Calibration-free estimation of field dependent aberrations for single molecule localization microscopy across large fields of view

Isabel Droste[1], Erik Schuitema[2], Sajjad Khan[3], Stijn Heldens[4], Ben van Werkhoven[4], Keith A. Lidke[3], Sjoerd Stallinga[1*†], and Bernd Rieger[1*†]

1 Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands

2 ESSD Software Engineering, Rotterdam, The Netherlands

3 Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico, USA

4 Netherlands eScience Center, Amsterdam, The Netherlands.

* equal contribution

†**For correspondence:** s.stallinga@tudelft.nl (S.S.), b.rieger@tudelft.nl (B.R.)

Abstract

Image quality in single molecule localization microscopy (SMLM) depends largely on the accuracy and precision of the localizations. While under ideal imaging conditions the theoretically obtainable precision and accuracy are achieved, in practice this changes if (field dependent) aberrations are present. Currently there is no simple way to measure and incorporate these aberrations into the Point Spread Function (PSF) fitting, therefore the aberrations are often taken constant or neglected all together. Here we introduce a model-based approach to estimate the field-dependent aberration directly from single molecule data without a calibration step. This is made possible by using nodal aberration theory to incorporate the field-dependency of aberrations into our fully vectorial PSF model. This results in a limited set of aberration fit parameters that can be extracted from the raw frames without a bead calibration measurement, also in retrospect. The software implementation is computationally efficient, enabling fitting of a full 2D or 3D dataset within a few minutes. We demonstrate our method on 2D and 3D localization data of microtubuli and nuclear pore complexes over fields of view (FOV) of up to 180 μm and compare it with spline-based fitting and a deep learning based approach.

Main text

Single-molecule localization microscopy (SMLM) [Betzig2006, Rust2006, Sharonov2006, Heilemann2008] enables routine imaging at the nano-scale. The need for high throughput and much data implies that a large Field Of View (FOV) must be imaged, which is enabled by modern sCMOS sensors that have up to 10 Mpixels and advanced illumination schemes [Douglass2016, Deschamps2016, Diekmann2017, Archetti2019, Mau2021, Li2022, Ma2024, Nelson2024]. A key problem that arises is that aberrations depend on the position in the FOV. This is problematic for both the precision and accuracy of in-plane (xy) localization, but even more so for 3D (xyz) localization. Up to now aberrations, let alone field dependent aberrations, are rarely considered in SMLM as it is time consuming and cumbersome to i) measure them and ii) include them into the fitting. For that reason, the default method for estimating the positions of the emitters is to fit a simplified Point Spread Function (PSF) model to the recorded data. Typically, this simplified model is a Gaussian [Smith2010], which reduces the computational load to estimate the positions of millions of emitters to reconstruct a super-resolution image. Known drawbacks, however, are position biases in case of asymmetric aberrations and of emitters with partially fixed orientation [Stallinga2010], and an underestimation of the photon count [Thorsen2018]. An actual data driven PSF model is the spline model [Li2019], which requires calibration of the PSF by beads, potentially at various field and depth positions.

1

Advantageously, a physically correct fully vectorial PSF model could be used, so that high NA and polarization effects are properly considered. Moreover, (field dependent) aberrations of the microscope can then automatically be incorporated, and emission dipole orientation as well, if needed. Up to now such a full vectorial model is not used in practice for SMLM due to the computational load [Hullemann2021].

The current practical standard to measure field dependent aberrations is a calibration in the form of an axial scan of many fluorescent beads distributed over the whole FOV [Li2019, Hulleman2021, Fu2023]. An alternative is provided by the localization data itself, as millions of localization events are in fact millions of measurements of the microscope's PSF across the FOV. The localization data itself thus provides a wealth of information on the optical system, and the challenge is to unlock that information. The key obstacle is that each individual localization event does not allow for fitting of aberration parameters next to the position, photon count and background, because the data is much too noisy for a robust estimation of many parameters. Previously, Xu et al. applied pupil phase retrieval to an averaged set of 3D localization data, but only so using an approximate scalar PSF model and for global aberrations independent of the position in the FOV [Xu2020]. Recently, Liu et al. proposed a framework for inverse PSF modelling, enabled by automatic differentiation, and showed its applicability to vectorial PSF fitting of 3D localization datasets with field dependent aberrations [Liu2024]. Here, we present an alternative method to estimate the field dependent aberrations from 2D or 3D single molecule images alone, without any calibration measurement. We make this possible by fitting the localization data with a global optical aberration model, with a limited number of fit parameters, instead of fitting the full aberration content for each measured PSF separately. This global aberration model is derived from the so-called Nodal Aberration Theory (NAT) [Shack1980], which provides explicit relations of the field dependency of the appearing aberration coefficients via low-order polynomials of the field coordinates. This eliminates the need for ad-hoc smoothing of field dependent aberrations as done in the inverse PSF modeling approach and enables application to 2D localization data.

Aberrations can be modelled by expressing the phase aberration function in the pupil plane in terms of Zernike modes. Using NAT, the Zernike coefficients A_{nm} can be expressed as low-order polynomials of the field coordinates (x, y) (see Supplementary Note for theory):

$$A_{nm}(x,y) = \sum_{l,q} \gamma_{lqnm} Q_{lq}(x,y),$$

where γ_{lqnm} are the NAT coefficients, and $Q_{lq}(x,y)$ are products of (Legendre) polynomials of order l and q. We adapted NAT to the square shape of the FOV by using Legendre polynomials to describe the field dependence. For describing the field dependency of the first order aberrations, that is, defocus, and primary astigmatism, coma, and spherical aberration, it turns out that only 13 NAT coefficients are needed. To include second order aberrations (trefoil and secondary astigmatism, coma, and spherical aberration), an additional 43 NAT coefficients are needed. This number of free parameters, though large, is still orders of magnitude less than the number of free parameters in data driven approaches [Li2018, Bates2022].

Figure 1 schematically depicts our method. The raw blinking time series data is segmented into regions of interests (ROI) containing fluorescent emitters as usual for SMLM. From these millions of ROIs, a very small subset of $M_s \sim 10^3$ ROIs across the FOV is randomly selected to be used as input for the

field dependent aberration estimation. The estimation process consists out of an alternation of local and global updates. During a global update, the NAT coefficients are updated while keeping the locations, photon count and local background of the emitters fixed, while during a local update, the locations are updated with fixed NAT coefficients. Subsequently, all ROIs are fitted with a vector PSF model ("Vectorfit", https://gitlab.tudelft.nl/imphys/ci/vectorfit) using the found field dependent aberrations as input. We have implemented the vector PSF fitting on GPU for speedup and have devised several algorithmic improvements for additional steps in efficiency. First, we utilize the phasor method to provide a fast, initial estimate of the lateral position that is robust against high background [Martens2018]. Second, we compute the initial estimate of photon count N_{ph} and background count bg with linear regression given the data and the model. Third, pre-computing the Optical Transfer Function (OTF) reduces the number of needed Fourier Transformations per iteration by a factor of 6. These different algorithmic innovations are outlined in the Methods section.

We have tested our method on SMLM data of microtubuli acquired over a large 97 x 97 µm FOV (see Methods for experimental details). The same sample was imaged with and without astigmatism to obtain 2D and 3D localization data, respectively (see Figure 2). The aberrations found from the single molecule data via our method give very similar aberration maps across the FOV (orange in Fig. 2a,e) as the aberration maps obtained by interpolating between aberrations retrieved from z-stack bead calibration data (light blue in Fig. 2a,e), with the largest difference in the slope of the coma aberration maps. We performed a chi-square goodness of fit test [Siemons2018] to check if the PSF model with single molecule derived aberrations gives rise to a better fit than the PSF model with bead calibration derived aberrations (Supplementary Figure 1). For both the 2D and 3D data we find on average about 1% smaller chi-square values for the single molecule derived aberrations, up to about 5% and 3% smaller at the edges of the FOV for the 2D and 3D datasets, respectively, indicating a slightly better fit of the single molecule derived aberrations. The overall chi square values are about 20% (2D) and 16% (3D) higher than the theoretical values based on shot noise statistics [Siemons2018] (Supplementary Figure 3), indicating a residual model mismatch, possibly due to higher order aberrations and amplitude aberrations. A further validation of our approach is found by a comparison of the modelled PSF to the measured single molecule spots in 4x4 patches across the FOV, as a function of the axial position (Supplementary Video 1). This comparison shows that the modelled PSF matches very well with the measured data.

We also found a difference in the estimated bead aberrations between consecutive days of measurement with different bead samples. The beads measured on the second day show a uniform shift in coma coefficient A_{3-1} (Supplementary Figure 2). By estimating the z-locations of the beads, we found that the tilt of the sample was different between the two measurement instances, most likely caused by remounting of a different bead sample on the stage. We attribute the difference in the estimated coma to this sample tilt, as passing a high NA emission beam through a tilted medium leads to coma that is constant over the FOV. Apparently, the aberration estimation from bead calibration data can be affected by small changes in the imaging conditions, like the sample tilt. Similar differences in imaging conditions can potentially arise between the bead calibration measurement and the actual single molecule data acquisition, implying that estimation of aberrations from the data itself not only makes the bead calibration measurement spurious, but is also inherently more reliable.

We also used the chi square goodness of fit test to assess the added value of taking the field dependency of the aberrations into account. We applied Vectorfit with all the aberrations set to zero, except for a constant amount of astigmatism ($A_{2-2} = 87\ m\lambda$, equal to the average amount over the FOV estimated from the bead calibration) for estimating the axial location for the 3D data. Supplementary Figure 3 shows that the PSF model with single molecule data derived field dependent aberrations fits the data better than the model without aberrations, especially towards the edges of the FOV. A comparison of the estimated locations using the two models shows that the location difference increases with the distance from the center of the FOV (Supplementary Fig. 3d). Moreover, for the 3D data, the shift in location is different for emitters with z < 0 than for emitters with z > 0 (Supplementary Fig. 3f), which could lead to a distortion of the imaged structures when field dependent aberrations are not considered.

We compared our reconstructions to existing methods, namely Picasso for the 2D data and SMAP for the 3D data. Picasso [Picasso] uses a Gaussian PSF model where the horizontal and vertical standard deviations σ_x and σ_y are estimated for each ROI [Smith2010], whereas SMAP [SMAP] uses a spline PSF model that is uniform across the FOV, with parameters estimated from the experimental bead data. Figures 2c,d show that 2D reconstructions with Picasso are very similar to reconstructions with Vectorfit. The average localization precision, estimated by linking localizations in consecutive frames [Cnossen2020], is 10.0 nm for Vectorfit with a CRLB of 9.3 nm, better than the value of 13.5 nm found for Picasso with a CRLB of 14.2 nm. Figures 2g-j show that 3D reconstructions with SMAP are similar to reconstructions with Vectorfit, where Vectorfit shows a somewhat smaller axial spread of the localizations. The linking analysis gives here an average localization precision of 12.1 nm (xy) and 29.9 nm (z), with a CRLB of 12.1 nm (xy) and 31.1 nm (z), compared to a precision of 12.1 nm (xy) and 35.7 nm (z) for SMAP. FRC image resolution [Nieuwenhuizen2013] estimation (Supplementary Figs. 4 and 5) show a comparable performance between Vectorfit and the other used softwares.

Next, we applied our method to 3D data of NUP96 in the nuclear pore complex (NPC) in U2OS cells (FOV of about 180x180 µm) of Fu et al. [Fu2023], where the axial direction was encoded using astigmatism, and compared the outcome to that of their method FD-DeepLoc in Figure 3. The aberrations maps retrieved from the single molecule data matches well with the aberration maps obtained from the bead calibration used in FD-DeepLoc, towards the edges of the FOV, however, the bead calibration derived aberration maps show large and erratic variations, which is unlikely from the optical point of view. Figure 3b shows the experimentally found standard deviation for the estimation of the different aberrations, obtained from 30 estimations for different randomly selected subsets of 5,000 localizations, as well as their Cramer-Rao lower Bound (CRLB) values. The experimental standard deviations are on the small $m\lambda$ scale across the whole FOV, but do not reach the CRLB values in the sub-$m\lambda$ level. We also tested the dependence of the experimental precision on the number of localization events M_s, and found a scaling inversely proportional to $M_s^{1/2}$, in agreement with expectations (Supplementary Figs. 6 and 7). The comparison of the modelled PSF to the measured single molecule spots in 4x4 patches across the FOV, as a function of the axial position (Supplementary Video 2) shows a good match between the modelled PSF with the measured data. These validation tests indicate that our approach enables unbiased and precise aberration estimation.

We compared the Vectorfit reconstruction with field dependent aberrations retrieved from the single molecule data to a reconstruction with Vectorfit without aberrations, were all the aberrations are set

to zero, except for a constant amount of astigmatism ($A_{22} = 103\ m\lambda$, in accordance with the reported rms value from the cylindrical lens) for estimating the axial location, and to the FD-DeepLoc reconstruction. In the center of the FOV, close to the optical axis, the reconstructions from the three methods are comparable (Supplementary Fig. 8). Towards the edges of the FOV, however, the aberrations become larger, leading to sizeable differences in the axial localization (see Figure 3c-f). In one corner of the FOV, the FD-DeepLoc reconstruction shows a large spread of z-locations within a single cell nucleus (Fig. 3d(iii)). The xz cross-section in Fig. 3f(iii) shows a large variation of z-locations in a single NPC, which seems unrealistic. In our reconstructions (Fig. 3d(i-ii) and Fig. 3f(i-ii)), the range of z-locations matches much better with the expected thickness of the NPC ring of around 50 nm [Wang2023, Thevasthasa2019]. At the top of the FOV we observe that Vectorfit with field dependent aberrations leads to a better reconstruction than without aberrations. (Supplementary Figure 9). A chi-square goodness of fit test (Supplementary Figure 10) shows that the Vectorfit model with field dependent aberrations outperforms Vectorfit with constant aberrations, with a chi-square ratio reaching 1.5 at the edges of the FOV. We also observe a shift in the lateral localizations up to a few tens of nanometres (Supplementary Fig. 10b) and a shift in the axial localizations of up to 150 nanometres (Supplementary Fig. 10c). The distribution of localization precision values, estimated using the method of linking localizations in consecutive frames, has a mean of 11.7 nm (xy) and 23.5 nm (z) for Vectorfit with field dependent aberrations, with a CRLB of 10.1 nm (yz) and 20.6 nm (z). For FD-Deeploc we find by the same method a localization precision of 7.9 nm (xy) and 23.2 nm (z). With Vectorfit, we achieve the CRLB and therefore attain the optimal localization precision based on the signal-to-noise ratio of the data. The axial localization precision of Vectorfit is comparable to FD-Deeploc, but the lateral localization precision is a bit worse. We speculate that the learning-based approach of FD-Deeploc could bias localizations from the same emitter, which would introduce spurious correlations in subsequent time frames, resulting in a lower measured lateral localization precision than the CRLB. Analysis of the image resolution in a grid of 4x4 patches across the FOV with FRC thus shows a slightly better image resolution for FD-DeepLoc than for VectorFit (Supplementary Figure 11).

Fully vectorial PSF fitting in a Maximum Likelihood Estimation (MLE) framework has not found its way into the practice of the SMLM field up to now. Software fitting is mostly dominated by fitting of Gaussians or spline-based models. The reason is solely that the fitting speed was too slow for the typical number of localizations (~10^6) in a super-resolution acquisition. CPU implementations of 2D Gaussian MLE fitting achieve currently about 10^4 fits/s and GPU implementations even more than 10^6 fits/s (7x7 pixel ROI). Only the phasor-based approach is equally quick on CPU [Martens2018]. Our OTF based algorithmic improvements for the fitting together with a high-end GPU (Nvidia A100) allow now 10^5 fits/s (see Supplementary Table 1), while a good desktop GPU (Nvidia RTX A4000) has $6 \cdot 10^4$ fits/s and a multi-core CPU still achieves $9 \cdot 10^3$ fits/s. This brings vectorial PSF fitting times for a typical data set to less than a minute on a GPU in 2D. In 3D our algorithmic improvements are less effective (for details see Supplementary Note) which results in $6.8 \cdot 10^3$ fits/s for a 17x17 pixel ROI on our desktop GPU which is about 30x slower than 3D spline-based fitting which achieves $2 \cdot 10^5$ fits/s on a desktop GPU for a 17x17 pixel ROI [Li2024], but better than the less than 1 fit/second reported for a 17x17x41 pixel bead stack [Fu2023].

Our approach of estimating aberrations directly from single molecule data makes taking the inclusion of field dependent aberrations to the localization much more feasible because it saves doing an

additional bead calibration experiment, which can also be prone to errors as shown above. The recently proposed method from Liu et al. to estimate aberrations from single molecule data can take more than half an hour on a GPU [Liu2024], while our model-based approach takes less than a minute on a CPU to estimate aberrations. In addition our work is applicable to 2D localization data while theirs is not. In 2D, however, the effect of field dependent aberrations on the localizations is less impactful than in 3D.

In future work, a user interface should be developed to make vectorial PSF fitting accessible for a broader audience. Furthermore, the NAT driven approach of estimating field dependent aberrations directly from single molecule data could be applied to other imaging modalities such as 4Pi SMLM or localization microscopy with fixed dipole emitters, and also for scanning but camera-based localization microscopy [Radmacher2024].

Author contributions
I.D. developed the analysis model and analysed all datasets, E.S., S.H. and B.v.W. developed GPU code, S.K. and K.A.L. performed imaging of biological samples. S.S. and B.R. devised key concepts and supervised the study. S.S., B.R. and I.D. wrote the paper. All authors read and approved the manuscript.

Acknowledgements

Data and code availability
Code and a software demo are available at https://gitlab.tudelft.nl/imphys/ci/vectorfit.

References
- 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**:1642-1645, 2006.
- 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. *Angewandte Chemie*, **47**(33):6172-6176, 2008.
- Sharonov and R.M. Hochstrasser. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proceedings of the National Academy of Science USA*, **103**(50):18911-18916, 2006.
- M.J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**(10):793-795, 2006.
- K.M. Douglass, C.M. Sieben, A. Archetti, A. Lambert, and S. Manley. Super-resolution imaging of multiple cells by optimized flat-field epi-illumination. *Nature Photonics*, **10**:705-710, 2016.
- J. Deschamps, A. Rowald, and J. Ries. Efficient homogeneous illumination and optical sectioning for quantitative single-molecule localization microscopy. *Optics Express*, **24**(24):28080-28090, 2016.

