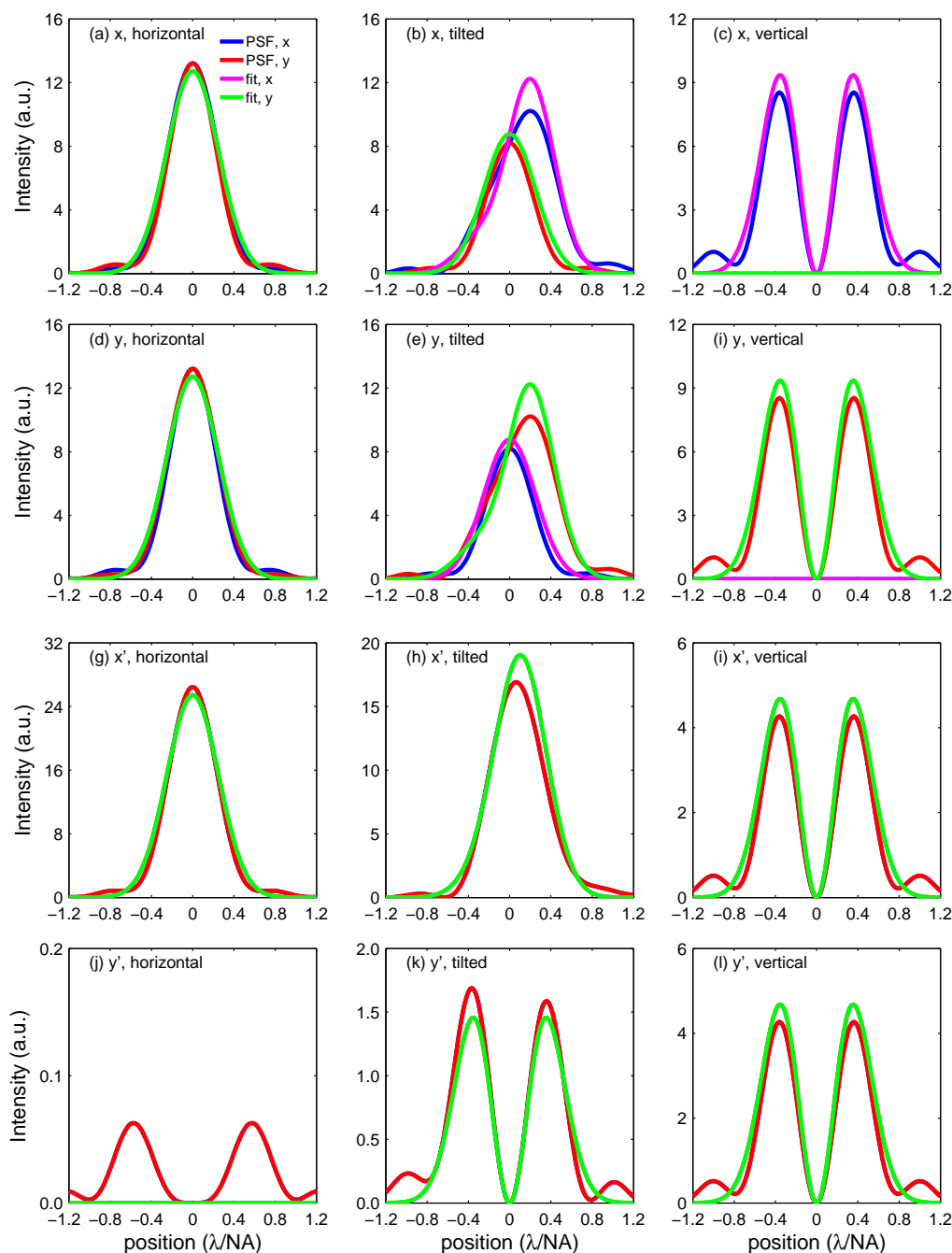
Fig. 4. Cross-sections of realistic vectorial PSFs for diffraction limited defocus (72 mλ rms) and Hermite PSF model fits in the four polarization detection channels, arranged by rows, for three different dipole orientations (horizontal, tilted, vertical), arranged by columns. The horizontal and tilted dipoles have azimuthal angle $\pi/4$. Note that the vertical axes in the figures are adapted to the peak height of each image so as to make the spot shapes better visible. The x and y cross-sections for the two rotated polarization channels overlap exactly, thus making only two of the four curves visible in figures (g) to (i) (the red curve overlaps the blue curve, the green curve overlaps the magenta curve).

the precise form of the Hessian does not affect the optimum in parameter space, only the path of convergence to the optimum. There is also a practical advantage to using this simplified Hessian H^s, as now the computation of the second order derivatives of the Poisson-rates is avoided. It is noted that there is no fundamental reason not to calculate the second order derivatives. If it is wanted or perhaps needed the full expression for the Hessian may be used as well, which thus comes at the price of finding and evaluating expressions for the second order derivatives of the Poisson-rates.

The update rule for the parameters in each iteration step of the Levenberg-Marquardt optimization algorithm is:

$$\theta \rightarrow \theta - [H^s + \lambda \text{diag}\,(H^s)]^{-1} \cdot G. \tag{32}$$

In case the parameter update does not result in an increase of the log-likelihood it is tried again with a Levenberg-Marquardt parameter λ that is increased by a factor η. In case the parameter update does result in an increase of the log-likelihood the parameter update is accepted and the Levenberg-Marquardt parameter λ is decreased by the factor η. We take as starting value $\lambda = 1$ and a multiplication factor $\eta = 10$. The iteration is terminated when the relative increase in the merit function is less than 5×10^{-5}. Convergence is typically achieved in 6-10 iterations. In our experience the speed of convergence is better compared to the modified Newton-Raphson method of [20] (which uses $\lambda = 1$ and only the diagonal part of the Hessian), and the stability of convergence is an improvement over the pure Newton-Raphson method (which uses the setting $\lambda = 0$).

Initial values for the parameters are calculated with an algorithm described in appendix C. The method is based on fitting the model to the lowest (up to second) order moments of the light distribution across the ROI. This results in values that are already close to the optimum, especially for low background, thus improving convergence of the MLE procedure.

The Fisher information matrix is:

$$F_{ij} = \sum_l \sum_k \frac{1}{\mu_k^l} \frac{\partial \mu_k^l}{\partial \theta_i} \frac{\partial \mu_k^l}{\partial \theta_j}. \tag{33}$$

In the next section we present results not for the vector components (D_x, D_y, D_z) but instead for the photon count N and for the orientational fluctuations (Q_p, Q_s). Such a change in variables $(D_x, D_y, D_z) \rightarrow (N, Q_p, Q_s)$ means we must replace the Fisher matrix following $F \rightarrow F' = JFJ^\dagger$, with J the Jacobian of the coordinate transformation. The Cramér-Rao Lower Bound (CRLB) on the parameter covariances is found by inverting the Fisher information matrix. The CRLB gives the best possible accuracy in estimating the parameters θ_i, provided the model fits the underlying reality. Clearly, in our case there is no exact match between the two as we work with a simplified PSF model based on Hermite functions. We will show, however, that our estimator is to a great extent unbiased, in the sense that the parameter averages over all statistical realizations corresponds to the assumed ground truth values of those parameters, thus lending support to the application of the CRLB concept.

