Supplementary materials for: Aperture interference and the volumetric resolution of light field fluorescence microscopy

Isaac Kauvar ikauvar@stanford.edu
Julie Chang jchang10@stanford.edu
Gordon Wetzstein gordonwz@stanford.edu
Department of Electrical Engineering, Stanford University, 350 Serra Mall, Stanford, CA

Abstract

This document provides supplementary information to 'Aperture interference and the volumetric resolution of light field fluorescence microscopy'. We provide more information about: the derivation of the defocus OTF, additional ALF-ring schemes, frequency space invertibility, the forward model and reconstruction algorithm, and Monte Carlo simulations for calculating noise thresholds. In addition, we use this space to describe our implementation and calibration in more detail and to include an expansion of the results shown in the primary document.

The 3D defocus OTF

In two dimensions, the OTF is simply the 2D autocorrelation of the coherent amplitude transfer function, $ATF(k_x, k_y)$, which is the Fourier transform of the coherent point spread function (the square of which is the incoherent point spread function). Ignoring the phase associated with the pupil function of the microscope objective, $ATF(k_x, k_y)$ is equivalent to the aperture mask function, and thus the 2D OTF is the autocorrelation of the aperture [3, 4].

$$OTF(k_x, k_y) = \mathcal{F}_{x,y}\{PSF(x, y)\} \quad (1)$$
$$= \mathcal{F}_{x,y}\{|\mathcal{F}^{-1}_{k_x,k_y}\{ATF(k_x, k_y)\}|^2\} \quad (2)$$
$$= ATF(k_x, k_y) \star ATF(k_x, k_y) \quad (3)$$

where $PSF(x, y)$ is the incoherent point spread function.

In three dimensions, the OTF is the 3D autocorrelation of the *generalized* aperture [4]. The generalized aperture is derived easily from the depth dependent or defocused aperture, $ATF(k_x, k_y; z)$, which incorporates the phase associated with the propagation of light as it defocuses (which with the Fresnel approximation yields the well known quadratic phase associated with defocus) [3]:

$$ATF(k_x, k_y; z) = ATF(k_x, k_y)e^{i2\pi z\sqrt{k^2-(k_x^2+k_y^2)}} \quad (4)$$
$$\overset{Fresnel}{\approx} ATF(k_x, k_y)e^{i2\pi kz}e^{-i\pi\frac{z}{k}(k_x^2+k_y^2)} \quad (5)$$

The generalized aperture $ATF(k_x, k_y, k_z)$ is the Fourier transform along the axial direction of $ATF(k_x, k_y; z)$. From Eq. 4, using the Fourier shift theorem:

$$ATF(k_x, k_y, k_z) = \int ATF(k_x, k_y; z)e^{-i2\pi zk_z}dz \quad (6)$$
$$= \int ATF(k_x, k_y)e^{i2\pi z(\sqrt{k^2-k_x^2-k_y^2}-k_z)}dz \quad (7)$$
$$= ATF(k_x, k_y)\delta(k_z - \sqrt{k^2 - k_x^2 - k_y^2}) \quad (8)$$

The generalized aperture can be visualized as the projection of the 2D aperture onto the surface of a spherical shell in (k_x, k_y, k_z) space, also known as the Ewald sphere. The radius of the shell is determined by the wavelength of light, and represents the fact that, in order to satisfy Maxwell's equations, light of wavelength λ propagating in a medium of refractive index n is required to have a specific total momentum, $k = \sqrt{k_x^2 + k_y^2 + k_z^2} = 2\pi n/\lambda$.

Thus, while we originally defined the defocus OTF as the 3D Fourier transform of the incoherent depth-dependent point spread function,

$$OTF(k_x, k_y, k_z) = \mathcal{F}_{x,y,z}\{PSF(x, y, z)\} \quad (9)$$
$$= \mathcal{F}_{x,y,z}\{|\mathcal{F}^{-1}_{k_x,k_y,k_z}\{ATF(k_x, k_y, k_z)\}|^2\} \quad (10)$$

by following the same logic that the 2D OTF can be derived as the autocorrelation of $ATF(k_x, k_y)$, it can be shown that the 3D defocus OTF is equivalent to the autocorrelation of

the generalized aperture $ATF(k_x, k_y, k_z)$ [4], as schematized in the top of Fig. 2 of the main paper:

$$OTF(k_x, k_y, k_z) = ATF(k_x, k_y, k_z) \star_3 ATF(k_x, k_y, k_z) \tag{11}$$

ALF-ring

One additional sequence, based on ALF-rings, is what we call the *ALF-ring*. Here, instead of using all of the grid points and rings with varying diameters, we use only one ring, half the diameter of the aperture, centered a quarter-aperture-diameter from the optical axis and rotated to different positions around the optical axis. This simplifies the hardware implementation, while still enabling us to capture the highest achievable axial frequencies, even if some lateral frequency is given up.