- R. Diekmann, Ø.I. Helle, C.I. Øie, P. McCourt, T.R. Huser, M Schütpelz, and B.S. Ahluwalia. Chip-based wide field-of-view nanoscopy. *Nature Photonics*, 11:322--328, 2017.
- A. Archetti, E. Glushkov, C. Sieben, A. Stronganov, A. Radenovic, and S. Manley. Waveguide-PAINT offers an open platform for large field-of-view super-resolution imaging. *Nature Communications*, **10**:1267, 2019.
- A. Mau, K. Friedl, C. Leterrier, N. Bourg, and S. Lévêque-Fort. Fast widefield scan provides tunable and uniform illumination optimizing super-resolution microscopy on large fields. *Nature Communications* **12**:3077, 2021.
- Q. Li, C.N. Hulleman, R.J. Moerland, E. Mailvaganam, S. Ganapathy, D. Brinks, S. Stallinga, and B. Rieger. Waveguide-based total internal reflection fluorescence microscope enabling cellular imaging under cryogenic conditions. *Optics Express*, 29(21):34097-34108, 2021.
- H. Ma, M. Chen, P. Nguyen, and Y. Liu. Toward drift-free high-throughput nanoscopy through adaptive intersection maximization. *Science Advances* **10**(21):eadm7765, 2024.
- T. Nelson, S. Vargas-Hernández, M. Freire, S. Cheng, and A. Gustavsson, "Multimodal illumination platform for 3D single-molecule super-resolution imaging throughout mammalian cells," *Biomedical Optics Express* **15**:3050-3063, 2024.
- R.Ø. Thorsen, C.N. Hulleman, M. Hammer, D. Grünwald, S. Stallinga, and B. Rieger. Impact of optical aberrations on axial position determination by photometry. *Nature Methods*, **15**:989-993, 2018.
- S. Stallinga and B. Rieger. Accuracy of the Gaussian point spread function model in 2D localization microscopy. *Optics Express*, **18**(24):24461-24476, 2010.
- C.S. Smith, N. Joseph, B. Rieger, and K.A. Lidke. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nature Methods*, 7(5):373-375, 2010.
- Y. Li, M. Mund, P. Hoess, J. Deschamps, U. Matti, B. Nijmeijer, V. Jimenez Sabinian, J. Ellenberg, I. Schön, and J. Ries. Real-time 3D single molecule localization using experimental point spread functions. *Nature Methods*, **15**:367-369, 2018.
- C.N. Hulleman, R.Ø. Thorsen, E. Kim, C. Dekker, S. Stallinga, and B. Rieger. Simultaneous orientation and 3D localization microscopy with a Vortex point spread function. *Nature Communications*, **12**:5934, 2021.
- M. Siemons, C. N. Hulleman, R. Ø. Thorsen, C. S. Smith, and S. Stallinga, High precision wavefront control in point spread function engineering for single emitter localization, *Optics Express,* **26**:8397-8416, 2018.
- S. Fu, W. Shi, T. Luo, Y. He, L. Zhou, J. Yang, Z. Yang, J. Liu, X. Liu, Z. Guo, C. Yang, C. Liu, Z. Huang, J. Ries, M. Zhang, P. Xi, D. Jin, and Y. Li. Field-dependent deep learning enables high-throughput whole-cell 3D super-resolution imaging. *Nature Methods*, **20**:459-468, 2023.
- M. Li, W. Shi, S. Liu, S. Fu, Y. Fei, L. Zhou, K. Fang, and Y. Li, Fast and universal single-molecule localization using multi-dimensional point spread functions. *Optics Express,* **32**:42601-42615 , 2024.
- R.V. Shack and K. Thompson, Influence of alignment errors of a telescope system on its aberration field. *Optical Alignment I*, Vol. 0251, 146–153. International Society for Optics and Photonics (SPIE, 1980).
- K. Martens, A.N. Bader, S. Baas, B. Rieger, and J. Hohlbein. Phasor based single-molecule localization microscopy in 3D (pSMLM-3D): an algorithm for MHz localization rates using standard CPUs. *Journal of Chemical Physics*, **148**:123311, 2018

- [Picasso] https://github.com/jungmannlab/picasso
- [SMAP] https://github.com/jries/SMAP
- J. Cnossen, T. Hinsdale, R.Ø. Thorsen, M. Siemons, F. Schueder, R. Jungmann, C.S. Smith, B. Rieger, and S. Stallinga. Localization microscopy at doubled precision with patterned illumination. *Nature Methods,* **17**:59–63, 2020.
- R.P.J. Nieuwenhuizen, K.A. Lidke, M. Bates, D.L. Puig, D. Grünwald, S. Stallinga, and B. Rieger. Measuring image resolution in optical nanoscopy. *Nature Methods,* **10**:557–562, 2013.
- W. Wang, A. Jakobi, Y.-L. Wu, J. Ries, S. Stallinga, and B. Rieger. Particle fusion of super-resolution data reveals the unit structure of Nup96 in nuclear pore complex. *Scientific Reports*, **13**:13327, 2023.
- J.V. Thevasthasan, M. Kahnwald, K. Cieslinski, P. Hoess, S. Kumar Penetie, M. Reitberger, D. Heid, K. Chitanaya Kasuba, S. Janice Hoerner, Y. Li, Y.-L. Wu, M. Mund, U. Matti, P. Matos Pereira, R. Henriques, B. Nijmeijer, M. Kuehlbeck, V. Jimenez Sabinina, J. Ellenberg, and J. Ries. Nuclear pores as versatile reference standards for quantitative superresolution microscopy. *Nature Methods*, **16**:1045-1053, 2019.
- Xu, F., Ma, D., MacPherson, K.P. *et al.* Three-dimensional nanoscopy of whole cells and tissues with in situ point spread function retrieval. *Nature Methods* **17**, 531–540, 2020.
- Liu, S., Chen, J., Hellgoth, J. *et al.* Universal inverse modeling of point spread functions for SMLM localization and microscope characterization. *Nature Methods* **21**, 1082–1093, 2024.
- [Vectorfit] https://gitlab.tudelft.nl/imphys/ci/vectorfit
- N. Radmacher, O. Nevskyi, J. Ignacio Gallea, J.C. Thiele, I. Gregor, S.O. Rizzoli, and J. Enderlein. Doubling the resolution of fluorescence-lifetime single-molecule localization microscopy with image scanning microscopy. *Nature Photonics*, 18(10):1059-1066, 2024.
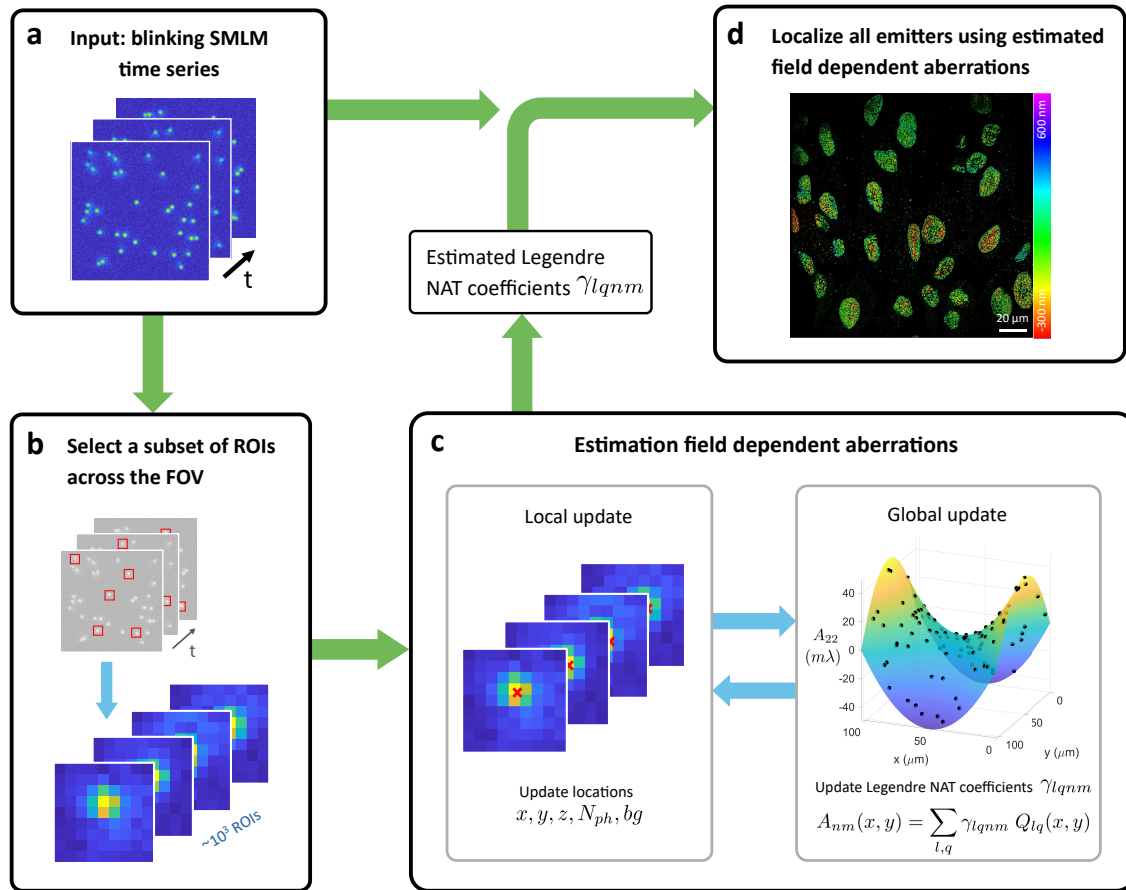
Fig. 1 | Schematic of fitting field dependent aberrations from single molecule data. a, The raw single molecule blinking frames are used as input. **b,** After segmentation into ROIs, a subset of ~10^3 ROIs across the FOV is selected for field dependent aberration estimation. **c,** The estimation of field dependent aberrations uses maximum likelihood estimation and consists of two optimization loops: a local update that updates the locations, photon count and background of each ROI while keeping the NAT coefficients constant, and a global update that updates the Legendre NAT coefficients while keeping the locations, photon count and background constant. By selecting different ROIs in **b** multiple times and repeating the estimation (**c**) an estimation precision for the aberrations can be calculated. **d,** Finally, in all ROIs emitters are localized using the vectorial PSF model (Vectorfit) with the estimated NAT coefficients as input.
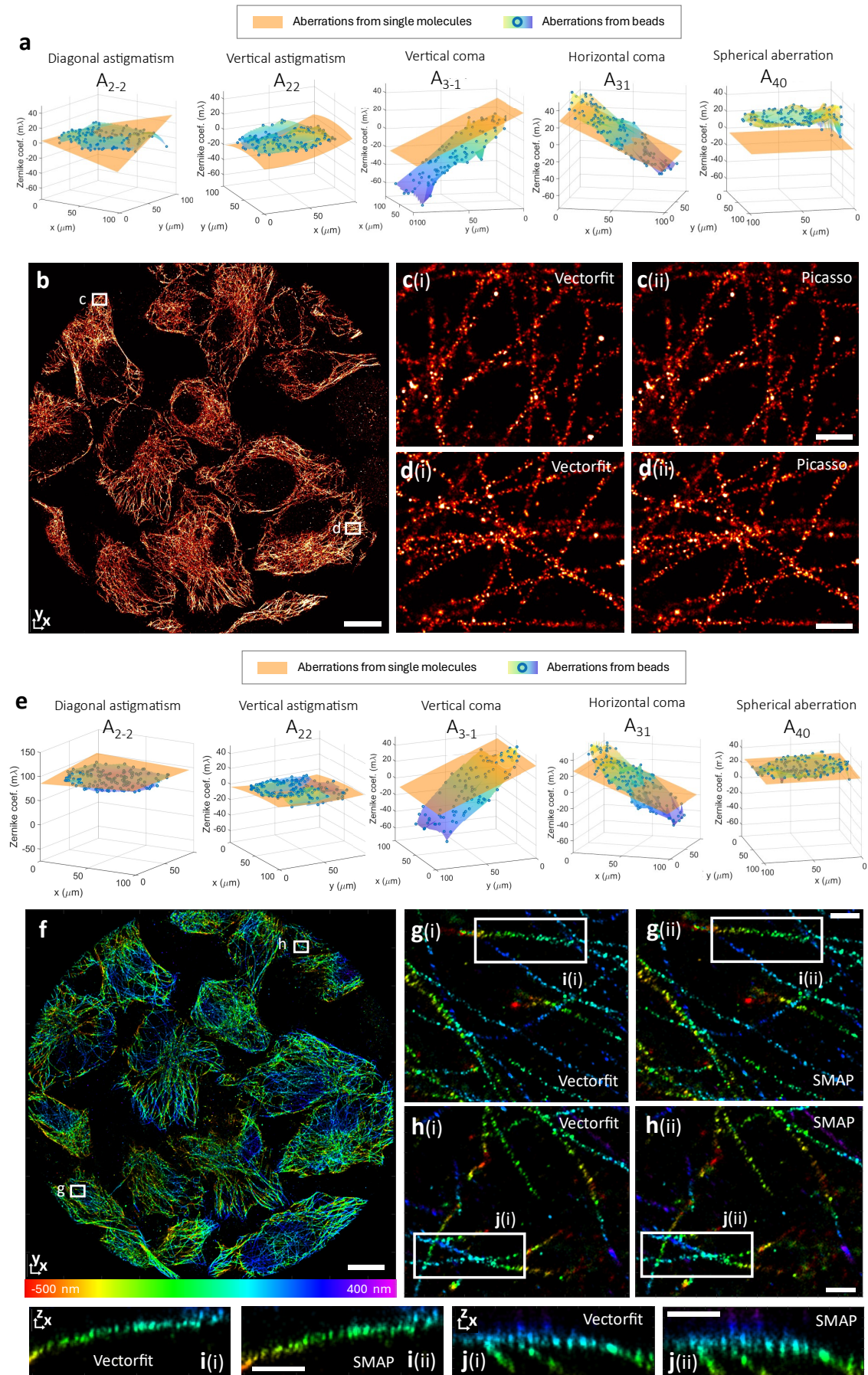
Fig. 2 | Aberration estimation from 2D and 3D experimental data of microtubili. A sample was imaging in 3D with and in 2D without astigmatism**. a,** Estimated Zernike aberration surfaces from single molecules of the 2D data, compared to interpolated aberrations from bead z-stack calibration. **b**, Full FOV image reconstructed using Vectorfit and fitted aberrations from single molecules. **c,** and **d,** insets of **b** where (i) shows vector fitting and (ii) Gaussian fitting by Picasso [Picassco] **e,** Estimated Zernike aberration surfaces from single molecules of 3D data, compared to interpolated aberrations from bead z-stack calibration. **f**, Full 3D FOV image reconstructed using Vectorfit and fitted aberrations from single molecules. The 3D image is shifted 14 μm to the upper right compared to the 2D image. The false color indicates the axial distance from the nominal focus. In the 3D image, the microtubules that overlay the nucleus are visible in blue, while in the 2D image, these regions appear black, as the emitters are too much out of focus. **g,h** insets of **f** where (i) shows the Vectorfit result and (ii) with SMAP [SMAP]. **i,j,** xz-cross sections of the regions indicated in **g,h**. Scale bars: 10 μm (**b,f**), 500 nm (**c,d,g,h,i,j**)
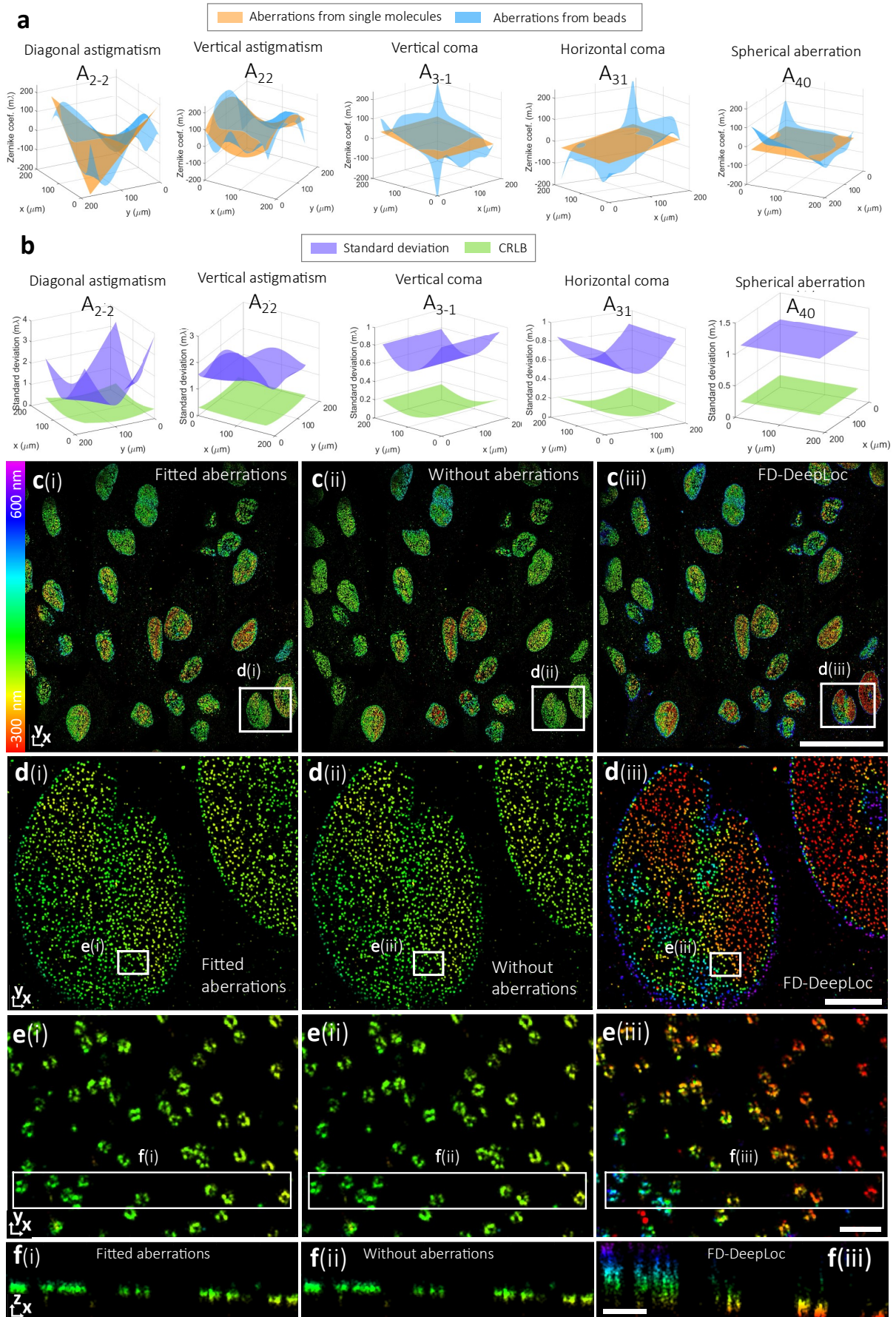
Fig. 3 | Aberration estimation and reconstructions for experimental 3D astigmatic data of NPCs. a, Estimated Zernike aberration surfaces from single molecules, compared to interpolated aberrations from bead z-stack calibration. **b,** Standard deviation and CRLB for the estimated Zernike coefficients. The standard deviations were calculated by repeating the estimation process 30 times with different randomly selected subsets of 5,000 localizations. **c**(i)-(iii), Full FOV images reconstructed using fitted aberrations from single molecules, without aberrations ($A_{22} = 103\ m\lambda$ and other Zernike modes set to zero) and FD-DeepLoc. **d**(i)-(iii), Zoomed views of the regions indicated by the boxes **d**(i)-(iii) in **c**. **e**(i)-(iii), Zoomed views of the regions indicated by the boxes **e**(i)-(iii) in **d**. **f**(i)-(iii), xz-cross sections of the regions indicated by the boxes **f**(i)-(iii) in **e**. Scale bars: 50 μm (**c**), 5 μm (**d**), 500 nm (**e,f**).