5. Numerical results

Results for the localization performance of the MLE algorithm are plotted in Fig. 5. In these simulation sets four polarization PSFs are generated according to the model described in the appendix. The NA is taken to be 1.25 (water immersion), the wavelength is 500 nm, each pixel is 80 nm in object space (slightly oversampled, Nyquist would imlpy 100 nm pixels), and the ROI is 11×11 pixels large. The emitter coordinates are drawn from a normal distribution with a width equal to one pixel. The four polarization images are distorted by Poisson-noise and background. A total number of 500 signal photons and 25 background photons (making
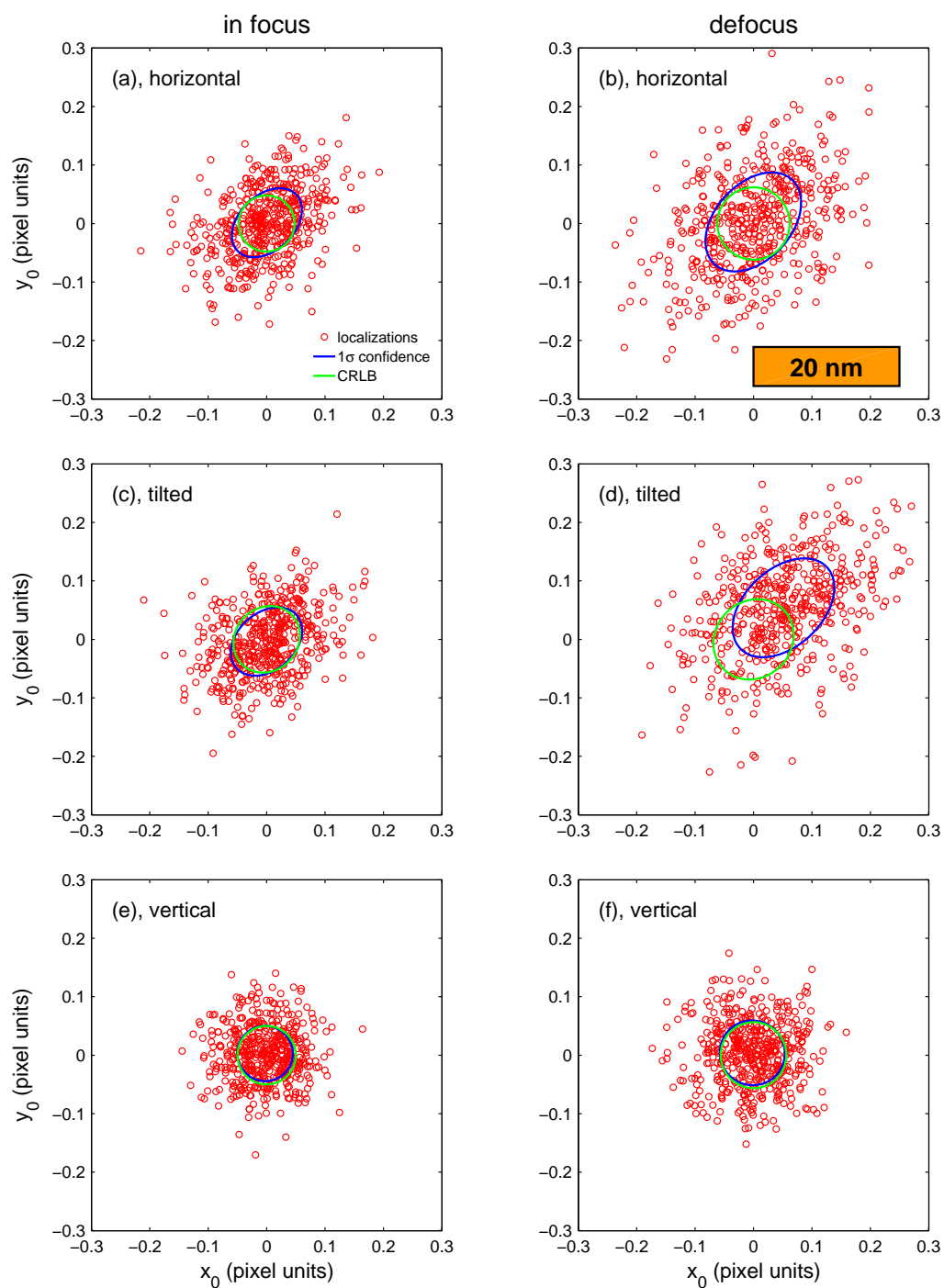
Fig. 5. Scatter plots of localizations for in focus images (left column) and for images with diffraction limited defocus (right column), for horizontal dipoles (top row), tilted dipoles (middle row), and vertical dipoles (bottom row).
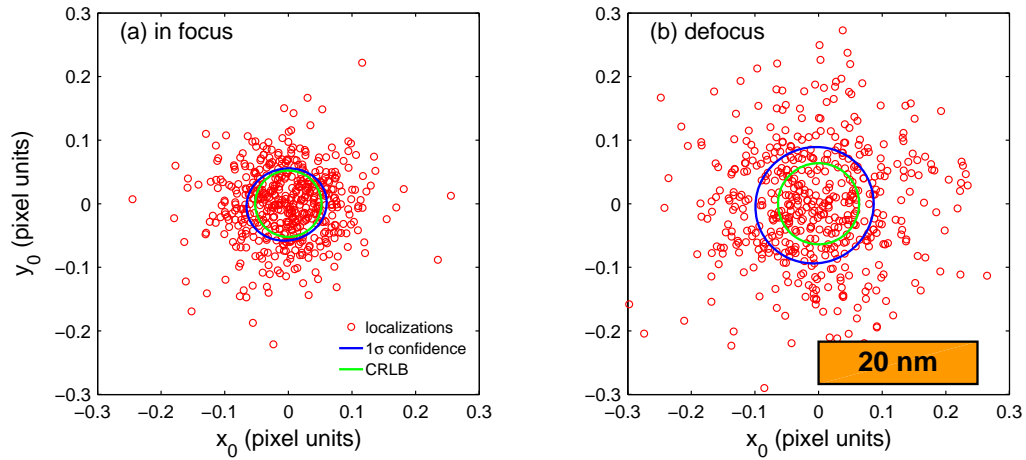
Fig. 6. Scatter plots of localizations for images of fixed dipole emitters with random dipole orientation in focus (left) and with diffraction limited defocus (right).

the signal-to-background ratio equal to 20) is distributed over the four images following Poisson statistics. So, the sum of the four images contains on average 500 signal photons and 25 background photons. The simulation is run for 500 instances. Fits with a final log-likelihood deviating more than three standard deviations from the mean are designated as outliers and are removed from the data set. Typically less than 1% of the runs results in an outlier, indicating that the fitting routine is quite robust. Figure 5 shows scatter plots of the found position errors for horizontal dipoles ($\theta_d = \pi/2$, $\phi_d = \pi/4$), tilted dipoles ($\theta_d = \pi/4$, $\phi_d = \pi/4$), vertical dipoles ($\theta_d = 0$) for the case of zero defocus and for defocus equal to the diffraction limit. For the zero defocus case localization errors are in the range 4-6 nm, where the largest localization error is for horizontal and tilted dipoles in the tilting direction. For comparison, rapidly rotating dipoles, for which the averaged dipole PSF can be well fitted with a Gaussian, give rise to a localization error of 4-5 nm for the same parameters and for the same photon count. Clearly, the maximum achievable localization precision is hardly compromised by dividing the total number of photons over multiple images. The diffraction limited defocus case gives rise to localization errors in the range 5-9 nm, where the worst case occurs for horizontal and tilted dipoles in the tilting direction. The bias for tilted dipoles is reduced to about 5 nm, which is a reduction with an order of magnitude compared to the state-of-the-art Gauss fitting [6]. The typical bias in each of the two coordinates is less by a factor $\sqrt{2}/3$, and amounts to about 2 nm. This shows up in the localization uncertainty for random fixed dipoles (see Fig. 6), which is somewhat above CRLB. In conclusion, the proposed MLE estimation achieves near optimum localization unertainty with a small residual localization bias of 5 nm/2 nm (worst case/typical case), and hardly underperformes compared to the much more simple case of rapidly rotating dipoles, for which Gaussian PSF fitting works very well.