Nonoverlapping ALF-ring simulations The ring-based aperture masks could also be designed to avoid any overlaps. This could enable single-snapshot implementations that exclude overlapping aperture supports. In Fig. 1, we perform simulations showing that it is possible to achieve the benefits of ALF, even with a design that never reuses any regions of the aperture during a sequence of aperture masks.

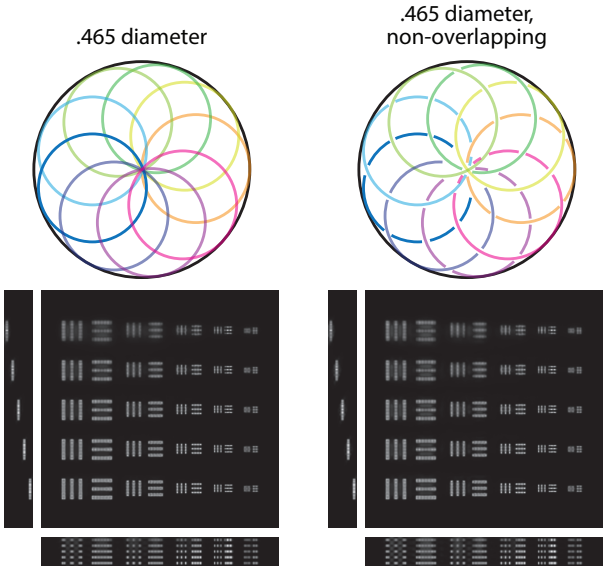

Figure 1. Simulation results using ALF-ring, and ALF-ring where the masks are chosen such that they never overlap (and thus each point in the aperture is open for only a single mask in the sequence). (top) ALF-ring aperture mask patterns, where each color represents a mask opening. (bot) Maximum intensity projections (MIPs) of volume reconstructions of the same test volume as in Fig. 6 of the main paper. The volume (1024 1024 15) consisted of non-overlapping test resolution bars every third slice, with 6.67 μm between each slice.

Frequency space invertibility

Here we derive the normalized frequency domain measurement matrix used in computing the invertibility of an aperture mask sequence, and we discuss the general idea of volumetric invertibility.

Starting with $OTF(\mathbf{k})$ from Eq. 11 in the main paper, we define the transfer function, $H^{(q)}(\mathbf{k})$ for each mask q, by normalizing $OTF^{(q)}$ and then scaling according to the amount of light transmitted through the corresponding aperture mask.

$$H^{(q)}(\mathbf{k}) := \frac{OTF^{(q)}(\mathbf{k})}{OTF^{(q)}(\mathbf{0})} \cdot \frac{\sum(ATF^{(q)}(k_x, k_y))}{\sum(ATF_{NA}(k_x, k_y))} \tag{12}$$

where $ATF^{(q)}$ is the aperture mask profile, and ATF_{NA} represents an open aperture with the maximum numerical aperture. This normalization enables us to make a connection with actual photon counts, as described in the main body of the paper.

Next, define $\tilde{V} \in \mathbb{R}^{n_k}$ to be the vectorized 3D frequency space representation of the specimen discretized into $n_k = n_{kx} \times n_{ky} \times n_{kz}$ elements. Define $B_q \in \mathbb{C}^{n_k \times n_k}$ to be a diagonal matrix form of H_q from Eq. 12. Define $P \in \mathbb{R}^{n_{kx} n_{ky} \times n_k}$ as a matrix operator that sums along k_z for each (k_x, k_y). If N is the number of aperture masks in a sequence, the measurement matrix $M \in \mathbb{C}^{n_{kx} n_{ky} N \times n_k}$ is a stacked matrix of blocks PB_1 to PB_N. We measure $\tilde{I} = M\tilde{V}$ where $\tilde{I}$ is the Fourier transform of the sensor images. As described in the main body of the paper, we then use the singular value decomposition of M to describe the invertibility of the acquired data.