13

Methods

Processing pipeline

First, the set of acquired images were offset and gain corrected to convert ADUs into photon counts [Heintzmann2018, vanVliet1998a]. ROIs of size of 9×9 pixels for 2D data and 17×17 pixels for 3D data pixels are identified by a two-stage filtering process to reduce photon noise and local background followed by an intensity threshold [Huang2011, Smith2015]. In short, we apply uniform filters to the raw images with filter size 3 and 7 pixels and take the difference. We then compute the local maximum in a 7×7 pixels neighbourhood for all pixels and accept the central pixel as candidate for a single-molecule spot if its value is the local maximum and is higher than a threshold directly above the background noise level (different for each dataset). Now the 9×9 or 17×17 ROI is segmented out for all candidates and each ROI, labelled with index s, is extracted, leading to a total of N_{loc} candidate localization events.

A randomly selected subset of $M_s \ll N_{loc}$ of localizations is processed to determine the Legendre NAT coefficients that characterize the variation of the Zernike aberration coefficients across the FOV. To ensure an even distribution over the FOV, the localizations are selected on a 10x10 grid over the FOV where the same number of localizations is selected from each grid cell. The selected localizations are fitted using a Maximum Likelihood Estimation (MLE) algorithm with alternating local and global iteration steps. In the local iteration step the emitter position $\vec{r}_s = (x_s, y_s, z_s)$, signal photon count N_{ph} and background bg of all M_s localizations are updated until they no longer improve, given the current set of NAT coefficients. For 2D data, the axial position is fitted as well, ensuring that $z_s \geq 0$, because this leads to an improved aberration estimation compared to keeping the z_s value fixed. In each global update step, 20% of the M_s localizations are selected randomly. Then, the NAT coefficients are updated until they no longer improve, given the estimated positions, signal photon counts and background level of this subset of localization events. By using only 20% of the emitters in each step, the process is sped up significantly, while, with high probability, each of the M_s localization events will be used at least once in the full optimization process. The process typically converges after 6-10 steps and takes 30-45 seconds with $M_s = 5,000$ on a CPU using 12 cores (Intel Xeon w5-2455X). The NAT coefficients are randomly initialized such that the system aberration value W_{SYS} (see Supplementary Note) is equal to $40 \, m\lambda$, because a small value different from zero appeared to result in a faster convergence. In the case of 3D data with astigmatism, an additional amount of $150 \, m\lambda$ of constant astigmatism is incorporated in the initial NAT coefficients. For 2D data, the procedure is repeated 10 times with different random initializations for the same emitters, after which the solution with the highest likelihood is selected to rule out convergence to a local maximum. This is needed to mitigate the risk of getting stuck in a local optimum during optimization. The precision of the method is estimated by repeating the process $P_s = 30$ times by taking different randomly selected subsets of M_s localization events. We found that several thousands of localization events are typically sufficient to reliably find the NAT coefficients (compare Supplementary Figures 6 and 7). Unless indicated otherwise, we use the NAT expansion to first order, giving 13 independent NAT coefficients that we denote by $\bar{\gamma}_k$ with $k = 1, 2 \dots, 13$, taking into account the dependencies between the γ_{lqnm} as described in Supplementary Note.

For each set of NAT coefficients, there is another set of NAT coefficients that leads to the same PSF, in combination with flipping the sign of the axial coordinate z_s. This symmetric set of coefficients is

obtained by flipping the sign of the NAT coefficients that belong to Zernike coefficients A_{nm} for which n is even and keeping the other NAT coefficients the same. This is because flipping the sign of these even Zernike coefficients in combination with flipping the sign of z_s leads to the same PSF, while for the Zernike coefficients for which n is uneven, only flipping the sign of z_s leads to the same PSF. For the first-order Zernike coefficients, the even n modes are defocus, astigmatism and spherical aberration ($A_{20}, A_{2-2}, A_{22}, A_{40}$), while the odd modes are tilt and coma ($A_{11}, A_{1-1}, A_{3-1}, A_{31}$) where tilt is not fitted in our model because it comes down to a shift in x_s or y_s. Therefore, the symmetric solution is obtained by flipping the sign of $\bar{\gamma}_k$ for all $k = 1,2,\dots,13$, except for $k = 10,11,12$, which are the NAT coefficients that correspond to coma. To make sure our method gives an unambiguous solution, we choose the solution for which γ_5 is positive, which corresponds to the constant term of A_{22} (vertical astigmatism).

In a next step all N_{loc} localization events are fitted with the vectorial PSF model including the found field dependent aberrations. In these fits, as well as for the local update step in the estimation of the aberration field, we achieve major steps in efficiency by execution on GPU, as well as by several algorithmic improvements. The first algorithmic improvement is the application of phasor fitting [Maertens2018] to provide an initial estimate of the lateral position (x_0, y_0) of the emitter. In short, the measured photon counts n_k ($k = 1,2,\dots,K^2$) of all $a \times a$ sized pixels in the ROI with pixel coordinates (x_k, y_k) are used to compute the four phasor components:

$$C_x = \sum_k n_k \cos\left(\frac{2\pi}{K}\frac{x_k}{a}\right), \tag{1}$$

$$S_x = \sum_k n_k \sin\left(\frac{2\pi}{K}\frac{x_k}{a}\right), \tag{2}$$

$$C_y = \sum_k n_k \cos\left(\frac{2\pi}{K}\frac{y_k}{a}\right), \tag{3}$$

$$S_y = \sum_k n_k \sin\left(\frac{2\pi}{K}\frac{y_k}{a}\right), \tag{4}$$

after which we find the emitter position as:

$$x_0 = \frac{Ka}{2\pi} \tan^{-1}\left(\frac{S_x}{C_x}\right), \tag{5}$$

$$y_0 = \frac{Ka}{2\pi} \tan^{-1}\left(\frac{S_y}{C_y}\right). \tag{6}$$

Next, a first estimate is made of the constant background photon count per pixel $\bar{b}_0$ by taking the median of pixels at the rim of the ROI, and a first estimate is made for the signal photon count by $\bar{N}_0 = \sum_k n_k - K^2\bar{b}_0$. In case of 3D astigmatic fitting an initial estimate of the axial coordinate is found via computing the phasor magnitudes [Martens2018]:

$$A_{xx} = \sqrt{C_x^2 + S_x^2}, \tag{7}$$

$$A_{yy} = \sqrt{C_y^2 + S_y^2}, \tag{8}$$

and subsequently fitting two parabolas according to [Rieger2014]:

$$A_{xx,yy} = A_0\left[1 + \left(\frac{z_0 \pm l}{d}\right)^2\right], \tag{9}$$

15

where A_0 is an overall constant measuring spot width, $d = \lambda/\left(n - \sqrt{n^2 - NA^2}\right)$ is the depth of focus and $l = (2n\lambda/NA^2)\sqrt{8}A_{22}$ is half the distance between the two focal lines. Solving for the axial position gives:

$$z_0 = l\, \frac{(1 + d^2/l^2)\Delta A}{A_c + \sqrt{A_c{}^2 - (1 + d^2/l^2)\Delta A^2}}, \tag{10}$$

with $A_c = A_{xx} + A_{yy}$ and $\Delta A = A_{yy} - A_{xx}$. In case the astigmatism is not vertical, we can use the more general formulas $l = (2n\lambda/NA^2)\sqrt{8}\sqrt{A_{2-2}^2 + A_{22}^2}$ and:

$$\Delta A = \cos\left(2 \cdot \alpha_{\text{astig}}\right) \cdot (A_{yy} - A_{xx}) + 2\sin(2 \cdot \alpha_{\text{astig}}) \cdot A_{xy}, \tag{11}$$

where

$$A_{xy} = \frac{1}{2}\sum_k n_k \sin\left(\frac{2\pi}{K}\frac{x_k}{a}\right)\sin\left(\frac{2\pi}{K}\frac{y_k}{a}\right), \tag{12}$$

and where the astigmatism angle α_{astig} is determined by the amplitude of vertical and diagonal astigmatism:

$$\alpha_{\text{astig}} = \frac{1}{2}\tan^{-1}\left(\frac{A_{2-2}}{A_{22}}\right). \tag{13}$$

It appears that the accuracy of the initial value estimation of the axial coordinate depends on the magnitude of the astigmatism compared to size of the ROI, i.e. how well the two focal lines fit within the ROI.

The second algorithmic improvement is the use of linear regression for an improved initial estimate for signal and background photon count. To that end the vectorial PSF $H_k = H(x_k - x_0, y_k - y_0)$ for 2D fitting or $H_k = H(x_k - x_0, y_k - y_0, z_0)$ for 3D fitting is computed (see for details below) at all pixel positions in the ROI. The improved initial values for signal photon count N_0 and background photon count per pixel b_0 are then found by weighted linear regression, i.e. by minimizing:

$$E = \sum_k \frac{\left(n_k - N_0 H_k - \bar{b}_0\right)^2}{\bar{N}_0 H_k + \bar{b}_0} \tag{14}$$

The accuracy of the initial value estimation is such that the average number of iterations needed in the final vectorial PSF localization ranges for 2D fitting from 3.7 ± 0.6 at low photon count (500 photons) down to 1.1 ± 0.3 at high photon count (10^5 photons), and for 3D fitting from 7.1 ± 4.2 at low photon count (500 photons) down to 4.4 ± 0.8 at high photon count (10^5 photons), as determined from a simulation study (see below for details).

The third algorithmic improvement is the use of precomputation of the Optical Transfer Function (OTF) in order to reduce the number of Fourier Transformations (FTs) per iteration. The vectorial PSF for 2D-fitting can be written as:

$$H(x, y) = \sum_{j=1}^{3}\sum_{k}^{2}\left|F_{jk}(x, y)\right|^2 \tag{15}$$

Where the matrix $F_{jk}(x, y)$ can be expressed as the 2D FT over the objective lens pupil:

$$F_{jk}(x, y) = \iint P_{jk}(q_x, q_y)\, e^{iW(q_x, q_y)}e^{-2\pi i(q_x x + q_y y)}dq_x dq_y \tag{16}$$

with $W(q_x, q_y)$ the aberration function. The aberration function is computed from the NAT coefficients for the position of the center of the ROI in the FOV. Explicit expressions for the pupil matrix

16

$P_{jk}(q_x, q_y)$ can be found elsewhere [Stallinga2010]. The iterative optimization in the localization algorithm requires the computation of the PSF $H_k = H(x_k - x_0, y_k - y_0)$ for all pixels in the ROI, as well as the derivatives of the PSF w.r.t. the emitter coordinates. Applying the chain rule to eq. (10) implies that the derivatives of the matrix components $F_{jk}(x, y)$ are then needed. These derivatives give extra factors $2\pi i(q_x, q_y)$ in the integrands for the FT. As a consequence, we need to evaluate $3 \times 6 = 18$ FTs in each iteration of the localization algorithm, which takes the bulk of the computation time. This can be improved by pre-computing the PSF using eq. (1), and then computing the OTF via FT:

$$\widehat{H}(q_x, q_y) = \iint H(x, y)e^{-2\pi i(q_x x + q_y y)}dxdy \tag{17}$$

and storing the OTF in memory. For each iteration of the localization algorithm, we only need to compute the PSF and the derivatives w.r.t. x_0 and y_0 via the evaluation of 3 inverse FTs:

$$H(x - x_0, y - y_0) = \iint \widehat{H}(q_x, q_y)e^{-2\pi i(q_x x_0 + q_y y_0)}e^{2\pi i(q_x x + q_y y)}dq_x dq_y \tag{18}$$

$$\frac{\partial H(x - x_0, y - y_0)}{\partial x_0} = -2\pi i \iint q_x \widehat{H}(q_x, q_y)e^{-2\pi i(q_x x_0 + q_y y_0)}e^{2\pi i(q_x x + q_y y)}dq_x dq_y \tag{19}$$

$$\frac{\partial H(x - x_0, y - y_0)}{\partial y_0} = -2\pi i \iint q_y \widehat{H}(q_x, q_y)e^{-2\pi i(q_x x_0 + q_y y_0)}e^{2\pi i(q_x x + q_y y)}dq_x dq_y \tag{20}$$

We only need to compute 3 inverse FTs now, i.e. this saves a factor of 6 in number of FTs to execute and about a 2-3x improvement in computation time.

For the case of 3D fitting a similar approach can be followed. Now the matrix components F_{jk} depend on z as well via:

$$F_{jk}(x, y, z) = \iint P_{jk}(q_x, q_y)e^{2\pi i(q_x x + q_y y + q_z z)}dq_x dq_y \tag{21}$$

with the z-component of the spatial frequency vector:

$$q_z = \sqrt{\left(\frac{n_{med}}{\lambda}\right)^2 - q_x{}^2 - q_y{}^2} \tag{22}$$

with n_{med} the refractive index of the medium and λ the wavelength of the fluorescence emission. Now we need to do a one-time computation of the 3D-PSF $H(x, y, z)$, and subsequently of the 3D-OTF $\widehat{H}(q_x, q_y, q_z)$ by inverse FT, and store that in memory. Note that after this inverse FT, the z-component of the spatial frequency vector is not restricted to be dependent on the two lateral components as in the integration over the objective pupil plane. For each iteration step in the localization algorithm, we compute the 2D-OTFs from the 3D-OTF:

$$\widetilde{H}(q_x, q_y, z_0) = \int \widehat{H}(q_x, q_y, q_z)e^{2\pi i q_z z_0}dq_z \tag{23}$$

$$\widetilde{H}_z(q_x, q_y, z_0) = -2\pi i \int q_z \widehat{H}(q_x, q_y, q_z)e^{2\pi i q_z z_0}dq_z \tag{24}$$

and then the PSF and the derivatives w.r.t. all the coordinates via the 4 FTs:

$$H(x - x_0, y - y_0, z_0) = \iint \widetilde{H}(q_x, q_y, z_0)\, e^{-2\pi i(q_x x_0 + q_y y_0)}e^{2\pi i(q_x x + q_y y)}dq_x dq_y \tag{25}$$

$$\frac{\partial H(x - x_0, y - y_0, z_0)}{\partial x_0} = -2\pi i \iint q_x \widetilde{H}(q_x, q_y, z_0)\, e^{-2\pi i(q_x x_0 + q_y y_0)}e^{2\pi i(q_x x + q_y y)}dq_x dq_y \tag{26}$$

$$\frac{\partial H(x - x_0, y - y_0, z_0)}{\partial y_0} = -2\pi i \iint q_y \widetilde{H}(q_x, q_y, z_0)\, e^{-2\pi i(q_x x_0 + q_y y_0)}e^{2\pi i(q_x x + q_y y)}dq_x dq_y \tag{27}$$

17

$$\frac{\partial H(x - x_0, y - y_0, z_0)}{\partial z_0} = -2\pi i \iint \widetilde{H}_z(q_x, q_y, z_0) \, e^{-2\pi i (q_x x_0 + q_y y_0)} e^{2\pi i (q_x x + q_y y)} dq_x dq_y \quad (28)$$

Full direct computation of the PSF and the 3 derivatives via computation of the matrix elements $F_{jk}(x, y, z)$ and its derivatives needs $4 \times 6 = 24$ FTs, so that again a factor of 6 is gained in number of FTs. The sixfold reduction of the number of FTs to execute, which make up a substantial portion of the total computational costs, leads to a 2 to 3-fold improvement in the computation time for 2D and 3D CPU and 2D GPU implementations (Supplementary Table 1). For 3D fitting on the GPU, however, no improvement is not realized. The exact cause of this remains unclear, but maybe be debit to inefficiencies in our GPU implementation of the 3D OTF computation.

In principle, the 2D or 3D OTF needs to be pre-computed for all different ROI positions in the FOV. The Zernike aberration coefficients, however, vary slowly across the FOV. In practice therefore, it is sufficient to pre-compute the OTF for 20x20 patches in the FOV. All FTs are executed with the Chirp Z-Transform (CZT) algorithm [Bakx2002], which allows for a free choice of number of samples and spacing of samples in both the spatial and the Fourier domain. For direct vectorial PSF fitting we use a pupil plane sampling with 42×42 for the 2D data (7×7) ROI and 48×48 for the 3D data (17×17) ROI. In order to perform the Fourier transforms, the pupil size N_{pupil} is chosen such that $N_{ROI} + N_{pupil} - 1$ is either a power of 2, or 3 times a power of 2.

For each ROI, optimization is stopped when the growth rate of the likelihood becomes smaller than a stopping tolerance ε_{stop}. When a ROI has not converged after a maximum number of iterations of 10 for 2D data and 20 for 3D data, this ROI is indicated as outlier and is discarded. For experimental data we use $\varepsilon_{stop} = 5 \cdot 10^{-10}$, because we found that larger values of ε_{stop} lead to a worse localization precision because the optimization is stopped too early, while for smaller values, too many ROIs are incorrectly indicated as outlier. For simulation data a larger value of $\varepsilon_{stop} = 10^{-6}$ appeared to be sufficient because with this the CRLB is already reached.

Simulation setup

To determine the fitting speed of the vectorial PSF (Supplementary Table 1), a total of 1 million emitters were simulated using the vectorial PSF model with 2,500 photons and 10 background photons per pixel using Poisson statistics to generate noisy instances of computed single molecule spots. The wavelength was 550 nm, the numerical aperture 1.33 and the pixel size was set to $\lambda/(4 \cdot NA)$. For 2D fitting, the ROIs were fitted in batches 500,000 on a Nvidia A100 GPU, and in batches of 60,000 on other devices. The sizes for the ROIs, pupil, OTF, and the PSF for calculating the OTF were 7×7, 42×42, 42×42, and 129×129 pixels, respectively. For 3D fitting, the ROIs were fitted in batches of 20,000, and the used sizes for the ROI, pupil, OTF, and the PSF for calculating the OTF were 17×17, 48×48, $48 \times 48 \times 40$, and $129 \times 129 \times 71$ pixels, respectively. The simulated lateral location was randomly chosen in the central 2×2 pixels of the ROI. For the 3D fitting, 150 mλ of astigmatism was used in both the forward and the fitting PSF model, and the simulated axial location was randomly chosen between -500 and 500 nm. The maximum number of iterations was set to 5 for 2D fitting and 10 for 3D fitting. The computations on the Nvidia A100 GPU were performed using the DelftBlue supercomputer [DHPC 2024]. To determine the number of iterations for different photon counts as described in the processing pipeline above, the average number of iterations was computed from localizing 10,000 emitters that were simulated using the vectorial PSF model with 500 and 100,000 photons. The ROI size for the 2D fitting was 9×9 with a pupil size of 40×40 and for 3D fitting the

18

ROI size was 17×17 with a pupil size of 48×48 pixels. The other parameters were the same as described above.

Post-processing

Sample drift was corrected on the localization data following the method of Cnossen et al. [Cnossen2021]. All images are rendered by histogram binning on a grid with super-resolution pixel size set to half the average CRLB of the localization precision, with additional Gaussian blurring with kernel size (standard deviation) equal to 2 super-resolution pixels.

For calculating the localization precision by linking of nearby localizations in subsequent frames, first the localization events with an estimated photon count between 500 and 10000 and an estimated precision (CRLB) below 60 nm (xy) and 120 nm (z) were selected. Histograms of these parameters show that only a few extreme outlier events are filtered in this way. Localizations in subsequent frames closer than 60 nm (xy) and 120 nm (z), and closer than 2× the smallest estimated precision (CRLB) of the events in all 2 or 3 spatial directions were considered to originate from a single on-event. The unbiased variance over the total run length of linked localizations is considered as the measure for the experimental localization precision. This method can induce a bias for small run lengths, especially if one of the linked localizations is of low quality (e.g. the first or last event where the molecule may be on for only a part of the frame time). For that reason only longer on-events are used in the statistical evaluation ($5 \leq$ run length ≤ 13).