The dipole orientation is determined from the dipole vector components (D_x, D_y, D_z) via Eq. (18). Figure 7 shows data connecting the found dipole vector components to the ground truth dipole direction along the unit vector (d_x, d_y, d_z). The simulation was done for a set of 500 dipole emitters with random but fixed orientation in a watery medium ($n_{med} = 1.33$), both for conditions corresponding to a water immersion objective (NA = 1.25) and to an oil immersion

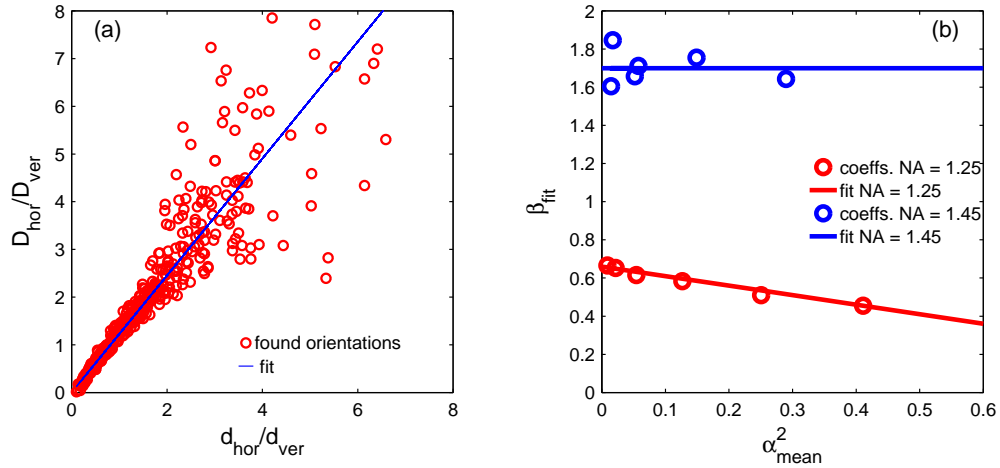Fig. 7. (a) Plot of the found ratio D_{hor}/D_{ver} as a function of the ground truth ratio d_{hor}/d_{ver} for a set of random fixed dipoles, and for a water immersion objective with NA = 1.25, and a linear fit through the data. (b) Plot of the found coefficient of linearity as a function of the average value of the squared asymmetry parameter for a water immersion objective (NA = 1.25) and an oil immersion objective (NA = 1.45).

objective (NA = 1.45, for capturing more light via supercritical fluorescence in TIRF-mode). The linear relation between $D_{hor}/D_{ver} = \sqrt{D_x^2 + D_x^2}/D_z$ and $d_{hor}/d_{ver} = \sqrt{d_x^2 + d_x^2}/d_z$ following from Eq. (18) works rather well. It is mentioned that the fitted value of β is based only on the orientations with intermediate tilt angles, as the scatter for orientations very close to the vertical or horizontal orientation tends to dominate the outcome of the fit. For the case of a water immersion objective, the coefficient of linearity turns out to depend quadratically on the asymmetry parameter: $\beta_{fit} = \beta_0 - \alpha^2/2$, where the parameter β_0 is found as $\beta_0 = 0.66$. Other values for the NA would necessitate a calibration similar to the one presented in Fig. 7 in order to find the right numerical value for β_0. For the case of an oil immersion objective the best fit indicates a constant value $\beta = 1.7$. The difference with the water immersion case is probably related to the evanescent wave contributions to the overall spot. It may be expected that the proper value of β depends not only on NA but also on the refractive index of the medium. Prior knowledge of this latter parameter is thus also needed to calibrate the parameter β. An alternative to the phenomenological relationships presented here is to use a look-up table for connecting $\vec{D}$ to $\vec{d}$.

Figure 8 shows scatter plots for the orientational fluctuations for the same cases considered as for the localization. For the vertical dipole case the angular uncertainty is around 0.8 deg in all directions, which makes sense as nothing breaks the rotational symmetry around the vertical axis. For the intermediate, tilted dipole case the polar angle uncertainty is slightly larger, around 2.5 deg in focus and around 3.0 deg with defocus equal to the diffraction limit. For the horizontal dipole case polar angle uncertainty is around 5 deg in focus and around 4 deg with defocus equal to the diffraction limit. The uncertainty in the azimuthal angle for both the tilted and horizontal dipole cases is around 1.6 deg both in and out of focus. The results for fixed dipoles with random orientation, shown in Fig. 9, give around 1.7 deg in the azimuthal direction and 4-5 deg in the polar direction, depending on the amount of defocus. Clearly, the nominal level of accuracy achieved, and the tolerance for defocus, are quite reasonable. The performance for
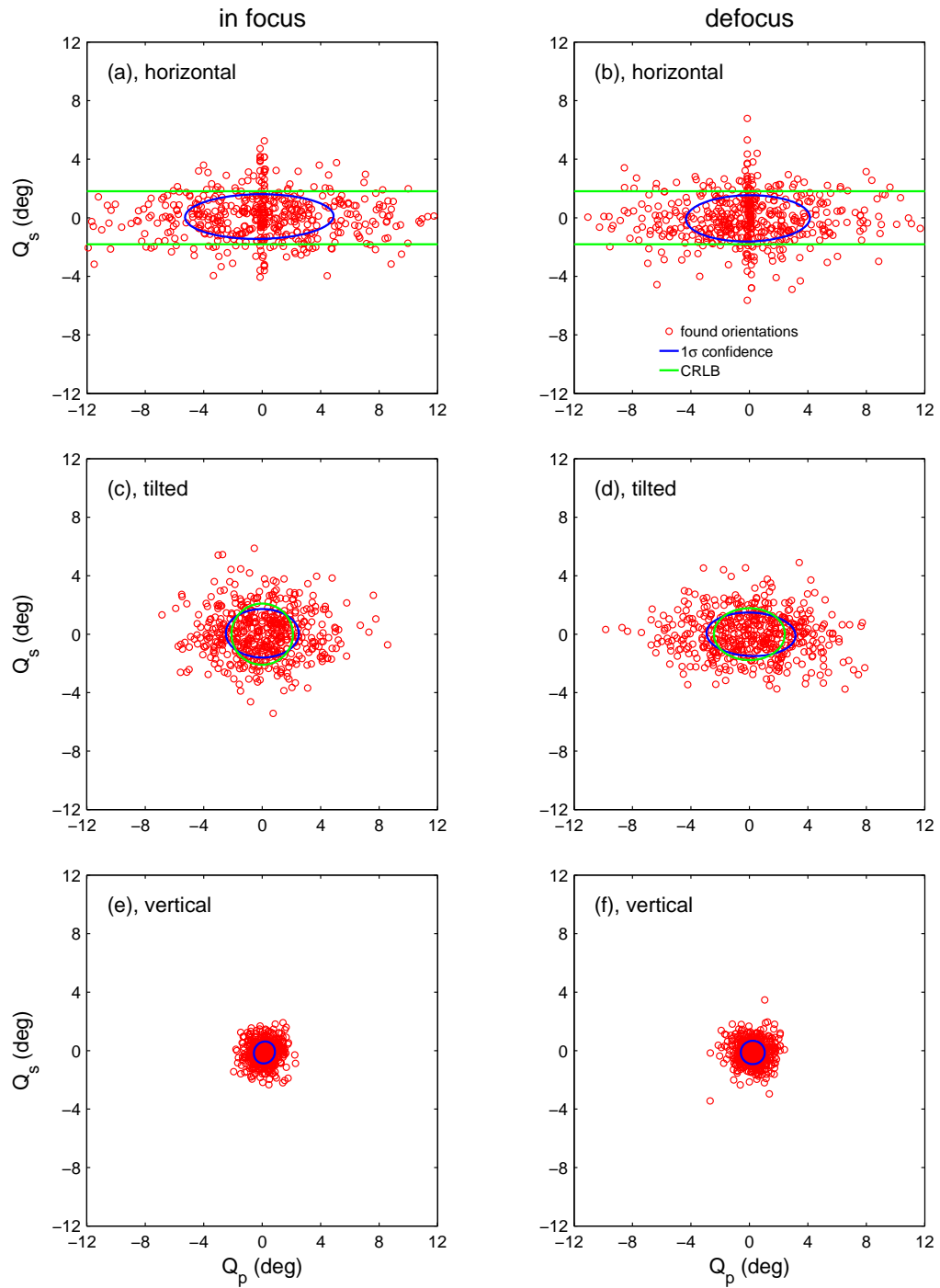
Fig. 8. Scatter plots of the variations of the found orientation around the ground truth orientation in the polar direction (Q_p) and in the azimuthal direction (Q_s) for in focus images (left column) and for images with diffraction limited defocus (right column), for horizontal dipoles (top row), tilted dipoles (middle row), and vertical dipoles (bottom row).
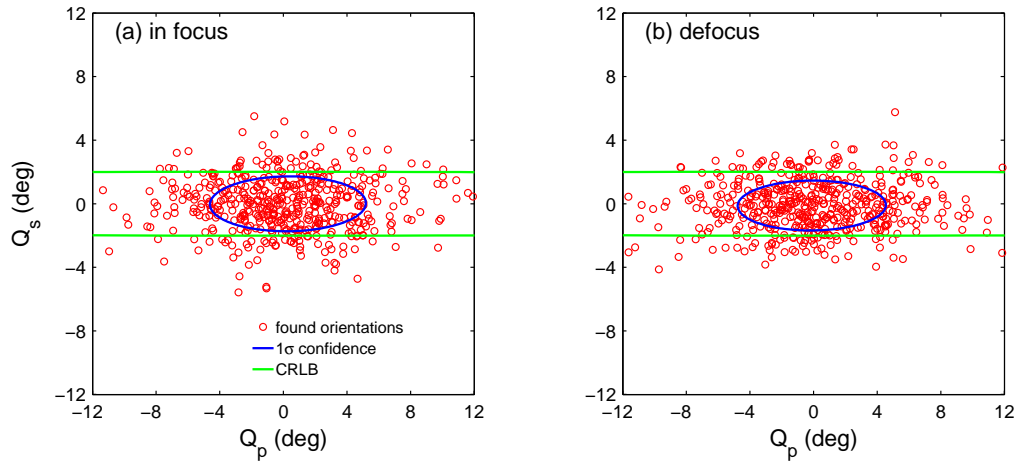
Fig. 9. Scatter plots of the variations of the found orientation around the ground truth orientation for images of fixed dipole emitters with random dipole orientation in focus (left) and with diffraction limited defocus (right).

most orientations is comparable to the results reported in [16], with the exception of vertical dipoles, for which the current results are an improvement.

The comparison between the achieved accuracy and the averaged CRLB of the Hermite PSF model at the found parameters reveals several striking features. For the tilted dipole case it appears that the angular accuracy is a bit above the average CRLB, according to prior expectations. For the horizontal dipole case the CRLB fits with the azimuthal uncertainty, but is much larger than the polar uncertainty. For the vertical dipole case the CRLB is much larger than the achieved uncertainty in both angular directions. The reason for this behaviour is that the Fisher matrix of the Hermite PSF model is singular for certain parameter values. This occurs when the asymmetry coefficient $\alpha = 0$ and when $D_z = 0$ (horizontal dipole) or $D_x = D_y = 0$ (vertical dipole). Close to these singular cases the CRLB for the polar direction (horizontal dipole) or for both angular directions (vertical dipole) grows very large. There are a number of reasons why the CRLB diverges while the realized uncertainty found from the simulations remains finite. First of all, the fact that the Fisher matrix is singular at these special points does not imply that the actual Hessian is singular, as it differs from the Fisher matrix by a noise term proportional to the difference in observed and model pixel values $n_k - \mu_k$, and by the additional diagonal Levenberg-Marquardt term. In our simulations it appears that for the vertical dipole case the Levenberg-Marquardt parameter λ indeed does not get small upon convergence to the optimum, as would be expected for a non-singular Hessian. A consequence of a (near) singular Hessian would be a deviation in the distribution of the found parameters from the normal distribution, which would be especially visible under low noise conditions, as then the difference between the Hessian and the singular Fisher-matrix vanishes. We have done Kolmogorov-Smirnov tests on the distribution of the polar angle fluctuation Q_p to check this. It turns out that indeed the distribution of found Q_p is not a normal distribution for horizontal and vertical dipoles at high photon counts (above about 10^3 for the vertical dipole case and even lower for the horizontal dipole case). A second reason for the difference between the averaged CRLB and the realized uncertainty is the variation in the asymmetry parameter α. For the horizontal and vertical case this parameter cannot be determined accurately, so values in the entire range $-1 \leq \alpha \leq 1$ will

be found. For non-zero α the Fisher matrix is not singular, thus regularizing the CRLB. We suspect that this is also the reason why the polar uncertainty for the horizontal dipole improves with defocus, as then the tendency for non-zero α increases. Yet a third reason may be found from the simplified nature of the Hermite PSF model. It may be the case that the Fisher matrix for the exact dipole PSF model is not singular for the horizontal and vertical dipole cases.

We have not developed a full understanding of why the realized uncertainty in the polar angle has the finite value it appears to have. We have tried several methods to regularize the calculation of the average CRLB in order to be able to predict the outcome of the simulations. For example, averaging the Fisher matrix and then taking the inverse instead of the other way around turns out to give a good fit with the observed polar uncertainty for the horizontal dipole case. However, this positive result does not extend to other values of the polar angle. Another regularization method that has been tried is to evaluate the CRLB of a suitable function of the polar angle, say $\cos(2\theta_p)$, which has a zero derivative at $\theta_p = 0$ and $\theta_p = \pi/2$. Although such a procedure gives finite results for the CRLB at these two orientations, it does not give a proper description of the realized uncertainty in the numerical simulations. Summarizing, the good quality of the realized angular uncertainty may be somewhat fortuitous and at present beyond description by a predictive model. Finally, we speculate that on a fundamental level the use of defocus for determining dipole orientation may provide an advantage, as then a formally diverging CRLB may be avoided.

Figure 10 shows the rms localization error (RMSLE), the rms orientational error in the s-direction (RMSOE$_s$), and in the p-direction (RMSOE$_p$) as a function of the signal photon count for different angles and defocus values and for zero background. These error measures incorporate both bias and variance effects in a single number. As benchmark for the localization results the Gaussian localization CRLB equal to $\sqrt{2}\sigma/\sqrt{N_{\mathrm{ph}}}$ (the factor $\sqrt{2}$ comes from adding the x and y contribution to the RMSLE) is plotted as well. In a recent paper Foreman and Török [12] show that the azimuthal angle of in-plane dipoles can be measured with a polarimeter architecture like the one considered in this paper with a CRLB equal to $1/\sqrt{2N_{\mathrm{ph}}}$. This line is plotted in the orientational error curves as a benchmark. It appears that the localization uncertainty in focus follows the Gaussian CRLB quite closely. The error is slightly larger in case of defocus for horizontal and tilted dipoles, in particular due to the small residual bias of a few nm for the tilted dipole case. This bias turns up at high photon counts as a plateau the localization error flattens off to. The azimuthal orientation error is excellently described by the bound derived by Foreman and Török for the tilted dipole case, both in focus and out of focus. For the horizontal dipole case the agreement is reasonably well, the error is a bit below the bound for low photon counts and a bit above the bound for high photon counts. The vertical dipole case is well below the bound for low photon counts, but gives an increase in uncertainty for high photon counts. The uncertainty in the polar orientation is significantly larger than the benchmark, for intermediate tilt angles by a factor of about 2, for larger tilt angles by a factor of about 3. Both the relatively large uncertainty in the polar angle and the increase in angular uncertainty of vertical dipoles for large photon counts is attributed to the near-singular Fisher matrix close to the horizontal or vertical dipole orientation, giving rise to non-normal data distributions around the mean with long tails to high values for the orientational error. Another deteriorating effect is a small bias in the polar angle for the tilted dipole case at high photon counts.