Experimental PSF data (Supplementary Video 1) was assembled by binning the localization events over 4×4 patches over the FOV and over 45 bins in the axial direction. The acquired spots for the different ROIs in each bin were 5× upsampled using Fourier space zero padding, centered at the estimated (x_0, y_0) position and finally summed.

Sample preparation

To acquire the data of Figure 2, HeLa cells were plated on 25 mm coverslips in 6-well plate (Warner instruments) and allowed to adhere overnight (12-16 hours). First, cells were briefly washed with warm 1xPBS and then fixed using a solution containing 0.6% paraformaldehyde, 0.1% glutaraldehyde, and 0.25% Triton, all diluted in PBS for 60 seconds. This was followed by a secondary fixation using a buffer made up of 4% paraformaldehyde and 0.2% glutaraldehyde diluted in PBS for two hours. The cells were then washed twice with PBS, treated with 0.1% $NaBH_4$ for 5 minutes to minimize background fluorescence due to glutaraldehyde, and further washed twice with PBS. To quench any reactive cross-linkers, the samples were incubated in 10 mM Tris for 10 minutes, followed by one PBS wash. Finally, the samples were permeabilized through a 20 minutes incubation in a solution containing 5% BSA and 0.05% Triton X-100 diluted in PBS. At the end, the samples were washed once with PBS and prepared for the labeling procedure. Cells were incubated with 15 µg/ml in volume of 350 uL of Monoclonal Anti-β-Tubulin, Mouse IgG clone AA2 (Sigma #T8328, concentration: 2mg/ml) in 2% BSA + 0.05% Triton X-100 for 5 hours. Then cells were washed once with Massive Photonics DNA-PAINT washing buffer (1×). For secondary antibody labeling, cells were incubated with 15 µg/ml in volume of 350 uL per coverslip of DBCO-DAMIG-DNA Dock-2 (Acronym of DAMIG: Donkey anti-mouse IgG) secondary antibodies (stock concentration is 500 µg/ml) in 2% BSA + 0.05% Triton X-100 for 12 hours. This is followed by 3 times wash with Massive Photonics DNA-PAINT washing buffer (1×). Before imaging, cells were washed once with Massive Photonics imaging buffer. The cell sample was placed in an

19

Attofluor chamber then Atto655 imager-2 strands at concentration of 10 pM diluted in Massive Photonics imaging buffer was added for DNA-PAINT imaging.

Data acquisition

To acquire the data of Figure 2, the sample was mounted on a stage of a custom-built sequential microscope [Schodt2023] controlled by custom-written software (github.com/LidkeLab/matlab-instrument-control) in MATLAB (MathWorks Inc.) [Pallikkuth2018]. A high-power 647 nm laser (2RU-VFL-P-500-647-B1R, MPB Communications) was used as an excitation laser. A 100X silicon oil immersion objective (UPLSAPO100XS, Olympus) was used to collect emitted fluorescent light. A 708/75 nm bandpass filter (FF01-708/75-25-D, Semrock) was placed in the emitted fluorescence light path. A 4f relay system consisting of 175 mm and 125 mm focal lengths lenses was set up in the emission path. To acquire the 3D data, we have set up two cylindrical lenses at the Fourier plane of focal lengths 1 m and -1 m respectively, for introducing astigmatism. The sCMOS camera (C11440-22CU, Hamamatsu) was used to detect emitted fluorescence light. At first, brightfield reference image was taken using 660 nm LED (M660L3, Thorlabs) illumination lamp and saved. During data acquisition, 100 sequences of 800 frames (a total of 80,000) were collected at 10 Hz. The saved brightfield reference image was used to realign each cell after every 4th sequence of imaging [Wester2021]. The experimental settings for the 2D data are identical to those used for the 3D data, except that the pair of cylindrical lenses is removed.

Additional references

- K.J.A. Martens, A.N. Bader, S. Baas, B. Rieger, and J. Hohlbein. Phasor based single-molecule localization microscopy in 3D (pSMLM-3D): an algorithm for MHz localization rates using standard CPUs. *Journal of Chemical Physics*, **148**:123311, 2018.
- S. Stallinga and B. Rieger. Accuracy of the Gaussian point spread function model in 2D localization microscopy. *Optics Express*, **18**(24):24461-24476, 2010.
- J. L. Bakx, Efficient computation of optical disk readout by use of the chirp z transform, *Applied Optics* **41**, 4879-4903, 2002.
- J. Cnossen, T. Ju Cui, C. Joo, and C.S. Smith. Drift correction in localization microscopy using entropy minimization. *Optics Express*, **29**(18):27961-27974, 2021.
- D.J. Schodt, F. Farzam, S. Liu, and K.A. Lidke. Automated multi-target super-resolution microscopy with trust regions. *Biomedical Optics Express*, 14(1):429-440, 2023
- S. Pallikkuth, M. Meddens, M. Fazel, H. Farsibaf, F. Farzam, M. Wester, and K. Lidke, A Matlab-based instrument control package for fluorescence imaging. *Biophysical Journal* **114**, 532a 2018.
- M.J. Wester, D.J. Schodt, H. Mazloom-Farsibaf, M. Fazel, S. Pallikkuth, and K.A. Lidke. Robust, fiducial-free drift correction for super-resolution imaging. *Scientific reports*, **11**:23672, 2021.
- Delft High Performance Computing Centre (DHPC), DelftBlue Supercomputer (Phase 2), 2024, https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2
- R.V. Shack and K. Thompson, Influence of alignment errors of a telescope system on its aberration field. *Proc. SPIE*, **251**:146-153, 1980.
- K.P. Thompson, Description of the third-order optical aberrations of near-circular pupil optical systems without symmetry. *Journal Optical Society of America A* **22**:1389-1401, 2005.
- R. Tessieres, Analysis for alignment of optical systems, M.Sc. thesis, The University of Arizona, 2003.

- I.W. Kwee and J.J.M. Braat, Double Zernike expansion of the optical aberration function. *Pure Applied Optics* **2**:21-32, 1993.

- Y. Li, M. Mund, P. Hoess, J. Deschamps, U. Matti, B. Nijmeijer, V. Jimenez Sabinian, J. Ellenberg, I. Schön, and J. Ries. Real-time 3D single molecule localization using experimental point spread functions. *Nature Methods*, **15**:367-369, 2018.

- M. Bates, J. Keller-Findeisen, A. Przybylski, A. Hüper, T. Stephan, P. Ilgen, A.R. Cereceda Delgado, E. D'Este, S. Jakobs, S. Sahl, and S.W. Hell. Optimal precision and accuracy in 4Pi-STORM using dynamic spline PSF models. *Nature Methods*, **19**:603-612, 2022.
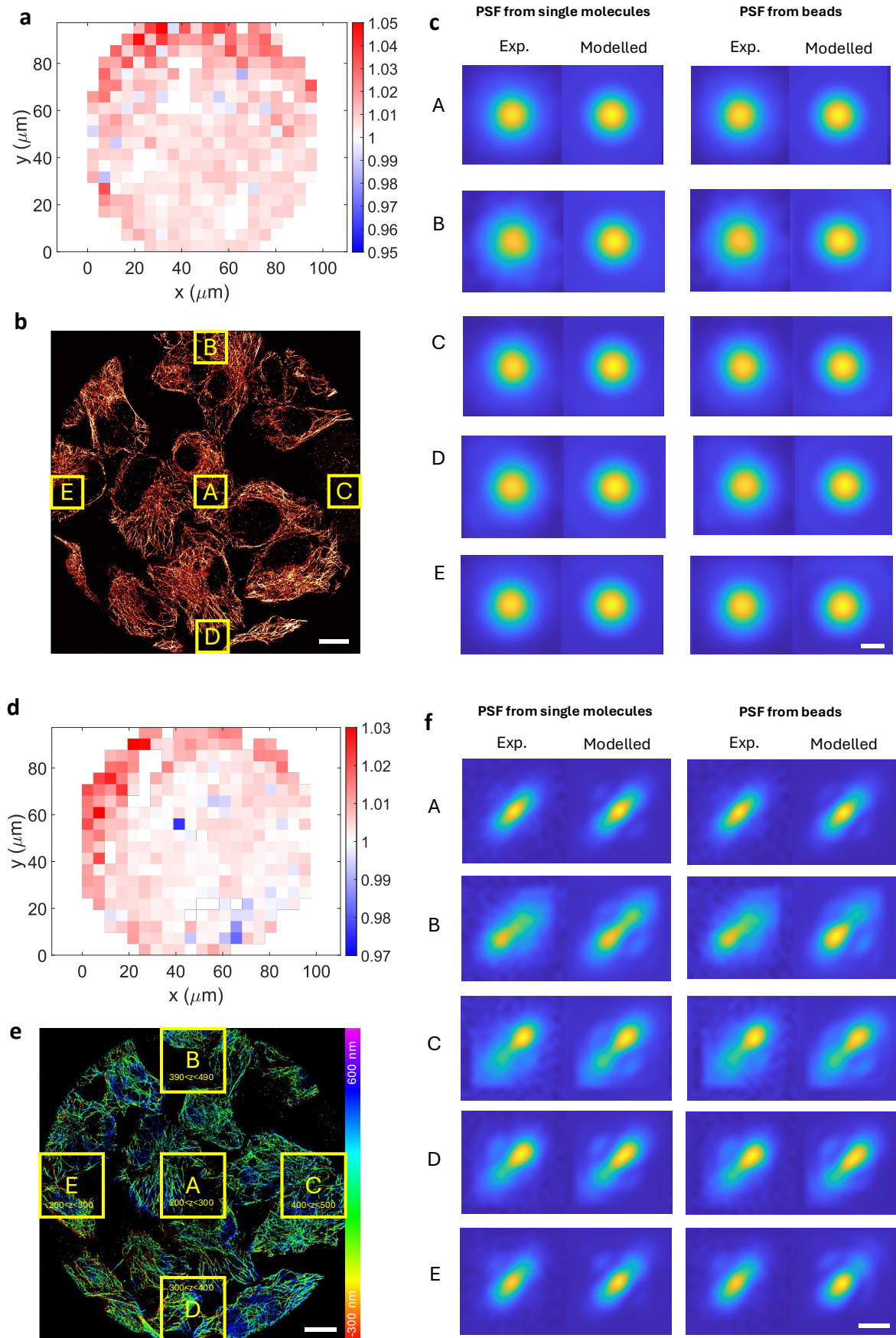
Supplementary Figure 1 | PSFs from single molecules compared to aberrations from bead calibration. a, The chi-squared goodness of fit (χ^2 GoF) value of the PSF model with bead calibration derived aberrations to the measured ROIs divided by the χ^2 GoF value of the PSF model with single molecule derived aberrations to the measured ROIs, averaged on a 20x20 tile grid over the whole FOV, for the 2D data shown in **b**. **b,** Full FOV 2D image indicating the 5 regions of the summed and upsampled PSFs shown in **c. c,** The average experimental and modelled PSFs within the regions as indicated in **b**, for both aberrations from single molecules and for aberrations from beads. The PSFs are obtained by centering and summing of PSFs, followed by 10-fold upsampling. Since the estimated location needed for the centering depends on the aberrations, a separate experimental PSF is shown for both the aberration choices. **d, e, f,** The same as **a, b, c**, but for the 3D data with astigmatism. Scale bars, 10 μm (**b,e**) 200 nm (**c**), 500 nm (**f**).
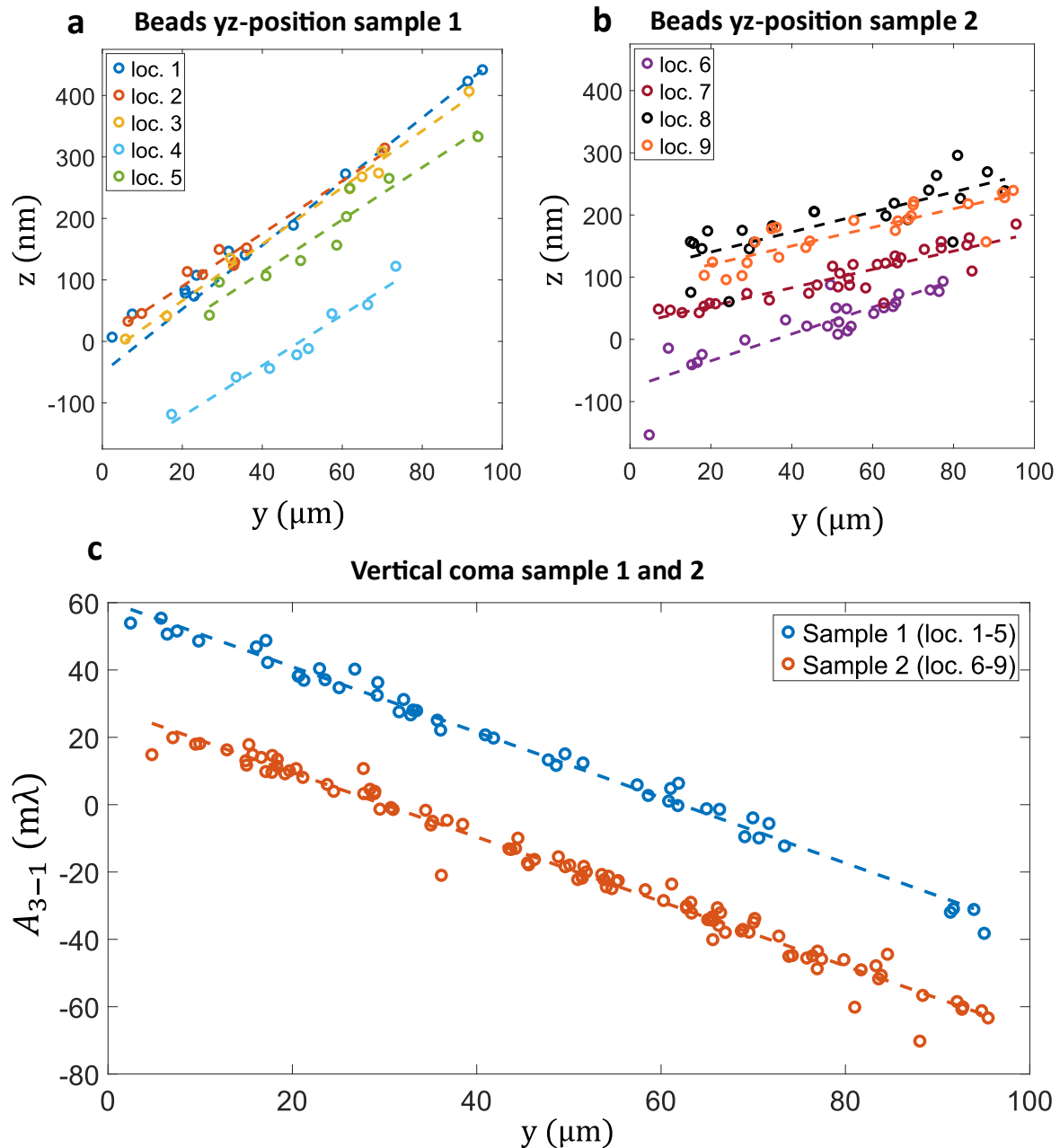
Supplementary Figure 2 | Aberration calibration with beads on two different samples Results of two bead calibration with two different bead samples, performed on two consecutive days. From sample 1, a z-stack was recorded at 5 different locations (loc. 1-5) and from sample 2, a z-stack was recorded at 4 different locations (loc. 6-9). **a,** Estimated yz-positions of sample 1 for each location. **b,** Estimated yz-positions of sample 2 for each location. **c,** Estimated Zernike coefficient A_{3-1} (vertical coma) for sample 1 and sample 2.
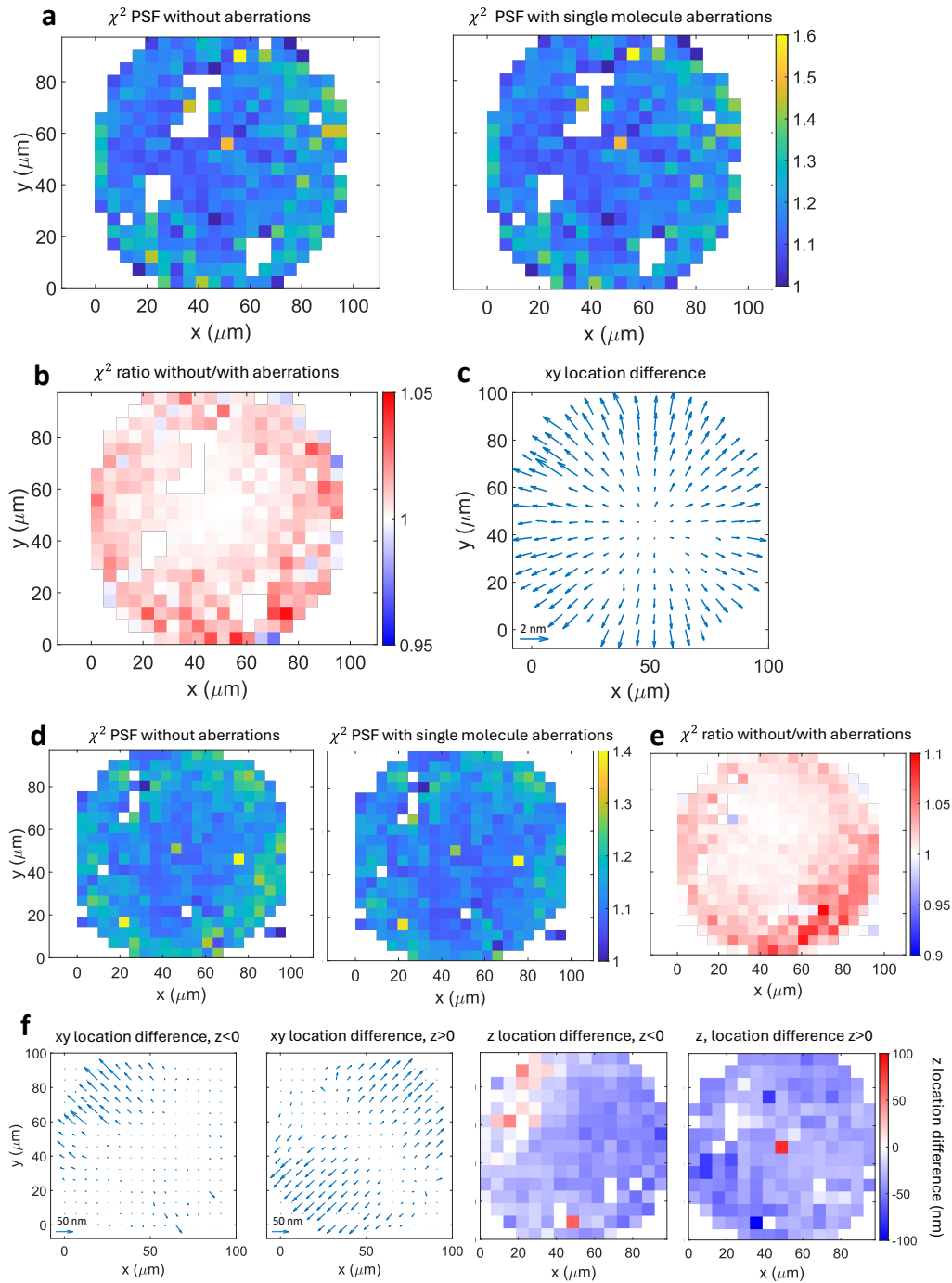
Supplementary Figure 3 | Comparison of goodness-of-fit and locations with and without aberrations. a, The chi-square goodness of fit value of the PSF model to the measured ROIs, averaged in 20×20 patched over the FOV, for the PSF model without aberrations, and the PSF model with field dependent aberrations from single molecules, for the 2D data. The values are normalized by the theoretically expected value for shot noise statistics. **b,** The ratio of the chi-square value without aberrations to the chi-square value with single molecules aberrations from **a**. **c,** The difference in the estimated xy locations that result from the PSF model without aberrations and the PSF model with single molecules aberrations, averaged over 15×15 patches over the FOV. **d-f,** show the same as **a-c,** but for the 3D data, where the xy and z location differences in **f** are broken down into emitters below ($z \leq 0$) and above the focal plane ($z \leq 0$).