6. Conclusion

In summary, we have shown that the realized localization error closely follows the CRLB derived from the Gaussian PSF model, for all dipole orientations, for all practical photon counts and for defocus values at least up to the diffraction limit. Clearly, this provides a solution for the tilted dipole bias problem. The uncertainty in the azimuthal orientation is reasonably well
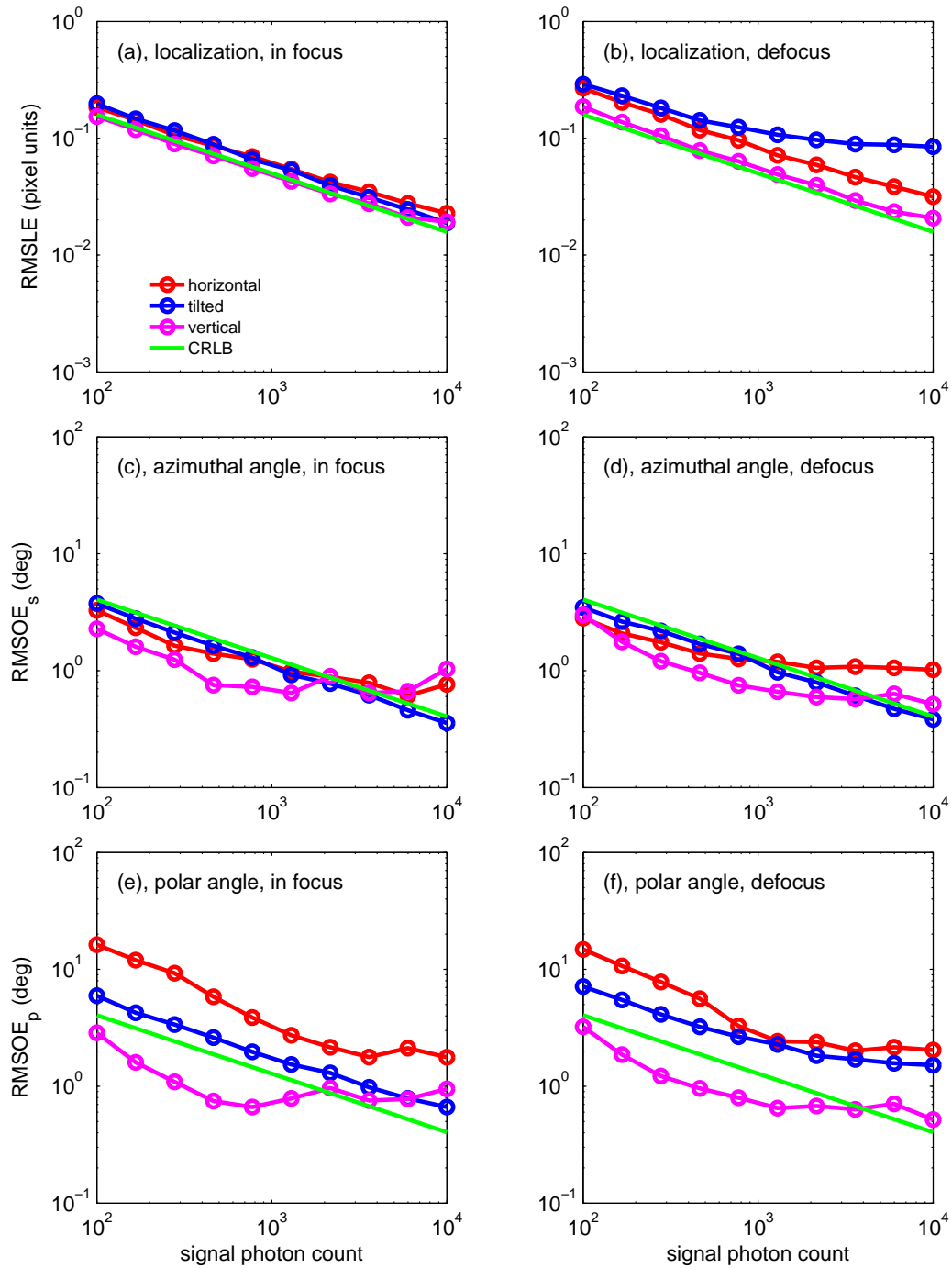
Fig. 10. Plots of the rms localization error (RMSLE, top row), the rms orientational error in the s-direction (RMSOE$_s$, middle row), and the rms orientational error (RMSOE$_p$, bottom row) as a function of the signal photon count.

described by the Foreman-Török formula, the uncertainty in the polar orientation appears to be somewhat worse. Our computational scheme is straightforward in view of the relative complexity of the shape variations of the polarization PSFs. The effective Hermite PSF model works well, and can be efficiently used in a MLE-fit for the unknown parameters. The Hermite PSF model provides significant analytical shortcuts for the integration over the pixel area and for the derivatives of the pixel Poisson rates with respect to the fit parameters.

A practical problem not considered here is the registration problem. The four images for the four polarization channels must be aligned with respect to each other with an accuracy typically better than 10 nm. Although this can be accomplished in principle with fiducials such as fluorescent beads, it requires careful experimentation and image analysis. Another aspect not considered here is the excess noise of the EMCCD. According to Mortensen et al. [17] MLE-fitting with a Poissonian noise based log-likelihood gives the correct results, only the predicted uncertainty is a factor $\sqrt{2}$ larger due to the noise associated with the electron multiplication process. A final issue introducing inaccuracies beyond the effects reported in this paper is birefringence of the sample. In case the emitter is located behind a layer of cell material, this birefringence may disturb the polarization of the emitted light, thus compromising the measurement of orientation and position.

An interesting topic for further study is the cross-over between fixed and free dipoles for coherent polarized illumination. In that case the ratio between the fluorescence lifetime and the rotational difusion time changes from zero to infinity. This will modify the PSF and the fitting procedure for position and orientation. Clearly, the uncertainty in determining the orientation will grow large upon approach of the free dipole case. The nature of this transition and a quantitative description of it are left for future work. It is noted that the ratio between the rotational diffusion time and the frame time is the relevant ratio if the illumination is unpolarized, e.g. if a lamp is used instead of a laser. A second topic of interest for future explorations is the generalization to the case of 3D-localization. It is not clear if the current approach with a simplified Hermite PSF model can be successfully applied to the bifocal [23, 24], astigmatic [25, 26], or helical [27] spot modifications needed for determining the axial position. Finally, an in-depth analysis of the effect of background on the uncertainty in position and orientation would provide a third topic worthy of investigation.

A. Summary of dipole image formation

The emitter dipole is located in a medium with refractive index n_{med} adjacent to a cover slip with refractive index n_{cov}. The emitter is imaged with an immersion objective lens with numerical aperture NA_{ob} designed for an immersion fluid with refractive index n_{imm}. The objective lens is assumed to be corrected for focusing onto the interface between the cover slip and the medium. The intersection of the optical axis with this interface is taken to be the origin of the coordinate system in object space. The emitted radiation is collected by the objective lens with focal length F_{ob} and focused by the tube lens with focal length F_{im} onto the detector. Both lenses are assumed to be aplanatic and the imaging system is assumed to be telecentric, i.e. the aperture stop is located at the back focal plane of the objective lens, which coincides with the front focal plane of the tube lens. The magnification of the imaging system is $M = F_{im}/F_{ob}$, thus making the numerical aperture in image space $NA_{im} = NA_{ob}/M$. In practice $M \gg 1$, so that $NA_{im} \ll 1$. The radius of the stop is given by $R = F_{ob}NA_{ob} = F_{im}NA_{im}$. The pupil coordinates are scaled with the stop radius R scaling the pupil to the unit circle, the object and image coordinates are scaled with the diffraction lengths λ/NA_{ob} (object space) and λ/NA_{im} (image space). The (2D) position of the emitter with respect to the focal point is $\vec{r}_d = (x_d, y_d)$. The scaled position is $\vec{u}_d = NA_{ob}\vec{r}_d/\lambda$. The electric field in the pupil plane at point $\vec{v} = (v_x, v_y, 0)$ corresponds to the plane wave in object space with wavevector along $(\sin\theta_{med}\cos\phi, \sin\theta_{med}\sin\phi, \cos\theta_{med})$,

so $v_x = n_{\text{med}} \sin\theta_{\text{med}} \cos\phi / NA_{\text{ob}}$ and $v_y = n_{\text{med}} \sin\theta_{\text{med}} \sin\phi / NA_{\text{ob}}$.

The electric field component $j = x,y$ proportional to dipole component $k = x,y,z$ is given by:

$$w_{jk}(\vec{u}) = \frac{1}{\pi} \int d^2 v\, C(\vec{v}) \frac{q_{jk}(\vec{v})}{\left(1 - v^2 \text{NA}_{\text{ob}}^2 / n_{\text{med}}^2\right)^{1/4}} \exp(-2\pi i \vec{u} \cdot \vec{v}), \tag{34}$$

where $C(\vec{v})$ is the pupil function ($C(\vec{v}) = 0$ for $|\vec{v}| > 1$). The polarization vectors are defined by:

$$
\begin{aligned}
q_{xk} &= \cos\phi\, T_p p_k - \sin\phi\, T_s s_k, & (35)\\
q_{yk} &= \sin\phi\, T_p p_k + \cos\phi\, T_s s_k, & (36)
\end{aligned}
$$

which depend on the p and s basis polarization vectors $\vec{p} = (\cos\theta_{\text{med}}\cos\phi, \cos\theta_{\text{med}}\sin\phi, -\sin\theta_{\text{med}})$ and $\vec{s} = (-\sin\phi, \cos\phi, 0)$ and on the Fresnel coefficients $T_a = T_{a,\text{med}-\text{cov}} T_{a,\text{cov}-\text{imm}}$ for $a = p,s$ and where the Fresnel coefficients for the two contributing interfaces are defined by $T_{a,1-2} = 2c_{a,1}/(c_{a,1} + c_{a,2})$ for $a = p,s$ and with $c_{p,l} = n_l/\cos\theta_l$ and $c_{s,l} = n_l \cos\theta_l$ for $l = \text{med}, \text{cov}, \text{imm}$.

In case of azimuthal symmetry in the pupil function (only defocus and/or spherical aberration) the functions $w_{jk}(\vec{u})$ can be written as:

$$
\begin{aligned}
w_{xx}(u,\psi) &= F_0(u) + F_2(u)\cos(2\psi), & (37)\\
w_{xy}(u,\psi) &= F_2(u)\cos(2\psi), & (38)\\
w_{xz}(u,\psi) &= iF_1(u)\cos\psi, & (39)\\
w_{yx}(u,\psi) &= F_2(u)\cos(2\psi), & (40)\\
w_{yy}(u,\psi) &= F_0(u) - F_2(u)\cos(2\psi), & (41)\\
w_{yz}(u,\psi) &= iF_1(u)\sin\psi, & (42)
\end{aligned}
$$

with:

$$F_0(u) = 2\int_0^1 dv \frac{v\left(T_s + T_p\sqrt{1 - v^2\text{NA}_{\text{ob}}^2/n_{\text{med}}^2}\right)}{2\left(1 - v^2\text{NA}_{\text{ob}}^2/n_{\text{med}}^2\right)^{1/4}} J_0(2\pi uv)\exp(iW(v)), \tag{43}$$

$$F_1(u) = 2\int_0^1 dv \frac{v^2 T_p \text{NA}_{\text{ob}}/n_{\text{med}}}{\left(1 - v^2\text{NA}_{\text{ob}}^2/n_{\text{med}}^2\right)^{1/4}} J_1(2\pi uv)\exp(iW(v)), \tag{44}$$

$$F_2(u) = 2\int_0^1 dv \frac{v\left(T_s - T_p\sqrt{1 - v^2\text{NA}_{\text{ob}}^2/n_{\text{med}}^2}\right)}{2\left(1 - v^2\text{NA}_{\text{ob}}^2/n_{\text{med}}^2\right)^{1/4}} J_2(2\pi uv)\exp(iW(v)). \tag{45}$$

B. Derivatives of the Poisson rate

The first order derivatives of the Poisson rates μ_k^x and μ_k^y of pixel k w.r.t. the parameters θ are:

$$
\begin{aligned}
\frac{\partial \mu_k^x}{\partial x_0} &= \frac{1}{\sqrt{2}\pi\sigma}\left[\left(D_{x'}^2 + \frac{1}{2}D_z^2\right)\Sigma_{10} + \alpha D_{x'}D_z(\cos\chi\Sigma_{20} + \sin\chi\Sigma_{11})\right.\\
&\quad \left. + \frac{1}{4}D_z^2\left(\cos^2\chi\Sigma_{30} + 2\sin\chi\cos\chi\Sigma_{21} + \sin^2\chi\Sigma_{12}\right)\right], \tag{46}\\
\frac{\partial \mu_k^y}{\partial x_0} &= \frac{1}{\sqrt{2}\pi\sigma}\left[\left(D_{y'}^2 + \frac{1}{2}D_z^2\right)\Sigma_{10} + \alpha D_{y'}D_z(-\sin\chi\Sigma_{20} + \cos\chi\Sigma_{11})\right.
\end{aligned}
$$

$$+\frac{1}{4}D_z^2\left(\sin^2\chi\Sigma_{30}-2\sin\chi\cos\chi\Sigma_{21}+\cos^2\chi\Sigma_{12}\right)\Bigg],\tag{47}$$

$$\frac{\partial\mu_k^x}{\partial y_0}=\frac{1}{\sqrt{2}\pi\sigma}\left[\left(D_{x'}^2+\frac{1}{2}D_z^2\right)\Sigma_{01}+\alpha D_{x'}D_z\left(\cos\chi\Sigma_{11}+\sin\chi\Sigma_{02}\right)\right.$$
$$\left.+\frac{1}{4}D_z^2\left(\cos^2\chi\Sigma_{21}+2\sin\chi\cos\chi\Sigma_{12}+\sin^2\chi\Sigma_{03}\right)\right],\tag{48}$$

$$\frac{\partial\mu_k^y}{\partial y_0}=\frac{1}{\sqrt{2}\pi\sigma}\left[\left(D_{y'}^2+\frac{1}{2}D_z^2\right)\Sigma_{01}+\alpha D_{y'}D_z\left(-\sin\chi\Sigma_{11}+\cos\chi\Sigma_{02}\right)\right.$$
$$\left.+\frac{1}{4}D_z^2\left(\sin^2\chi\Sigma_{21}-2\sin\chi\cos\chi\Sigma_{12}+\cos^2\chi\Sigma_{03}\right)\right],\tag{49}$$