25
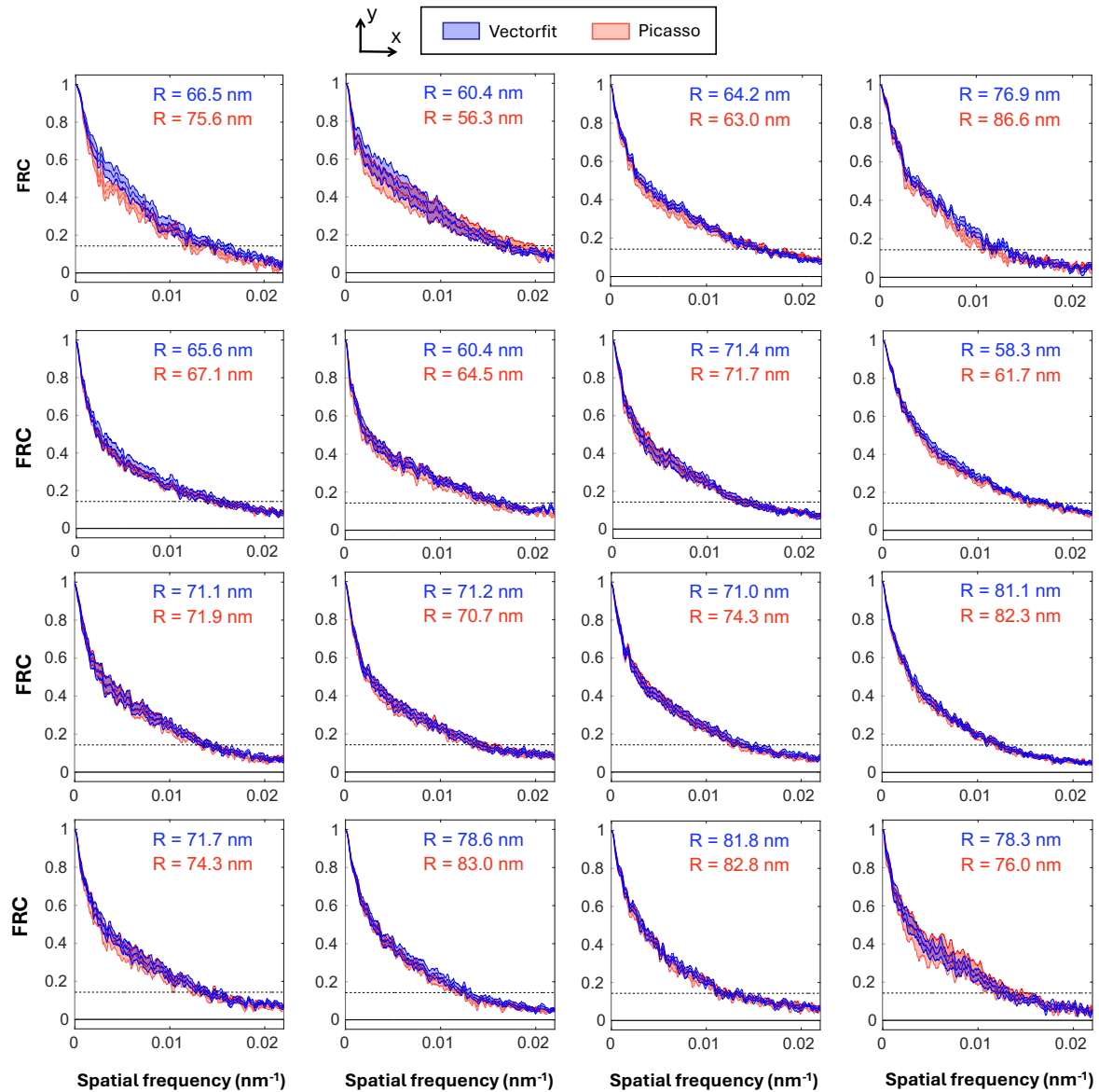
Supplementary Figure 4 | FRC analysis of 2D data of microtubili. FRC curves on a 4x4 grid of the FOV for localization data from Vectorfit and Picasso (Gaussian). The FRC analysis was done after linking localizations in consecutive frames and filtering out the brightest 10^{-6} fraction of super-resolution pixels, to eliminate hot pixels or persistently active emitters that could bias the computed resolution. Both the mean FRC curves and the spread quantified by the standard deviation over 10 repetitions are plotted. The FRC resolutions are shown in each subplot. The standard deviations are in the order of a few nm for all resolution values.
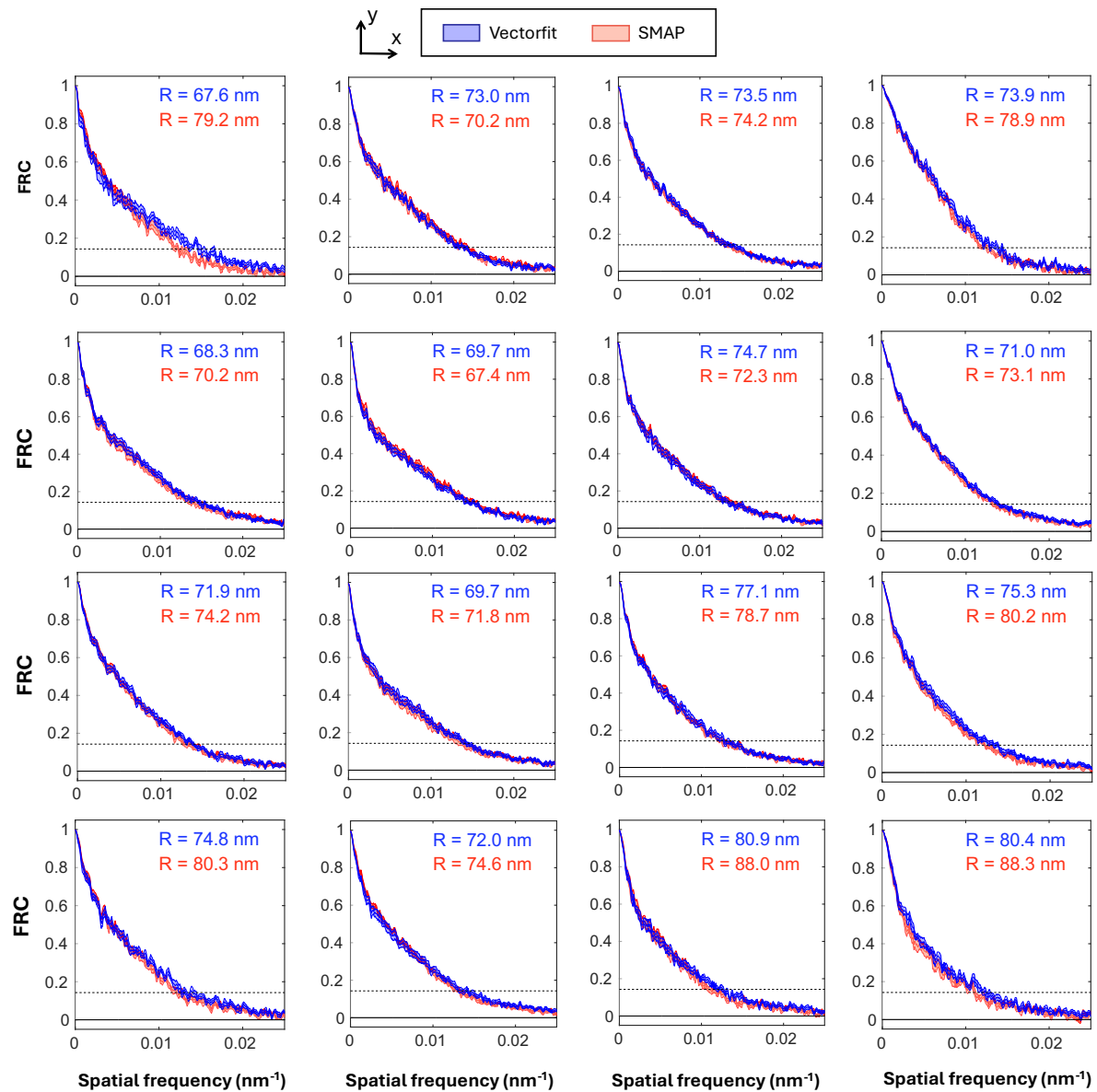
Supplementary Figure 5 | FRC analysis of 3D data of microtubili. FRC curves on a 4x4 grid of the FOV for localization data from Vectorfit and SMAP (spline). The FRC analysis was done after linking localizations in consecutive frames and filtering out the brightest $2 \cdot 10^{-5}$ fraction of super-resolution pixels, to eliminate hot pixels or persistently active emitters that could bias the computed resolution. Both the mean FRC curves and the spread quantified by the standard deviation over 10 repetitions are plotted. The FRC resolutions are shown in each subplot. The standard deviations are in the order of a few nm for all resolution values.
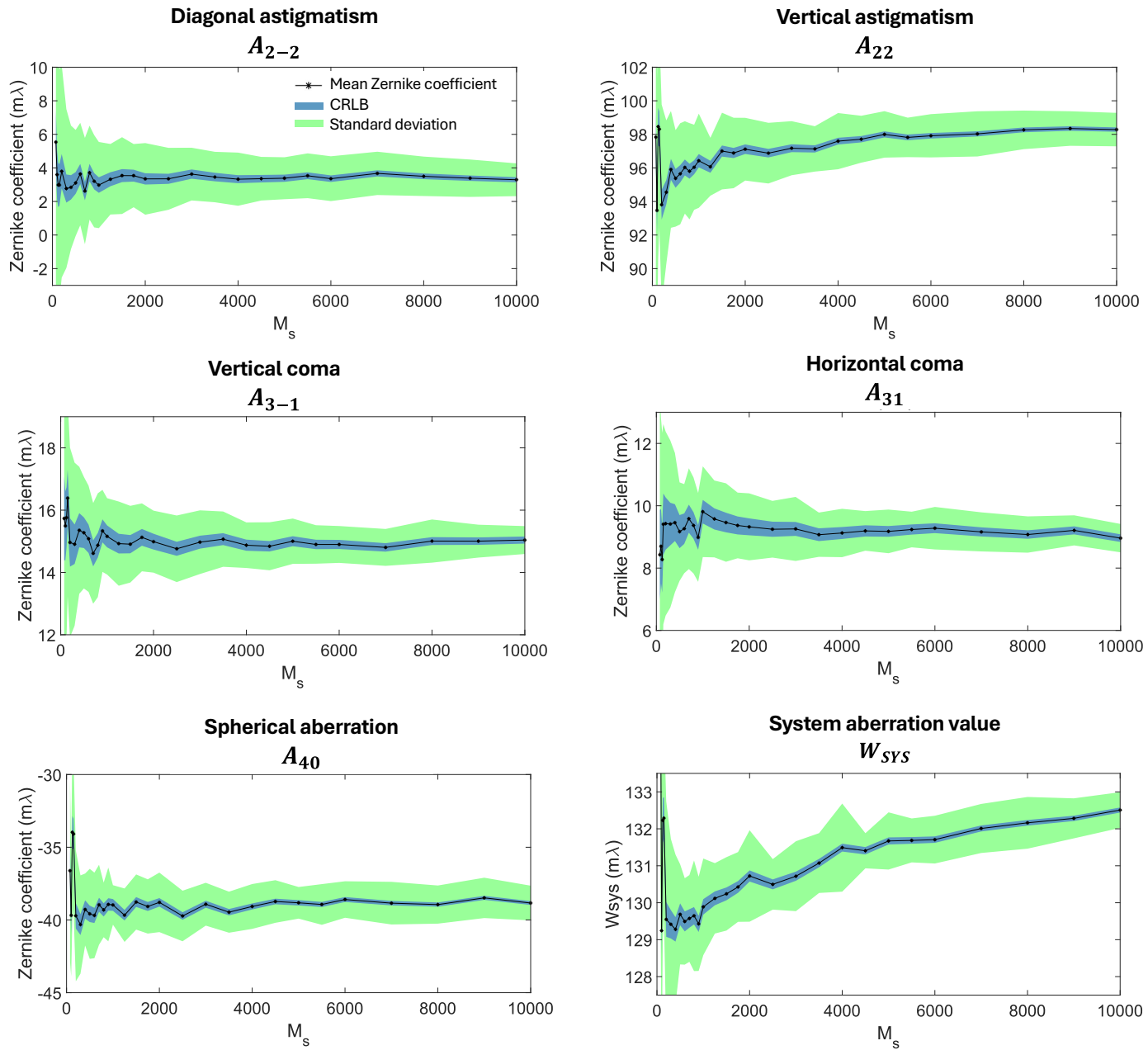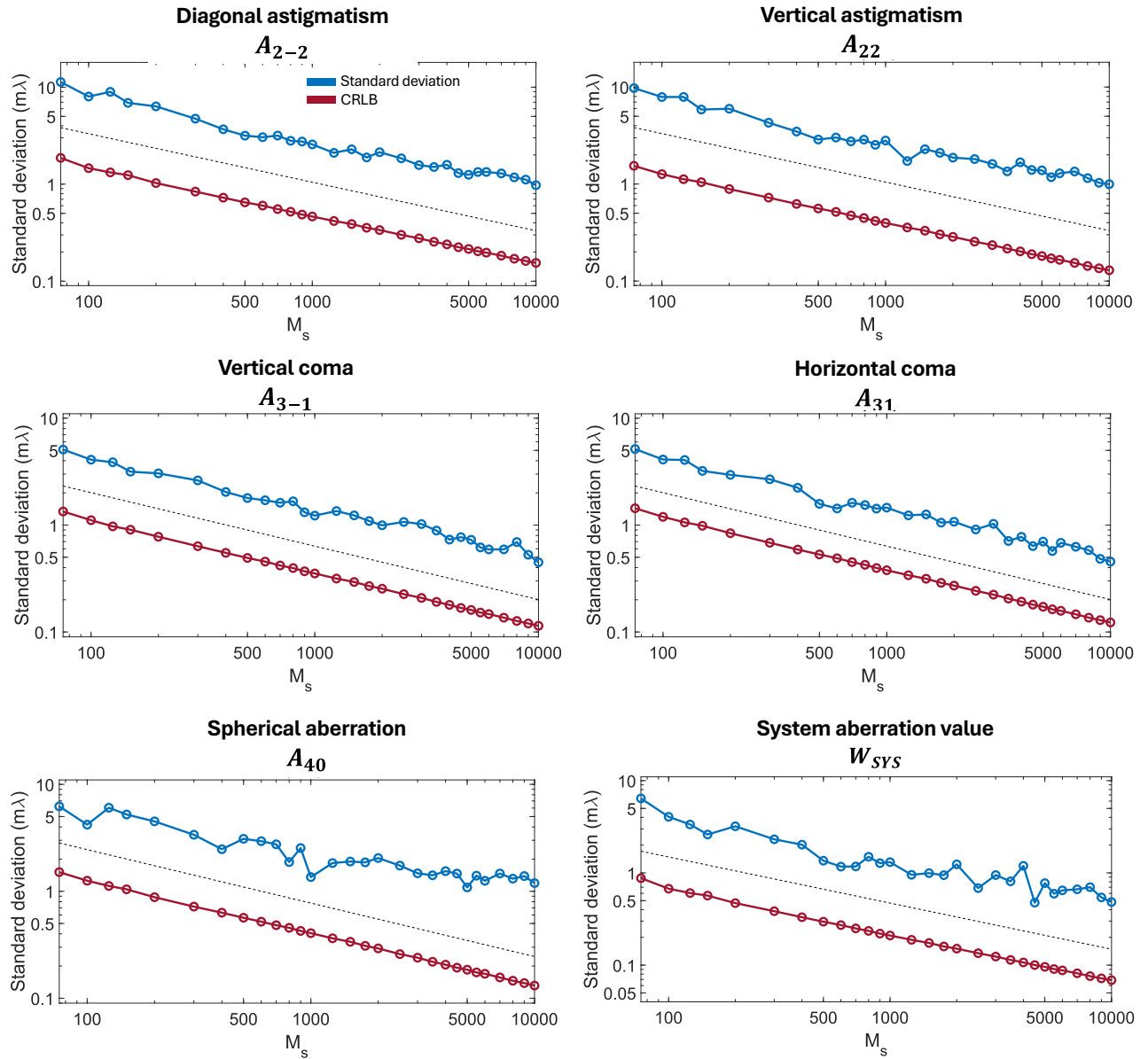
Supplementary Figure 6 | Estimated aberrations as a function of the number of spots M_s. The standard deviation of the estimated Zernike coefficients was calculated by repeating the estimation 30 times with different randomly selected subsets of M_s localization events. For each value of M_s, the Zernike coefficient CRLB was calculated. The Zernike coefficients, standard deviations and the Zernike coefficient CRLBs are averaged over the FOV. The bottom right panel shows the system aberration level (see Supplementary Note). . The 3D data of NPCs from Fig. 3 was used as input here.