$$\frac{\partial\mu_k^x}{\partial\sigma}=\frac{1}{\pi\sigma}\left[\left(D_{x'}^2+\frac{1}{2}D_z^2\right)\Lambda_{00}+\alpha D_{x'}D_z\left(\cos\chi\Lambda_{10}+\sin\chi\Lambda_{01}\right)\right.$$
$$\left.+\frac{1}{4}D_z^2\left(\cos^2\chi\Lambda_{20}+2\sin\chi\cos\chi\Lambda_{11}+\sin^2\chi\Lambda_{02}\right)\right],\tag{50}$$

$$\frac{\partial\mu_k^y}{\partial\sigma}=\frac{1}{\pi\sigma}\left[\left(D_{y'}^2+\frac{1}{2}D_z^2\right)\Lambda_{00}+\alpha D_{y'}D_z\left(-\sin\chi\Lambda_{10}+\cos\chi\Lambda_{01}\right)\right.$$
$$\left.+\frac{1}{4}D_z^2\left(\sin^2\chi\Lambda_{20}-2\sin\chi\cos\chi\Lambda_{11}+\cos^2\chi\Lambda_{02}\right)\right],\tag{51}$$

$$\frac{\partial\mu_k^x}{\partial D_x}=\frac{\cos\chi}{\pi}\left[2D_{x'}\Sigma_{00}+\alpha D_z\left(\cos\chi\Sigma_{10}+\sin\chi\Sigma_{01}\right)\right],\tag{52}$$

$$\frac{\partial\mu_k^y}{\partial D_x}=-\frac{\sin\chi}{\pi}\left[2D_{y'}\Sigma_{00}+\alpha D_z\left(-\sin\chi\Sigma_{10}+\cos\chi\Sigma_{01}\right)\right],\tag{53}$$

$$\frac{\partial\mu_k^x}{\partial D_y}=\frac{\sin\chi}{\pi}\left[2D_{x'}\Sigma_{00}+\alpha D_z\left(\cos\chi\Sigma_{10}+\sin\chi\Sigma_{01}\right)\right],\tag{54}$$

$$\frac{\partial\mu_k^y}{\partial D_y}=\frac{\cos\chi}{\pi}\left[2D_{y'}\Sigma_{00}+\alpha D_z\left(-\sin\chi\Sigma_{10}+\cos\chi\Sigma_{01}\right)\right],\tag{55}$$

$$\frac{\partial\mu_k^x}{\partial D_z}=\frac{1}{\pi}\left[D_z\Sigma_{00}+\alpha D_{x'}\left(\cos\chi\Sigma_{10}+\sin\chi\Sigma_{01}\right)\right.$$
$$\left.+\frac{1}{2}D_z\left(\cos^2\chi\Sigma_{20}+2\sin\chi\cos\chi\Sigma_{11}+\sin^2\chi\Sigma_{02}\right)\right],\tag{56}$$

$$\frac{\partial\mu_k^y}{\partial D_z}=\frac{1}{\pi}\left[D_z\Sigma_{00}+\alpha D_{y'}\left(-\sin\chi\Sigma_{10}+\cos\chi\Sigma_{01}\right)\right.$$
$$\left.+\frac{1}{2}D_z\left(\sin^2\chi\Sigma_{20}-2\sin\chi\cos\chi\Sigma_{11}+\cos^2\chi\Sigma_{02}\right)\right],\tag{57}$$

$$\frac{\partial\mu_k^x}{\partial\alpha}=\frac{1}{\pi}D_{x'}D_z\left(\cos\chi\Sigma_{10}+\sin\chi\Sigma_{01}\right),\tag{58}$$

$$\frac{\partial\mu_k^y}{\partial\alpha}=\frac{1}{\pi}D_{y'}D_z\left(-\sin\chi\Sigma_{10}+\cos\chi\Sigma_{01}\right),\tag{59}$$

$$\frac{\partial\mu_k^x}{\partial b}=1,\tag{60}$$

$$\frac{\partial\mu_k^y}{\partial b}=1,\tag{61}$$

where the arguments of the Σ_{lm} are suppressed for the sake of brevity, and where:

$$
\begin{aligned}
\Lambda_{lm} &= \Delta\left(\frac{x_k - x_0}{\sqrt{2}\sigma}\psi_{l+1}\left(\frac{x_k - x_0}{\sqrt{2}\sigma}\right)\right)\Delta\left(\psi_m\left(\frac{y_k - y_0}{\sqrt{2}\sigma}\right)\right) \\
&\quad + \Delta\left(\psi_l\left(\frac{x_k - x_0}{\sqrt{2}\sigma}\right)\right)\Delta\left(\frac{y_k - y_0}{\sqrt{2}\sigma}\psi_{m+1}\left(\frac{y_k - y_0}{\sqrt{2}\sigma}\right)\right).
\end{aligned}
\tag{62}
$$

In evaluating these derivatives it is used that:

$$
\frac{\partial}{\partial x_0}\Delta\psi_l\left(\frac{x_k - x_0}{\sqrt{2}\sigma}\right) = \frac{1}{\sqrt{2}\sigma}\Delta\psi_{l+1}\left(\frac{x_k - x_0}{\sqrt{2}\sigma}\right),
\tag{63}
$$

$$
\frac{\partial}{\partial\sigma}\Delta\psi_l\left(\frac{x_k - x_0}{\sqrt{2}\sigma}\right) = \frac{1}{\sigma}\Delta\left(\frac{x_k - x_0}{\sqrt{2}\sigma}\psi_{l+1}\left(\frac{x_k - x_0}{\sqrt{2}\sigma}\right)\right),
\tag{64}
$$

and similar expressions for the analogous derivatives.

C. Initial value estimation by the method of moments

Initial vlaues for the parameters are obtained from the lowest (up to second) order moments of the light distribution across the ROI. This method generalizes the centroid estimation of emitter location, which works well provided that the PSF is accurately described by a Gaussian, and provided there is no or only little background. It may therefore be expected that the generalization presented here also works well in the low-background limit. We will first describe how to extract estimates from each of the two pairs (with and without $\pi/4$ polarization rotation) of polarized images. Afterwards the two sets of values can be combined to the final estimate.

The Hermite PSF-model of Eqs. (11) and (12) gives rise to zeroth order moments (photon counts in the x and y polarization channels) as given by Eqs.(14) and (15). For the calculation of the first and second order moments it appears to be convenient to work with coordinates in a rotated frame, i.e. using $x' = x\cos\chi + y\sin\chi$ and $y' = -x\sin\chi + y\cos\chi$ and similar definitions for x_0' and y_0'. The first order moments are:

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy\, PSF_x(x,y)x' = \left(D_x^2 + \frac{1}{2}D_z^2\right)x_0' + \frac{\alpha}{\sqrt{2}}D_xD_z\sigma,
\tag{65}
$$

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy\, PSF_x(x,y)y' = \left(D_x^2 + \frac{1}{2}D_z^2\right)y_0',
\tag{66}
$$

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy\, PSF_y(x,y)x' = \left(D_y^2 + \frac{1}{2}D_z^2\right)x_0',
\tag{67}
$$

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy\, PSF_y(x,y)y' = \left(D_y^2 + \frac{1}{2}D_z^2\right)y_0' + \frac{\alpha}{\sqrt{2}}D_yD_z\sigma.
\tag{68}
$$

Clearly, the offset term does not affect the first order moment of x' in the y-polarization channel and the first order moment of y' in the x-polarization channel. The second order moments are:

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy\, PSF_x(x,y)x'^2 = \left(D_x^2 + \frac{1}{2}D_z^2\right)\left(\sigma^2 + x_0'^2\right) + D_z^2\sigma^2 + \sqrt{2}\alpha D_xD_z\sigma x_0',
\tag{69}
$$

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy\, PSF_x(x,y)y'^2 = \left(D_x^2 + \frac{1}{2}D_z^2\right)\left(\sigma^2 + y_0'^2\right),
\tag{70}
$$

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy\, PSF_y(x,y)x'^2 = \left(D_y^2 + \frac{1}{2}D_z^2\right)\left(\sigma^2 + x_0'^2\right),
\tag{71}
$$

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy\, PSF_y(x,y)y'^2 = \left(D_y^2 + \frac{1}{2}D_z^2\right)\left(\sigma^2 + y_0'^2\right) + D_z^2\sigma^2 + \sqrt{2}\alpha D_yD_z\sigma y_0'.
\tag{72}
$$

The moments of $x'y'$ turn out to be zero in the model.