Supplementary Figure 7 | Log-log plot of the precision of the Zernike coefficient estimation as a function of M_s compared to the CRLB. The standard deviation of the estimated Zernike coefficients was calculated by repeating the method 30 times with different randomly selected subsets of M_s localization events. For each value of M_s, the Zernike coefficient CRLB was calculated. The standard deviations and the Zernike coefficient CRLBs are averaged over the FOV. The dotted line has a slope of -0.5. The bottom right panel shows the standard deviation and CRLB of the system aberration level (see Supplementary Note). The 3D data of NPCs from Fig. 3 was used as input here.
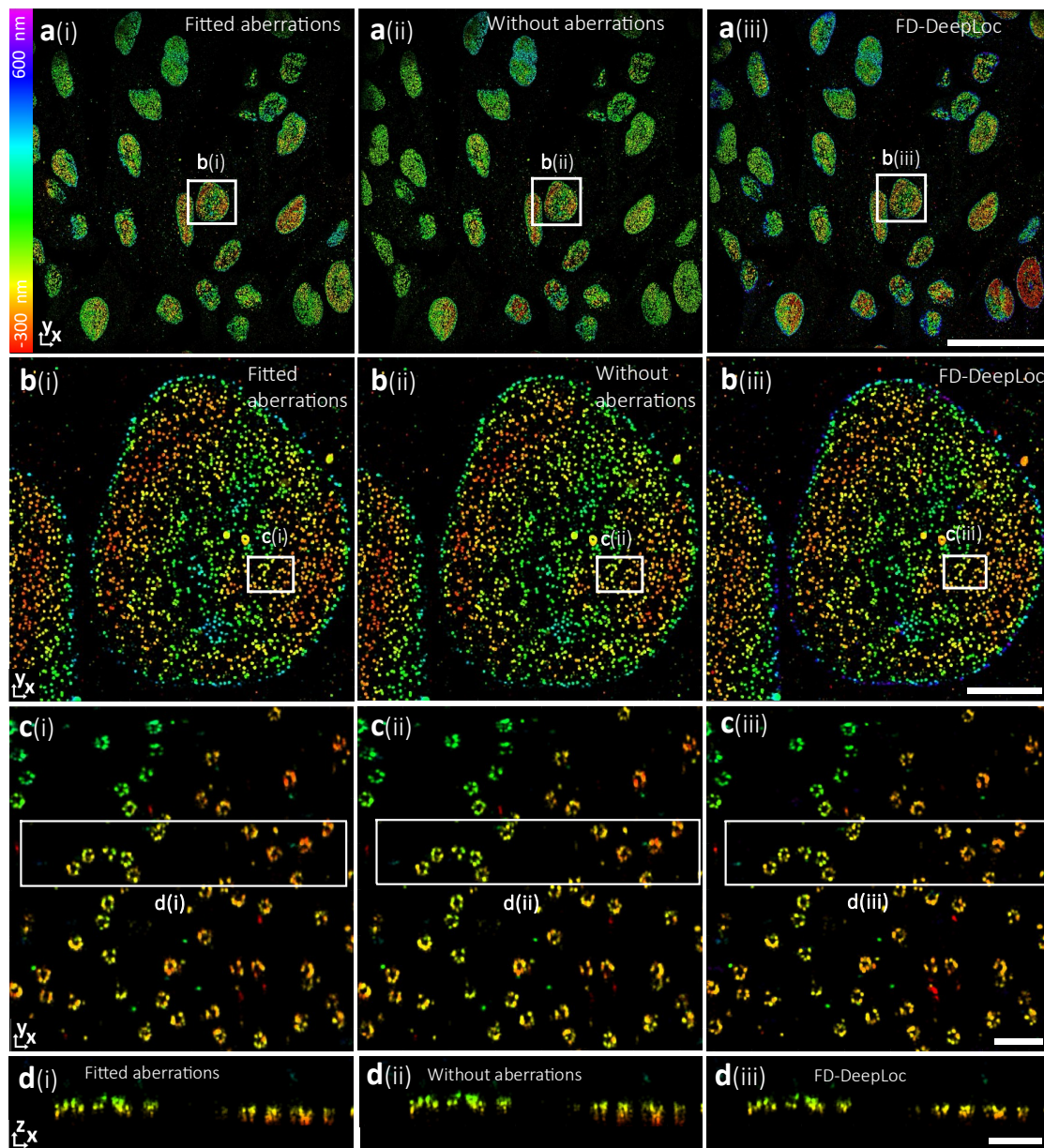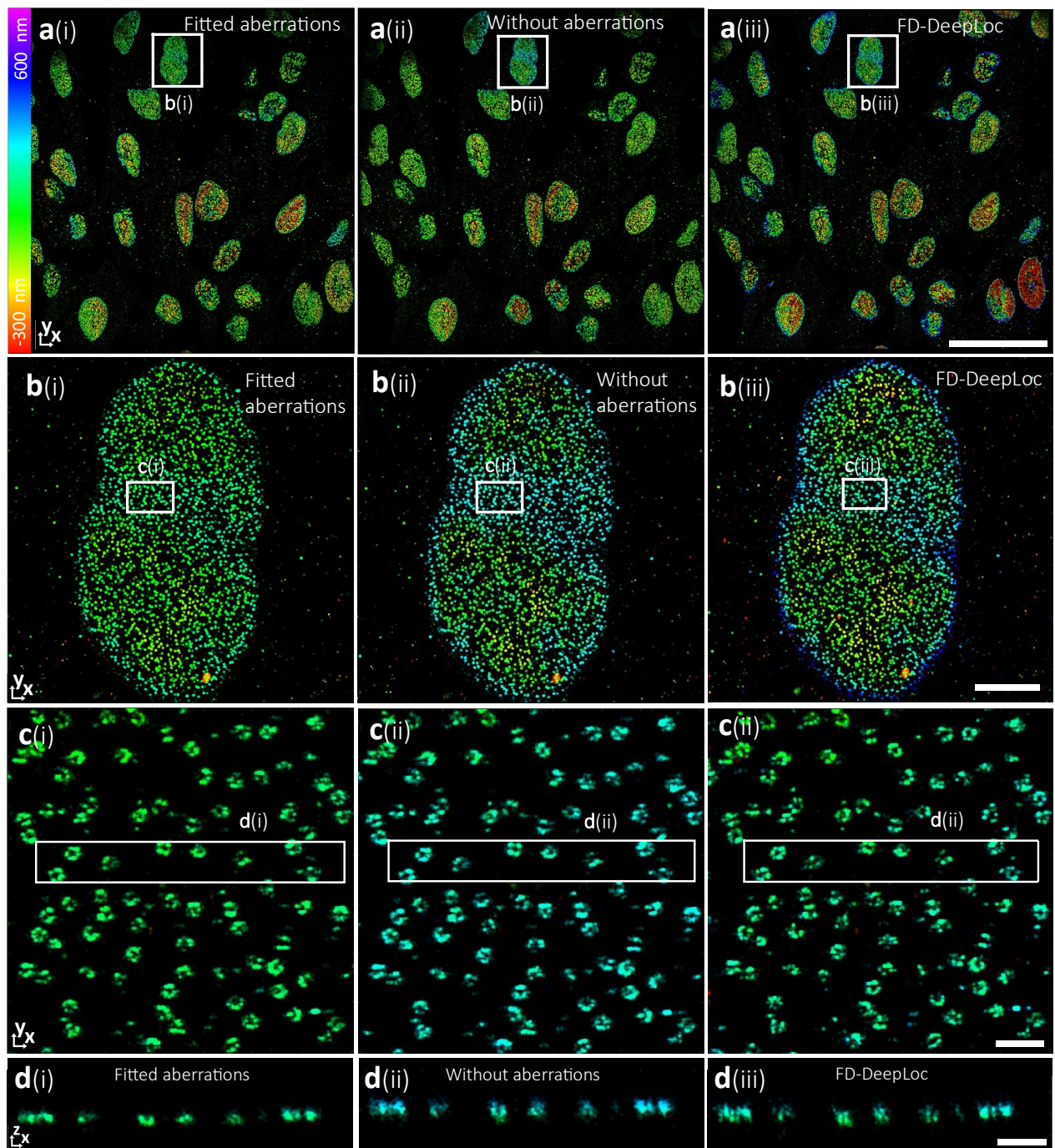
Supplementary Figure 8 | Reconstructions for experimental 3D astigmatic data of NPCs in the center of the FOV

a(i)-(iii), Full FOV images reconstructed with estimated aberrations from single molecules, without aberrations ($A_{22} = 103\ m\lambda$ and other Zernike modes set to zero), and FD-DeepLoc, **b**(i)-(iii), Zoomed views of the regions indicated by the boxes b(i)-(iii) in **a**. **c**(i)-(iii), Zoomed views of the regions indicated by the boxes c(i)-(iii) in **b**. **d**(i)-(iii), xz-cross sections of the regions indicated by the boxes d(i)-(iii) in **c**. Scale bars, 50 μm (**a**), 5 μm (**b**), 500 nm (**c,d**).

Supplementary Figure 9 | Reconstructions for experimental 3D astigmatic data of NPCs at the top of the FOV

a(i)-(iii), Full FOV images reconstructed with fitted aberrations from single molecules, without aberrations ($A_{22} = 103\ m\lambda$ and other Zernike modes set to zero), and FD-DeepLoc, **b**(i)-(iii), Zoomed views of the regions indicated by the boxes B(i)-(iii) in **a**. **c**(i)-(iii), Zoomed views of the regions indicated by the boxes **c**(i)-(iii) in **b**. **d**(i)-(iii), xz-cross sections of the regions indicated by the boxes **d**(i)-(iii) in **c**. Scale bars: 50 μm (**a**), 5 μm (**b**), 500 nm (**c,d**).
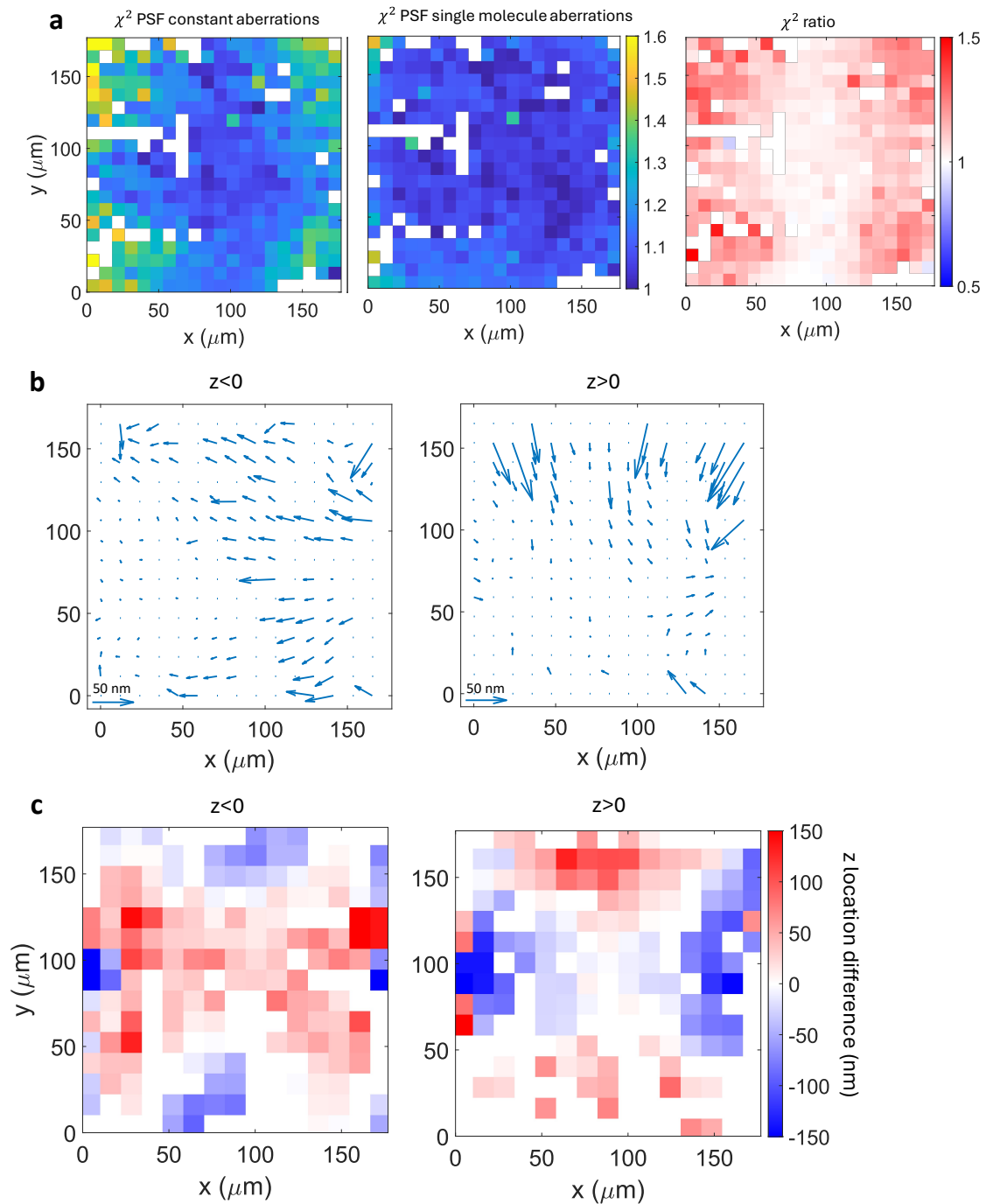
Supplementary Figure 10 | Comparison of goodness-of-fit and locations with and without aberrations. a, The chi-square goodness of fit value of the PSF model to the measured ROIs, averaged in 20×20 patched over the FOV, for the PSF model without aberrations ($A_{22} = 103\ m\lambda$ and other Zernike modes set to zero), and the PSF model with field dependent aberrations from single molecule data, and the ratio of the chi-square values without and with aberrations. The chi-square values are normalized by the theoretically expected value for shot noise statistics. **b,c,** The difference in the estimated xy-locations (**b**) and the z-locations (**c**) that result from the PSF model without aberrations and the PSF model with
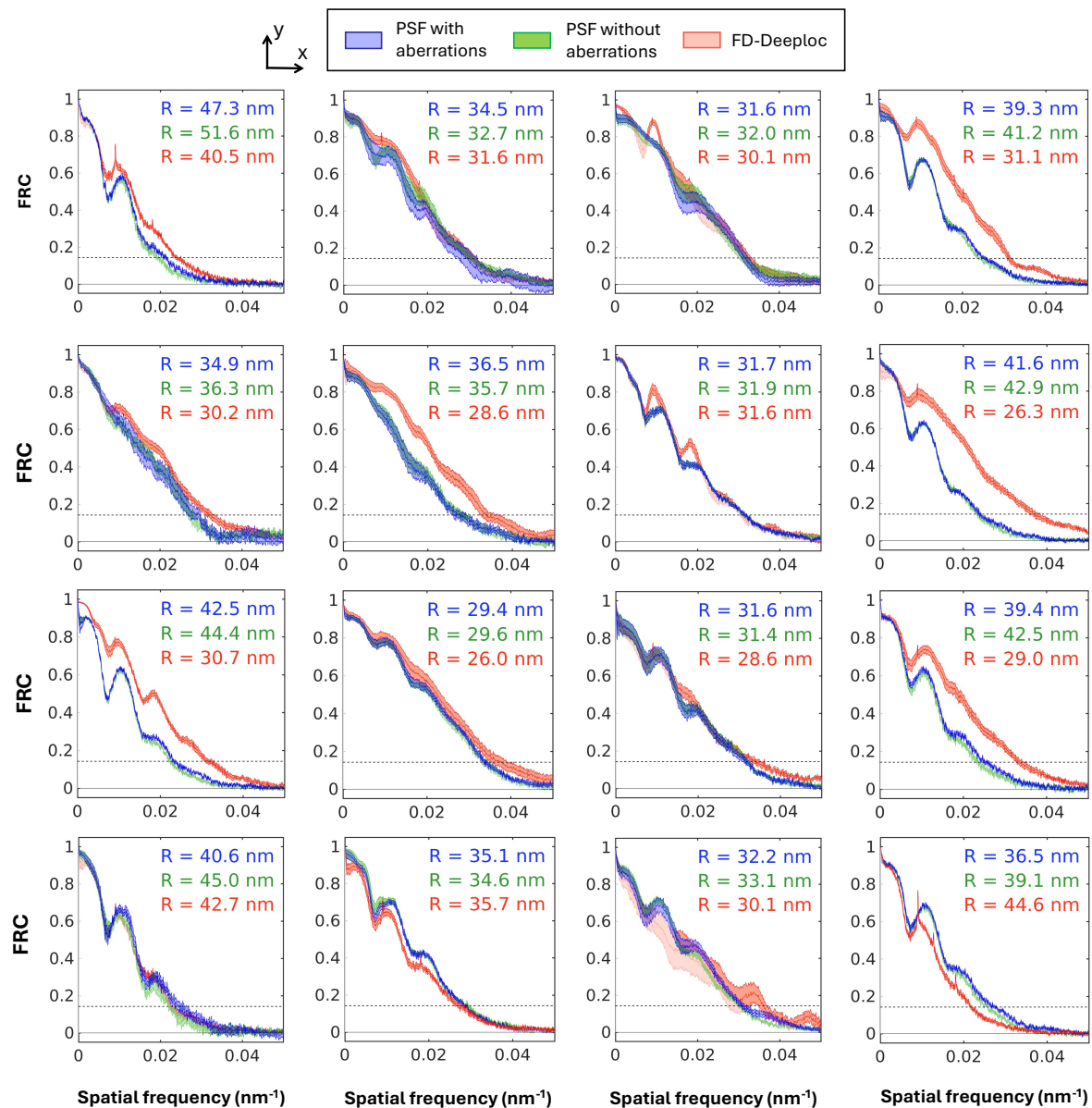
Supplementary Figure 11 | FRC analysis of 3D data of NUP96 in the NPC. FRC curves on a 4x4 grid of the FOV for localization data from Vectorfit with aberrations from single molecules, from Vectorfit without aberrations (with only 103 mλ constant astigmatism) and from FD-Deeploc. The FRC analysis was done after linking localizations in consecutive frames and filtering out the brightest 10^{-6} fraction of super-resolution pixels, to eliminate hot pixels or persistently active emitters that could bias the computed resolution. Both the mean FRC curves and the spread quantified by the standard deviation over 10 repetitions are plotted. The curves are not smooth, but have peaks that are most likely due to the dominant spatial frequency of the NPC structure (diameter around 100 nm), with first and second harmonic giving visible peaks. The FRC resolutions are shown in each subplot. The standard deviations are in the order of a few nm for all resolution values.

Supplementary Note

Implementation Nodal Aberration Theory (NAT) with Legendre polynomials

The key idea of NAT is to develop the Zernike aberration coefficients in a Taylor series of low order polynomials of the coordinates in the Field Of View (FOV), with the benefit of symmetry relations that reduce the number of independent coefficients. This applies to optical imaging systems with small field angles, such as microscopes or telescopes. In the following we will derive such Taylor series for 1st and 2nd order aberration theory, following the pioneering work of Shack and Thompson [Shack1980, Thompson2005] and Tessieres [Tessieres2003]. We will add a new principle to NAT, namely the use of expansions in Legendre polynomials to replace the plain Taylor series. Such an orthogonal basis function set is better adapted to the square shape of the FOV in camera based imaging, and enables a straightforward measure for the average aberration performance across the FOV, the so-called system aberration level.

Legendre polynomial expansion

In order to keep track of the typical or average performance across the FOV it is useful to know the system aberration value from:

$$W_{SYS}^2 \equiv \langle W_{RMS}^2 \rangle = \frac{1}{4} \int_{-1}^{1} \int_{-1}^{1} dx\,dy\, W_{RMS}^2(x,y) \tag{1}$$

where x and y are scaled coordinates in the FOV of size $L_x \times L_y$:

$$x = \frac{2x_c}{L_x} \tag{2}$$

$$y = \frac{2y_c}{L_y} \tag{3}$$

with x_c and y_c the physical FOV coordinates. The RMS aberration value at point (x,y) in the FOV is computed from the sum over the squares of the Zernike coefficients:

$$W_{RMS}(x,y) = \sqrt{\sum_{n,m} A_{nm}(x,y)^2} \tag{4}$$

The definition of the system RMS aberration value suggests that an expansion of the Zernike aberration coefficients $A_{nm}(x,y)$ in orthogonal functions may be useful. This has been done in terms of Zernike modes for circularly shaped FOVs [Kwee1993]. In typical imaging systems, however, we have a rectangular, often square, FOV, defined by the camera. This implies that expanding in a complete set of functions that are orthogonal on the unit square is more practical.

Such an expansion is possible using Legendre polynomials $P_n(x)$, which are a complete set of orthogonal functions on the interval $[-1,1]$. Starting from the lowest orders $P_0(x) = 1$ and $P_1(x) = x$, they can be easily computed using the recurrence relation:

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \tag{5}$$

This results in $P_2(x) = (3x^2 - 1)/2$, $\quad P_3(x) = (5x^3 - 3x)/2$, ... They satisfy the orthogonality relation:

$$\frac{1}{2} \int_{-1}^{1} dx \; P_n(x)P_m(x) = \frac{1}{2n+1} \delta_{nm} \tag{6}$$

with δ_{nm} the Kronecker delta. Legendre polynomial products can be used as an orthogonal and complete set of functions on the unit square:

$$Q_{lq}(x,y) = P_l(x)P_q(y) \tag{7}$$

The expansion of the Zernike coefficients across the FOV can now be written as:

$$A_{nm}(x,y) = \sum_{l,q} \gamma_{lqnm} Q_{lq}(x,y) \tag{8}$$

where the expansion coefficients γ_{lqnm} are henceforth called the Legendre NAT coefficients. The system RMS aberration follows from the orthogonality of the Legendre polynomials as:

$$W_{SYS}^2 = \langle W_{RMS}^2 \rangle = \sum_{l,q,m,n} \frac{\gamma_{lqnm}^2}{(2l+1)(2q+1)} \tag{9}$$

This enables a quantitative comparison of the importance of the Legendre NAT coefficients, both in absolute terms (compared to the Maréchal's diffraction limit) and in relative terms (compared to the whole set of Legendre NAT coefficients).

NAT for 1st order aberration theory

Aberration theory including the impact of misalignment is developed using the pupil coordinates $\vec{\rho} = (\rho_x, \rho_y) = \rho(\cos\theta, \sin\theta)$, the field coordinates $\vec{r} = (x,y) = r(\cos\psi, \sin\psi)$, and misalignment vectors of the form $\vec{\sigma} = (\sigma_x, \sigma_y)$ that describe lateral displacements from the aligned position. These can form 5 different rotational invariants that depend on $\vec{\rho}$ and/or $\vec{r}$, namely the inner products $\rho^2 = \vec{\rho} \cdot \vec{\rho}$, $r^2 = \vec{r} \cdot \vec{r}$, $\vec{\rho} \cdot \vec{r}$, $\vec{\rho} \cdot \vec{\sigma}$, $\vec{r} \cdot \vec{\sigma}$. The aberration function is expanded in powers of these 5 invariants with coefficients α_{pqsuv}:

$$W = \sum_{pqrsuv} \alpha_{pqsuv} (\rho^2)^p (r^2)^q (\vec{\rho} \cdot \vec{r})^s (\vec{\rho} \cdot \vec{\sigma}_{pqrsuv})^u (\vec{r} \cdot \vec{\sigma}'_{pqrsuv})^v \tag{10}$$

where the misalignment vectors $\vec{\sigma}_{pqrsuv}$ and $\vec{\sigma}'_{pqrsuv}$ are in principle different for each term in the series expansion. The expansion order is defined using the powers of $\vec{\rho}$ and $\vec{r}$ and $\vec{\sigma}$ combined. For order j we keep the terms that satisfy:

$$p + q + s + u + v = j + 1 \tag{11}$$

This implies that the radial order $n = 2p + s + u$ and the azimuthal order $m = s + u$ satisfy:

$$n + m = 2p + 2s + 2u \leq 2(p + q + s + u + v) = 2(j+1) \tag{12}$$

The combinatorial problem of finding integer values $\{p, q, s, u, v\}$ that satisfy Eq. (11) is the same as dividing $j + 1$ identical balls over $k = 5$ ordered bins. There are:

$$\frac{(j+k)!}{(j+1)!\,(k-1)!} = \frac{(j+5)!}{(j+1)!\,4!}$$

different terms (5, 15, 35, 140, … terms for $j = 0, 1, 2, 3, …$). Of these terms we can ignore the piston terms with radial order $2p + s + u = 0$. Then $p = s = u = 0$ and the combinatorial problem boils down to distributing $j + 1$ identical balls over $k = 2$ unordered bins, which has $j + 2$ different possibilities. We can also ignore the terms linear in $\vec{\rho}$, corresponding to tip/tilt that possibly depends on field coordinates, i.e. distortion. As distortion cannot be assessed without a reference object, which is usually absent, we will also ignore these terms. Now $2p + s + u = 1$ (terms linear in ρ), which has as solution $p = 0$ and either $s = 0$ and $u = 1$ or $s = 1$ and $u = 0$. There are $2(j + 1)$ of these terms. These restrictions reduce the number of independent terms to 1, 8, 25, 127, … for $j = 0, 1, 2, 3, …$ The number of independent parameters associated with these invariants is 1 if $u = v = 0$ (the pre-factor α_{pqsuv}), 2 if $u > 0$ and $v = 0$ (the misalignment vector components σ_x and σ_y) or $u = 0$ and $v > 0$ (the misalignment vector components σ'_x and σ'_y), and 4 if $u > 0$ and $v > 0$ (the misalignment vector components σ_x, σ_y, σ'_x and σ'_y), where the pre-factor α_{pqsuv} can be absorbed into the misalignment vectors in the cases where either $u > 0$ or $v > 0$ or both. Due to symmetry relations the actual number of independent parameters is actually less. Supplementary Table 2 summarizes all appearing terms up to and including order $j = 2$.

To lowest order $j = 0$ the aberration function only consists of defocus:

$$W = \alpha\rho^2 \tag{13}$$

To first order $j = 1$ we find an aberration function that is a linear combination of the 8 additional invariants listed in Supplementary Table 2. We recognize one term related to primary spherical aberration, two terms related to primary coma, and the remaining 5 terms related to primary astigmatism and defocus/field curvature. The whole set of terms can be decomposed into Zernike modes:

$$W = \sum_{n+|m|\leq 4} A_{nm}(\vec{r})Z_{nm}(\vec{\rho}) \tag{14}$$

Where we list, for the sake of completeness, the appearing Zernike modes:

$$Z_{20}(\vec{\rho}) = 2\rho^2 - 1$$

$$Z_{22}(\vec{\rho}) = \rho^2 \cos(2\theta)$$

$$Z_{2-2}(\vec{\rho}) = \rho^2 \sin(2\theta)$$

$$Z_{31}(\vec{\rho}) = (3\rho^3 - 2\rho)\cos\theta \tag{15}$$

$$Z_{3-1}(\vec{\rho}) = (3\rho^3 - 2\rho)\sin\theta$$

$$Z_{40}(\vec{\rho}) = 6\rho^4 - 6\rho^2 + 1$$

The participating terms per radial order n can be rewritten as follows. For primary spherical aberration we find (by adding an appropriate constant term and term in ρ^2):

$$W_4 = \alpha_s Z_{40}(\vec{\rho}) \tag{16}$$

so that:

$$A_{40} = \alpha_s \tag{17}$$

For primary coma we find:

$$W_3 = 3\alpha_c \rho^2 (\vec{\rho} \cdot \vec{r}) + 3\rho^2(\vec{\rho} \cdot \vec{\sigma}_c) = 3(\sigma_{cx} + \alpha_c x)\rho^3 \cos\theta + 3(\sigma_{cy} + \alpha_c y)\rho^3 \sin\theta \tag{18}$$

Adding tip/tilt (that can be added/subtracted at will since we will ignore this in the end anyway) gives Zernike coma coefficients:

$$A_{31} = \sigma_{cx} + \alpha_c x \tag{19}$$

$$A_{3-1} = \sigma_{cy} + \alpha_c y \tag{20}$$

For primary astigmatism we can rearrange terms to obtain:

$$W_{2a} = \alpha_a (2(\vec{\rho} \cdot \vec{r})^2 - \rho^2 r^2) + 2(\vec{\rho} \cdot \vec{r})(\vec{\rho} \cdot \vec{\sigma}_a) - \rho^2(\vec{r} \cdot \vec{\sigma}_a)^2 + 2(\vec{\rho} \cdot \vec{\tau})^2 - \rho^2 \tau^2 \tag{21}$$

where $\vec{\sigma}_a$ and $\vec{\tau}$ are misalignment vectors. The first term may be rewritten as:

$$\begin{aligned}
2(\vec{\rho} \cdot \vec{r})^2 - \rho^2 r^2 &= \rho^2 r^2 (2\cos(\theta - \psi)^2 - 1) = \rho^2 r^2 \cos(2\theta - 2\psi) \\
&= r^2 \cos(2\psi)\, \rho^2 \cos(2\theta) + r^2 \sin(2\psi)\, \rho^2 \sin(2\theta) \\
&= (x^2 - y^2)\rho^2 \cos(2\theta) + 2xy\, \rho^2 \sin(2\theta)
\end{aligned} \tag{22}$$

Similarly:

$$2(\vec{\rho} \cdot \vec{\tau})^2 - \rho^2 \tau^2 = (\tau_x{}^2 - \tau_y{}^2)\rho^2 \cos(2\theta) + 2\tau_x \tau_y \rho^2 \sin(2\theta) \tag{23}$$

and:

$$2(\vec{\rho} \cdot \vec{r})(\vec{\rho} \cdot \vec{\sigma}_a) - \rho^2(\vec{r} \cdot \vec{\sigma}_a)^2 = (\sigma_{ax} x - \sigma_{ay} y)\rho^2 \cos(2\theta) + (\sigma_{ay} x + \sigma_{ax} y)\rho^2 \sin(2\theta) \tag{24}$$

With $\beta_{22} = \tau_x{}^2 - \tau_y{}^2$ and $\beta_{2-2} = 2\tau_x \tau_y$ this gives Zernike astigmatism coefficients:

$$A_{22} = \beta_{22} + \sigma_{ax} x - \sigma_{ay} y + \alpha_a(x^2 - y^2) \tag{25}$$

$$A_{2-2} = \beta_{2-2} + \sigma_{ay} x + \sigma_{ax} y + 2\alpha_a xy \tag{26}$$

For defocus/field curvature, finally, we find:

$$W_{2d} = 2(\alpha_d r^2 + \vec{\sigma}_d \cdot \vec{r} + \beta)\rho^2 \tag{27}$$

Adding piston we find that:

$$A_{20} = \beta_{20} + \sigma_{dx} x + \sigma_{dy} y + \alpha_d(x^2 + y^2) \tag{28}$$

We can now count the number of parameters that describe the dependency of aberrations across the FOV. There are 4 parameters for defocus/field curvature (β_{20}, $\vec{\sigma}_d$, α_d), 5 parameters for astigmatism ($\beta_{22}, \beta_{2-2}, \vec{\sigma}_a, \alpha_a$), 3 parameters for coma ($\vec{\sigma}_c, \alpha_c$), and 1 for spherical aberration (α_s), giving a total of 13 parameters.

37

As all the aberration coefficients depend on the field coordinates up to 2nd order we can decompose them in terms of the Legendre polynomial products $Q_{00}(x,y)$, $Q_{10}(x,y)$, $Q_{01}(x,y)$, $Q_{20}(x,y)$, $Q_{11}(x,y)$, $Q_{02}(x,y)$. This finally gives the Legendre NAT coefficients as:

$$\gamma_{0020} = \beta_{20} + \frac{1}{3}\alpha_d$$

$$\gamma_{1020} = \sigma_{dx}$$

$$\gamma_{0120} = \sigma_{dy}$$

$$\gamma_{2020} = \gamma_{0220} = \frac{2}{3}\alpha_d$$

$$\gamma_{0022} = \beta_{22}$$

$$\gamma_{002-2} = \beta_{2-2}$$

$$\gamma_{1022} = \gamma_{012-2} = \sigma_{ax} \tag{29}$$

$$\gamma_{0122} = -\gamma_{102-2} = -\sigma_{ay}$$

$$\gamma_{2022} = \gamma_{0222} = \frac{1}{3}\gamma_{112-2} = \frac{2}{3}\alpha_a$$

$$\gamma_{0031} = \sigma_{cx}$$

$$\gamma_{003-1} = \sigma_{dy}$$

$$\gamma_{1031} = \gamma_{013-1} = \alpha_c$$

$$\gamma_{0040} = \alpha_s$$

We see that there is a total of 19 non-zero Legendre NAT coefficients, that satisfy 6 symmetry relations, implying that there are 13 free parameters.

In order to link the 13 fit parameters to the system RMS aberration value we use the parametrization $\bar{\gamma}_k$ for $k = 1,2 \dots, N_\gamma$ with $N_\gamma = 13$, and:

$$\gamma_{0020} = \bar{\gamma}_1$$

$$\gamma_{1020} = \sqrt{3}\,\bar{\gamma}_2$$

$$\gamma_{0120} = \sqrt{3}\,\bar{\gamma}_3$$

$$\gamma_{2020} = \gamma_{0220} = \sqrt{5}\,\bar{\gamma}_4$$

$$\gamma_{0022} = \bar{\gamma}_5$$

$$\gamma_{002-2} = \bar{\gamma}_6$$

$$\gamma_{1022} = \gamma_{012-2} = \sqrt{3}\,\bar{\gamma}_7 \tag{30}$$

$$\gamma_{0122} = -\gamma_{102-2} = -\sqrt{3}\,\bar{\gamma}_8$$

38

$$\gamma_{2022} = -\gamma_{0222} = \frac{1}{3}\gamma_{112-2} = \sqrt{\frac{15}{7}}\,\bar{\gamma}_9$$

$$\gamma_{0031} = \bar{\gamma}_{10}$$

$$\gamma_{003-1} = \bar{\gamma}_{11}$$

$$\gamma_{1031} = \gamma_{013-1} = \sqrt{3}\bar{\gamma}_{12}$$

$$\gamma_{0040} = \bar{\gamma}_{13}$$

With this parametrization the system RMS aberration is given by:

$$W_{SYS}^2 = \sum_{k=1}^{N_\gamma} M_k\,\bar{\gamma}_k^2 \tag{31}$$

Here, M_k gives the multiplicity factor of each independent $\bar{\gamma}_k$, i.e. how many different Legendre NAT terms are proportional to $\bar{\gamma}_k$. In this case $M_9 = 3$, $M_4 = M_7 = M_8 = M_{12} = 2$, and all other $M_k = 1$.

NAT for 2nd order aberration theory

In this case we have quite a bit more terms to cover, as there is an additional number of 25 invariants that have to be taken into account (see Supplementary Table 2). These invariants give rise to a total of 48 additional parameters, giving a total of 13+48 = 61 parameters. A careful analysis reveals that there is some redundancy in the invariants, reducing the number of free parameters to 56. The additional Zernike modes that appear up to this order are:

$$Z_{33}(\vec{\rho}) = \rho^3 \cos(3\theta)$$

$$Z_{3-3}(\vec{\rho}) = \rho^3 \sin(3\theta)$$

$$Z_{42}(\vec{\rho}) = (4\rho^4 - 3\rho^2) \cos(2\theta)$$

$$Z_{4-2}(\vec{\rho}) = (4\rho^4 - 3\rho^2) \sin(2\theta) \tag{32}$$

$$Z_{51}(\vec{\rho}) = (10\rho^5 - 12\rho^3 + 3\rho) \cos\theta$$

$$Z_{5-1}(\vec{\rho}) = (10\rho^5 - 12\rho^3 + 3\rho) \sin\theta$$

$$Z_{60}(\vec{\rho}) = 20\rho^6 - 30\rho^4 + 12\rho^2 - 1$$

The field coordinates are expanded up to 4th order, implying we need the Legendre polynomial products $Q_{30}(x,y)$, $Q_{21}(x,y)$, $Q_{12}(x,y)$, $Q_{03}(x,y)$, $Q_{40}(x,y)$, $Q_{31}(x,y)$, $Q_{22}(x,y)$, $Q_{13}(x,y)$, and $Q_{04}(x,y)$ in addition to the first order aberration expansion. The appearing gamma NAT-coefficients and their symmetry relations can be derived similarly as for the first order expansion. The derivation of these is cumbersome and has been done with the help of Mathematica (See funextra/gammaNATcoefs.nb in Vectorfit [Vectorfit]). First the set of 9 existing invariants up to order $j = 1$ is expanded with 25 new relevant invariants at order $j = 2$ to a total of 34 invariants. Next, this set is divided into the order n of the pupil coordinates (see Supplementary Table 3). This gives 15 invariants with 29 parameters for $n = 2$ (defocus and primary astigmatism), 10 invariants with 19 parameters for $n = 3$ (primary coma and trefoil), 6 invariants with 9 parameters for $n = 4$ (primary

spherical aberration and secondary astigmatism), 2 invariants with 3 parameters for $n = 5$ (secondary coma), and 1 invariant with 1 parameter for $n = 6$ (secondary spherical aberration). For each of these groups the invariants of Supplementary Table 2 are expanded in Legendre polynomial/Zernike mode products to find the set of non-zero Legendre NAT coefficients and to deduce the symmetry relations between these coefficients. This analysis shows that the number of truly free parameters is less than 61. The coma associated first order invariant $\rho^2(\vec{\rho} \cdot \vec{r})$ can be absorbed into the second order invariant $(\vec{\rho} \cdot \vec{r})(\vec{\rho} \cdot \vec{\sigma})(\vec{r} \cdot \vec{\sigma})$, reducing the number of free parameters for the third order aberrations from 19 to 18. The defocus and astigmatism associated first order invariants $\rho^2(\vec{r} \cdot \vec{\sigma})$, $\rho^2 r^2$, and $(\vec{\rho} \cdot \vec{r})^2$ can likewise be absorbed into second order invariants, reducing the number of free parameters for the second order aberrations from 29 to 25. The final list of non-zero gamma NAT coefficients, grouped per aberration type, and expressed in the remaining total of $N_\gamma = 56$ free parameters $\bar{\gamma}_k$ ($k = 1, \dots, N_\gamma$) are listed in the following.