This set of lowest order moments are estimated from the pixel values n_k^j for all pixels k in channel $j = x, y$ by calculating the sums:

$$M_0^j = \sum_k n_k^j, \tag{73}$$

$$\left(M_{1x}^j, M_{1y}^j\right) = \sum_k n_k^j (x_k, y_k), \tag{74}$$

$$\left(M_{2x}^j, M_{2y}^j\right) = \sum_k n_k^j \left(x_k^2, y_k^2\right). \tag{75}$$

For the sake of clarity we will first describe how to find values in case of zero offset $\alpha = 0$. In that case the zeroth and first order moments suffice for finding the emitter coordinates (centroid estimate):

$$x'_{0v} = \frac{M_{1x}^x + M_{1x}^y}{M_0^x + M_0^y}, \tag{76}$$

$$y'_{0v} = \frac{M_{1y}^x + M_{1y}^y}{M_0^x + M_0^y}. \tag{77}$$

In the case of non-zero offset $\alpha \neq 0$ this estimate is biased but still has minimum variance. We will henceforth refer to this estimate as the minimum-variance estimate with subscript 'v'. With the quantities:

$$A_{2x}^j = M_{2x}^j - 2M_{1x}^j x'_{0v} + M_0^j {x'_{0v}}^2, \tag{78}$$

$$A_{2y}^j = M_{2x}^j - 2M_{1x}^j y'_{0v} + M_0^j {y'_{0v}}^2, \tag{79}$$

an estimate for the spot width follows as:

$$\sigma^2 = \frac{A_{2x}^y + A_{2y}^x}{M_0^x + M_0^y}, \tag{80}$$

The vector component D_z can now be obtained from:

$$D_z^2 = \frac{A_{2x}^x + A_{2y}^y - A_{2x}^y - A_{2y}^x}{2\sigma^2}, \tag{81}$$

which allows to deduce the other two vector components from the zeroth order moments by:

$$D_{x'}^2 = M_0^x - \frac{1}{2}D_z^2, \tag{82}$$

$$D_{y'}^2 = M_0^y - \frac{1}{2}D_z^2. \tag{83}$$

In case of a non-zero offset term $\alpha \neq 0$ the first and zeroth order moments give rise to a second estimate for the emitter position:

$$x'_{0b} = \frac{M_{1x}^y}{M_0^y}, \tag{84}$$

$$y'_{0b} = \frac{M_{1y}^x}{M_0^x}. \tag{85}$$

This estimate is unbiased, hence the subscript 'b', and at first sight seems preferable over the centroid estimate. However, the distribution of photons over the two channels can be rather uneven giving one of the two estimates a relatively large variance. Clearly, it is more sensible to look for an optimum bias-variance trade-off which has the lowest possible overall localization error. This optimum is found via a weighted sum of the two estimates:

$$x_0' = (1 - \varepsilon) x_{0b}' + \varepsilon x_{0v}', \tag{86}$$

with ε a weighting parameter (and a similar expression for the y-coordinate). It is noted that translation invariance requires that the linear combination of the two fit values takes this form with only a single degree of freedom in the weighting parameter. The optimum value of ε is determined from mimimizing the (x-contribution to the) Mean Square Localization Error MSLE, which is the sum of the variance V and the square of the bias Δ:

$$\text{MSLE} = \langle (x_0' - x_{\text{true}}')^2 \rangle = \langle (x_0' - \langle x_0' \rangle)^2 \rangle + (\langle x_0' \rangle - x_{\text{true}}')^2 = V + \Delta^2. \tag{87}$$

The estimated variances of the minimum-variance and minimum-bias fit values are $V = \sigma / \sqrt{M_0^x + M_0^y}$ and $V_y = \sigma / \sqrt{M_0^y}$, where an initial estimate for σ is given by (80). The bias is estimated as $\Delta = \kappa (x_{0v} - x_{0b})$ with κ a parameter describing the confidence level of the bias estimate. It follows that the MSLE is given by:

$$\text{MSLE} = (1 - \varepsilon)^2 V_y + \varepsilon^2 V + \varepsilon^2 \Delta^2. \tag{88}$$

The miminum of this expression is obtained for:

$$\varepsilon = \frac{V_y}{V + V_y + \Delta^2}, \tag{89}$$

and the minimum MSLE is equal to:

$$\text{MSLE} = \frac{V_y (V + \Delta^2)}{V + V_y + \Delta^2}. \tag{90}$$

For vertical dipoles or horizontal dipoles with azimuthal angle in between the polarizer axes we typically expect $\Delta^2 \approx V_y \approx 2V$ giving $\text{MSLE} \approx V$. For tilted dipoles we can have $\Delta^2 \gg V_y \approx 2V$ giving $\text{MSLE} \approx 2V$. For horizontal dipoles with azimuthal angle close to the orthogonal axis we expect $V_y \gg V$ leading to $\text{MSLE} \approx V + \Delta^2$. For this last case a near zero bias confidence parameter κ is needed to keep Δ small in all cases. On the other hand it must be equal to unity for tilted dipoles with azimuthal angle in between the polarizer axes. A suitable choice is therefore:

$$\kappa = \frac{4 M_0^x M_0^y}{(M_0^x + M_0^y)^2}. \tag{91}$$

With this choice the worst case (tilted dipoles with large bias) gives a localization accuracy that is a factor $\sqrt{2}$ worse than the ideal $\sqrt{V}$ value.

Improved estimates can be found for the other parameters now that the position of the emitter is estimated with mimimum MSLE. To that end, the minimum variance estimate in Eqs. (78) and (79) must be replaced by the minimum localization error estimate. The equations for estimating the spot width and the vector components $(D_{x'}, D_{y'}, D_z)$ remain the same. Finally, the asymmetry coefficient α is obtained from:

$$\alpha = \sqrt{2} \frac{D_{x'} \left(M_{1x}^x - M_0^x x_0' \right) + D_{y'} \left(M_{1y}^y - M_0^y y_0' \right)}{\left(D_{x'}^2 + D_{y'}^2 \right) D_z \sigma}. \tag{92}$$

There is a connection between the method of moments presented here and the orthogonality relations satisfied by the Hermite polynomials:

$$\int_{-\infty}^{\infty} du \, H_n(u) H_m(u) \exp\left(-u^2\right) = \sqrt{\pi} 2^n n! \delta_{nm}. \tag{93}$$

This orthogonality implies that the coefficients for the expansion of a spot in Hermite functions can be found by integrating the product of the spot and the particular Hermite polynomial over the image coordinates. Such integrals can be expressed as linear combinations of the moments of the spot. Clearly, the Hermite expansion coefficients, and hence the model parameters, can be directly expressed in terms of the moments.