Second order terms (defocus, primary astigmatism): there are 15 invariants with 25 free parameters that give rise to 15 Legendre product polynomials for defocus and each type of primary astigmatism, so 45 terms in total, 8 of these are zero, and there are 12 symmetry relations, leading to 25 independent terms remaining:

$$\gamma_{0020} = \bar{\gamma}_1$$

$$\gamma_{0022} = \bar{\gamma}_2$$

$$\gamma_{002-2} = \bar{\gamma}_3$$

$$\gamma_{1020} = \sqrt{3}\,\bar{\gamma}_4$$

$$\gamma_{0120} = \sqrt{3}\,\bar{\gamma}_5$$

$$\gamma_{1022} = \sqrt{3}\,\bar{\gamma}_6$$

$$\gamma_{0122} = \sqrt{3}\,\bar{\gamma}_7$$

$$\gamma_{102-2} = \sqrt{3}\,\bar{\gamma}_8$$

$$\gamma_{012-2} = \sqrt{3}\,\bar{\gamma}_9$$

$$\gamma_{2020} = \sqrt{5}\,\bar{\gamma}_{10}$$

$$\gamma_{1120} = 3\,\bar{\gamma}_{11}$$

$$\gamma_{0220} = \sqrt{5}\,\bar{\gamma}_{12}$$

$$\gamma_{2022} = -\gamma_{0222} = \sqrt{5}\,\bar{\gamma}_{13} \tag{33}$$

$$\gamma_{1122} = 3\,\bar{\gamma}_{14}$$

$$\gamma_{202-2} = \sqrt{5}\,\bar{\gamma}_{15}$$

$$\gamma_{112-2} = 3\,\bar{\gamma}_{16}$$

$$\gamma_{022-2} = \sqrt{5}\,\bar{\gamma}_{17}$$

$$\frac{1}{3}\gamma_{3020} = \frac{1}{5}\gamma_{1220} = \sqrt{\frac{21}{31}}\,\bar{\gamma}_{18}$$

$$\frac{1}{5}\gamma_{2120} = \frac{1}{3}\gamma_{0320} = \sqrt{\frac{21}{31}}\,\bar{\gamma}_{19}$$

$$\gamma_{3022} + \frac{1}{5}\gamma_{1222} = \gamma_{032-2} + \frac{1}{5}\gamma_{212-2} = \sqrt{\frac{672}{377}}\,\bar{\gamma}_{20}$$

$$\frac{1}{3}\gamma_{3022} - \frac{1}{5}\gamma_{1222} = -\frac{1}{3}\gamma_{032-2} + \frac{1}{5}\gamma_{212-2} = \sqrt{\frac{672}{377}}\,\bar{\gamma}_{21}$$

$$\gamma_{0322} + \frac{1}{5}\gamma_{2122} = -\gamma_{302-2} - \frac{1}{5}\gamma_{122-2} = \sqrt{\frac{672}{377}}\,\bar{\gamma}_{22}$$

$$\frac{1}{3}\gamma_{0322} - \frac{1}{5}\gamma_{2122} = \frac{1}{3}\gamma_{302-2} - \frac{1}{5}\gamma_{122-2} = \sqrt{\frac{672}{377}}\,\bar{\gamma}_{23}$$

$$\frac{1}{9}\gamma_{4020} = -\frac{1}{9}\gamma_{0420} = \frac{1}{35}\gamma_{2220} = \sqrt{\frac{3}{67}}\,\bar{\gamma}_{24}$$

$$\gamma_{4022} = -\gamma_{0422} = \frac{1}{7}\gamma_{312-2} = \frac{1}{7}\gamma_{132-2} = \sqrt{\frac{9}{11}}\,\bar{\gamma}_{25}$$

Symmetry relations with combined NAT coefficients can be inverted to:

$$\gamma_{3022} = \frac{3}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{20} + \frac{3}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{21}$$

$$\gamma_{1222} = \frac{5}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{20} - \frac{15}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{21}$$

$$\gamma_{032-2} = \frac{3}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{20} - \frac{3}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{21}$$

$$\gamma_{212-2} = \frac{5}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{20} + \frac{15}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{21}$$

$$\gamma_{0322} = \frac{3}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{22} + \frac{3}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{23}$$

$$\gamma_{2122} = \frac{15}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{22} - \frac{5}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{23}$$

$$\gamma_{302-2} = -\frac{3}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{22} + \frac{3}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{23}$$

$$\gamma_{122-2} = -\frac{15}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{22} - \frac{5}{4}\sqrt{\frac{672}{377}}\,\bar{\gamma}_{23}$$

Third order terms (primary coma and trefoil): there are 10 invariants with 18 free parameters that give rise to 10 Legendre product polynomials for each type of primary coma and trefoil, so 40 terms in total, 8 of these are zero, and there are 14 symmetry relations, leading to 18 independent terms remaining. The 11 independent coefficients related to primary coma are:

$$\gamma_{0031} = \bar{\gamma}_{26}$$

$$\gamma_{003-1} = \bar{\gamma}_{27}$$

$$\gamma_{1031} = \sqrt{3}\,\bar{\gamma}_{28}$$

$$\gamma_{013-1} = \sqrt{3}\,\bar{\gamma}_{29}$$

$$\gamma_{0131} = \sqrt{3}\,\bar{\gamma}_{30}$$

$$\gamma_{103-1} = \sqrt{3}\,\bar{\gamma}_{31}$$

$$\frac{1}{2}(\gamma_{2031} + \gamma_{0231}) = \sqrt{5}\,\bar{\gamma}_{32} \tag{34}$$

$$\frac{1}{2}(\gamma_{2031} - \gamma_{0231}) = \frac{1}{3}\gamma_{113-1} = \sqrt{\frac{15}{7}}\,\bar{\gamma}_{33}$$

$$\frac{1}{2}(\gamma_{023-1} + \gamma_{203-1}) = \sqrt{5}\,\bar{\gamma}_{34}$$

$$\frac{1}{2}(\gamma_{023-1} - \gamma_{203-1}) = \frac{1}{3}\gamma_{1131} = \sqrt{\frac{15}{7}}\,\bar{\gamma}_{35}$$

$$\frac{1}{3}\gamma_{3031} = \frac{1}{3}\gamma_{033-1} = \frac{1}{5}\gamma_{1231} = \frac{1}{5}\gamma_{213-1} = \sqrt{\frac{42}{97}}\bar{\gamma}_{36}$$

Symmetry relations with combined NAT coefficients can be inverted to:

$$\gamma_{2031} = \sqrt{5}\,\bar{\gamma}_{32} + \sqrt{\frac{15}{7}}\,\bar{\gamma}_{33}$$

$$\gamma_{0231} = \sqrt{5}\,\bar{\gamma}_{32} - \sqrt{\frac{15}{7}}\,\bar{\gamma}_{33}$$

$$\gamma_{023-1} = \sqrt{5}\,\bar{\gamma}_{34} + \sqrt{\frac{15}{7}}\,\bar{\gamma}_{35}$$

$$\gamma_{203-1} = \sqrt{5}\,\bar{\gamma}_{34} - \sqrt{\frac{15}{7}}\,\bar{\gamma}_{35}$$

The 7 independent coefficients related to trefoil are:

$$\gamma_{0033} = \bar{\gamma}_{37}$$

$$\gamma_{003-3} = \bar{\gamma}_{38}$$

$$\gamma_{0133} = -\gamma_{103-3} = \sqrt{3}\,\bar{\gamma}_{39}$$

$$\frac{1}{2}\left(\gamma_{1033} + \gamma_{013-3}\right) = \sqrt{3}\,\bar{\gamma}_{40} \tag{35}$$

$$\gamma_{2033} = -\gamma_{0233} = \frac{1}{3}\gamma_{113-3} = \sqrt{\frac{15}{7}}\,\bar{\gamma}_{41}$$

$$-\gamma_{203-3} = \gamma_{023-3} = \frac{1}{3}\gamma_{1133} = \sqrt{\frac{15}{7}}\,\bar{\gamma}_{42}$$

$$\gamma_{3033} = -\gamma_{033-3} = -\frac{1}{5}\gamma_{1233} = \frac{1}{5}\gamma_{213-3} = \frac{1}{2}\left(\gamma_{013-3} - \gamma_{1033}\right) = \sqrt{\frac{21}{22}}\,\bar{\gamma}_{43}$$

Symmetry relations with combined NAT coefficients can be inverted to:

$$\gamma_{013-3} = \sqrt{3}\,\bar{\gamma}_{40} + \sqrt{\frac{21}{22}}\,\bar{\gamma}_{43}$$

$$\gamma_{1033} = \sqrt{3}\,\bar{\gamma}_{40} - \sqrt{\frac{21}{22}}\,\bar{\gamma}_{43}$$

Fourth order terms (primary spherical aberration, secondary astigmatism): there are 6 invariants with 9 free parameters that give rise to 6 Legendre product polynomials for primary spherical aberration and each type of secondary astigmatism, so 18 different terms in total. Of these 4 are zero, there are 5 symmetry relations, leading to 9 independent terms remaining:

43

$$\gamma_{0040} = \bar{\gamma}_{44}$$

$$\gamma_{0042} = \bar{\gamma}_{45}$$

$$\gamma_{004-2} = \bar{\gamma}_{46}$$

$$\gamma_{1040} = \sqrt{3}\, \bar{\gamma}_{47}$$

$$\gamma_{0140} = \sqrt{3}\, \bar{\gamma}_{48} \tag{36}$$

$$\gamma_{1042} = \gamma_{014-2} = \sqrt{3}\, \bar{\gamma}_{49}$$

$$\gamma_{0142} = -\gamma_{104-2} = \sqrt{3}\, \bar{\gamma}_{50}$$

$$\gamma_{2040} = \gamma_{0240} = \sqrt{5}\, \bar{\gamma}_{51}$$

$$\gamma_{2042} = -\gamma_{0242} = \frac{1}{3}\gamma_{114-2} = \sqrt{\frac{15}{7}}\, \bar{\gamma}_{52}$$

Fifth order terms (secondary coma): there are 2 invariants with 3 free parameters that give rise to 3 Legendre product polynomials for each type of secondary coma, so 6 different terms in total. Of these 2 are zero, there is 1 symmetry relation, leading to 3 independent terms remaining, with coefficients:

$$\gamma_{0051} = \bar{\gamma}_{53}$$

$$\gamma_{005-1} = \bar{\gamma}_{54} \tag{37}$$

$$\gamma_{1051} = \gamma_{015-1} = \sqrt{3}\, \bar{\gamma}_{55}$$

Sixth order term (secondary spherical aberration): there is only 1 invariant and 1 free parameter, that gives rise to only 1 Legendre product polynomial with coefficient:

$$\gamma_{0060} = \bar{\gamma}_{56} \tag{38}$$

With this parameterization Eq. (31) still holds, where the new multiplicity factors are $M_k = 2$ for $k = 13,18,19,32,34,39,40,49,50,51,55$, $M_k = 3$ for $k = 24,33,35,41,42,52$, $M_k = 4$ for $k = 20,21,22,23,25,36$, $M_k = 6$ for $k = 43$, and $M_k = 1$ for all other k.

An alternative to the current approach to find the redundant set of NAT coefficients for higher orders is to arrange all participating gamma NAT coefficients in a vector x of length N, and rewrite the set of K symmetry relations as:

$$Bx = 0 \tag{39}$$

where B is a $K \times N$ matrix. The projection operator on the non-null space is:

$$P = I - B^T(BB^T)^{-1}B \tag{40}$$

as $BP = B - B = 0$. Diagonalization of P leads to a set of eigenvectors $y^{(l)}$, $l = 1,2,\ldots,N-K$ with eigenvalue 1, and in the optimization we could use an expansion of the gamma NAT coefficient vector x in terms of these eigenvector:

44

$$x = \sum_{l=1}^{N-K} x_l y^{(l)} \tag{41}$$

where the x_l can subsequently be used as independent optimization parameters.

Supplementary Table 1: Fitting speed of Vectorial PSF

Fits in 10^3 fits/s for localizing a total of 1 million emitters that were simulated with the vectorial PSF model as described in the method section. The ROI size for 2D fitting is 7×7 and for 3D fitting 17×17. The GPU computations on the Nvidia A100 were performed using the DelftBlue supercomputer [DHPC 2024]. The reported times include the computation of one OTF, assuming aberrations are constant over the FOV. When a grid of 20x20 OTFs is used, an additional 75 seconds is required for the computation.

	CPU using 1 core Intel Xeon w5-2455X		CPU using all 12 cores Intel Xeon w5-2455X		Desktop GPU Nvidia RTX A4000		High end HPC GPU Nvidia A100	
	Direct	OTF	Direct	OTF	Direct	OTF	Direct	OTF
2D	0.6	1.5	4.0	8.7	17.4	61.5	50.3	100.0
3D	0.2	0.5	1.4	2.7	6.3	6.8	14.8	10.1

Supplementary Table 2: NAT invariants and number of parameters

The table shows a list of all possible invariants of the form $(\rho^2)^p (r^2)^q (\vec{\rho} \cdot \vec{r})^s (\vec{\rho} \cdot \vec{\sigma})^u (\vec{r} \cdot \vec{\sigma}')^v$ as defined in Eq. (10) for NAT order $j = p + q + s + u + v - 1$. The terms with overall order in the pupil coordinate $2p + s + u$ equal to 0 or 1 correspond to piston or wavefront tip/tilt and are ignored. These terms are marked with an asterisk.

j	p	q	s	u	v	invariant	$2p+s+u$	$2q+s+v$	#pars
0	1	0	0	0	0	ρ^2	2	0	1
	0	1	0	0	0	r^2	0*	2	1
	0	0	1	0	0	$\vec{\rho}\cdot\vec{r}$	1*	1	1
	0	0	0	1	0	$\vec{\rho}\cdot\vec{\sigma}$	1*	0	2
	0	0	0	0	1	$\vec{r}\cdot\vec{\sigma}$	0*	1	2
1	2	0	0	0	0	ρ^4	4	0	1
	0	2	0	0	0	r^4	0*	4	1
	0	0	2	0	0	$(\vec{\rho}\cdot\vec{r})^2$	2	2	1
	0	0	0	2	0	$(\vec{\rho}\cdot\vec{\sigma})^2$	2	0	2
	0	0	0	0	2	$(\vec{r}\cdot\vec{\sigma})^2$	0*	2	2
	1	1	0	0	0	$\rho^2 r^2$	2	2	1
	1	0	1	0	0	$\rho^2(\vec{\rho}\cdot\vec{r})$	3	1	1
	1	0	0	1	0	$\rho^2(\vec{\rho}\cdot\vec{\sigma})$	3	0	2
	1	0	0	0	1	$\rho^2(\vec{r}\cdot\vec{\sigma})$	2	1	2
	0	1	1	0	0	$r^2(\vec{\rho}\cdot\vec{r})$	1*	3	1
	0	1	0	1	0	$r^2(\vec{\rho}\cdot\vec{\sigma})$	1*	2	2
	0	1	0	0	1	$r^2(\vec{r}\cdot\vec{\sigma})$	0*	3	2
	0	0	1	1	0	$(\vec{\rho}\cdot\vec{r})(\vec{\rho}\cdot\vec{\sigma})$	2	1	2
	0	0	1	0	1	$(\vec{\rho}\cdot\vec{r})(\vec{r}\cdot\vec{\sigma})$	1*	2	2
	0	0	0	1	1	$(\vec{\rho}\cdot\vec{\sigma})(\vec{r}\cdot\vec{\sigma})$	1*	1	4
2	3	0	0	0	0	ρ^6	6	0	1
	0	3	0	0	0	r^6	0*	6	1
	0	0	3	0	0	$(\vec{\rho}\cdot\vec{r})^3$	3	3	1
	0	0	0	3	0	$(\vec{\rho}\cdot\vec{\sigma})^3$	3	0	2
	0	0	0	0	3	$(\vec{r}\cdot\vec{\sigma})^3$	0*	3	2
	2	1	0	0	0	$\rho^4 r^2$	4	2	1
	2	0	1	0	0	$\rho^4(\vec{\rho}\cdot\vec{r})$	5	1	1
	2	0	0	1	0	$\rho^4(\vec{\rho}\cdot\vec{\sigma})$	5	0	2
	2	0	0	0	1	$\rho^4(\vec{r}\cdot\vec{\sigma})$	4	1	2
	1	2	0	0	0	$\rho^2 r^4$	2	4	1
	0	2	1	0	0	$r^4(\vec{\rho}\cdot\vec{r})$	1*	5	1
	0	2	0	1	0	$r^4(\vec{\rho}\cdot\vec{\sigma})$	1*	4	2
	0	2	0	0	1	$r^4(\vec{r}\cdot\vec{\sigma})$	0*	5	2
	1	0	2	0	0	$\rho^2(\vec{\rho}\cdot\vec{r})^2$	4	2	1
	0	1	2	0	0	$r^2(\vec{\rho}\cdot\vec{r})^2$	2	4	1
	0	0	2	1	0	$(\vec{\rho}\cdot\vec{r})^2(\vec{\rho}\cdot\vec{\sigma})$	3	2	2
	0	0	2	0	1	$(\vec{\rho}\cdot\vec{r})^2(\vec{r}\cdot\vec{\sigma})$	2	3	2
	1	0	0	2	0	$\rho^2(\vec{\rho}\cdot\vec{\sigma})^2$	4	0	2
	0	1	0	2	0	$r^2(\vec{\rho}\cdot\vec{\sigma})^2$	2	2	2
	0	0	1	2	0	$(\vec{\rho}\cdot\vec{r})(\vec{\rho}\cdot\vec{\sigma})^2$	3	1	2
	0	0	0	2	1	$(\vec{\rho}\cdot\vec{\sigma})^2(\vec{r}\cdot\vec{\sigma})$	2	1	4
	1	0	0	0	2	$\rho^2(\vec{r}\cdot\vec{\sigma})^2$	2	2	2
	0	1	0	0	2	$r^2(\vec{r}\cdot\vec{\sigma})^2$	0*	3	2
	0	0	1	0	2	$(\vec{\rho}\cdot\vec{r})(\vec{r}\cdot\vec{\sigma})^2$	1*	3	2
	0	0	0	1	2	$(\vec{\rho}\cdot\vec{\sigma})(\vec{r}\cdot\vec{\sigma})^2$	1*	2	4
	1	1	1	0	0	$\rho^2 r^2(\vec{\rho}\cdot\vec{r})$	3	3	1
	1	1	0	1	0	$\rho^2 r^2(\vec{\rho}\cdot\vec{\sigma})$	3	2	2
	1	0	1	1	0	$\rho^2(\vec{\rho}\cdot\vec{r})(\vec{\rho}\cdot\vec{\sigma})$	4	1	2
	0	1	1	1	0	$r^2(\vec{\rho}\cdot\vec{r})(\vec{\rho}\cdot\vec{\sigma})$	2	3	2
	1	1	0	0	1	$\rho^2 r^2(\vec{r}\cdot\vec{\sigma})$	2	3	2
	1	0	1	0	1	$\rho^2(\vec{\rho}\cdot\vec{r})(\vec{r}\cdot\vec{\sigma})$	3	2	2
	0	1	1	0	1	$r^2(\vec{\rho}\cdot\vec{r})(\vec{r}\cdot\vec{\sigma})$	1*	4	2
	1	0	0	1	1	$\rho^2(\vec{\rho}\cdot\vec{\sigma})(\vec{r}\cdot\vec{\sigma})$	3	1	4
	0	1	0	1	1	$r^2(\vec{\rho}\cdot\vec{\sigma})(\vec{r}\cdot\vec{\sigma})$	1*	3	4
	0	0	1	1	1	$(\vec{\rho}\cdot\vec{r})(\vec{\rho}\cdot\vec{\sigma})(\vec{r}\cdot\vec{\sigma})$	2	2	4

Supplementary Table 3: Number of free parameters per radial aberration order

The table shows the number of invariants and associated number of parameters per radial aberration order n and per NAT order j as deduced from Supplementary Table 2. Due to redundancy in the lower order invariants when going to order $j = 2$, the number of free parameters is reduced.

n	# invariants				# parameters				
	$j=0$	$j=1$	$J=2$	total	$J=0$	$J=1$	$J=2$	total	independent
2	1	5	9	15	1	8	20	29	25
3	0	2	8	10	0	3	16	19	18
4	0	1	5	6	0	1	8	9	9
5	0	0	2	2	0	0	3	3	3
6	0	0	1	1	0	0	1	1	1
total	1	8	25	34	1	12	48	61	56