The sequence of aperture masks should sample all possible lateral and axial spatial frequencies, but we also want individual samples that finely tile the 3D frequency space. Our goal is to sample the 3D volume into voxels; similarly, we can think of sampling the 3D frequency space into 'frequency voxels'. In the following, we explore how this can be accomplished with a sequence of aperture masks. As schematized in Fig. S2, we can use the Fourier slice theorem to understand how a 2D sensor samples the defocus OTF. The key message is that k_z values for a given k_x are mixed together; importantly, however, only those k_z values at a given k_x value that are transmitted by the defocus OTF *for that specific aperture mask* are mixed together. Further, as illustrated in the bottom of Fig. 2, sampling k_z values individually is what yields unambiguous volumetric information. This is why a single image taken with a fully open aperture provides no depth information, whereas a sequence of aperture pinhole images does. We note that this is related to concepts from Magnetic Resonance Imaging (MRI), which directly samples in frequency space: in

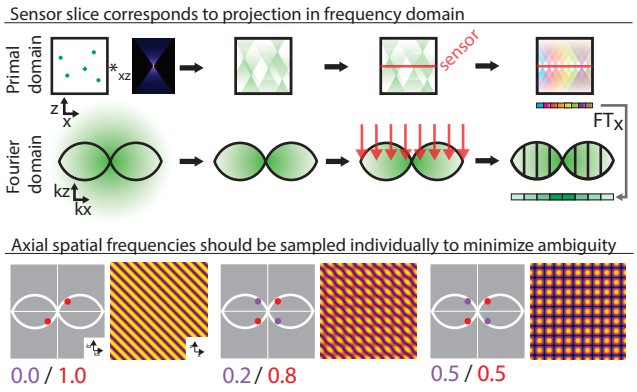

Figure 2. (top) A planar image sensor measures a slice of the 3D volume, convolved in 3D with the 3D PSF. In Fourier space, this is equivalent to multiplying the spectrum of the volume with the defocus OTF, integrating along the direction perpendicular to the sensor, and discretizing along the direction of the sensor. The result is the Fourier transform of the sensor image, which samples part of the 3D spectrum of the volume. In the illustrated case, for each k_x value, all k_z values are blurred together. (bottom) On the left: a fringe pattern in the primal domain is equivalent to two points in the frequency domain that are symmetric about the origin. Sampling a 3D volume is equivalent to knowing the weight assigned to each of these fringe patterns. If two frequency domain components are combined into one measurement, but the ratio of their contributions is unknown, then there will be ambiguity in the recovered volume. This is illustrated by different fringe patterns that could arise from different ratios of each component.

particular, it is well known that increasing the resolution of one's frequency domain measurements increases the axial field of view of the recovered volume [5]. Here too, finer sampling of axial frequency domain measurements enables recovery of information across a larger axial field of view. The conclusion is that we want to choose either aperture masks that transmit a small range of k_z values for each k_x, or choose a sequence of aperture masks such that if they do measure multiple k_z values for each k_x, the measurements of different masks overlap, allowing us to recover fine frequency domain sampling through demultiplexing.

Forward model and reconstruction

Forward model and vectorization The following describes how we computationally merge the information recorded in the images of an aperture mask sequence. The forward model of incoherent fluorescent image formation can be described by a filtering in the pupil plane with the 2D incoherent $OTF_z(k_x, k_y)$ for each depth z followed by an axial projection on to the camera sensor. The incoherent OTF for a given aperture mask with binary, numerical aperture-limited amplitude function $\alpha(k_x, k_y)$ is calculated via the ATF and incoherent PSF:

$$ATF_z(k_x, k_y) = e^{ikz} \exp\left(-i\pi\lambda z(k_x^2 + k_y^2)\right) \alpha(k_x, k_y) H(k_x, k_y) \tag{13}$$

$$PSF_z(x, y) = |\mathcal{F}^{-1}_{k_x,k_y}\{ATF_z(k_x, k_y)\}|^2 \tag{14}$$

$$OTF_z(k_x, k_y) = \mathcal{F}_{x,y}\{PSF_z(x, y)\} \tag{15}$$

where $k = n2\pi/\lambda$ is the magnitude of the wave vector and $H(k_x, k_y) = \sqrt{\cos(\sin^{-1}(\lambda(k_x^2 + k_y^2)^{1/2}))}$ is an apodization function defined by the Abbe sine condition [2].

For a specimen volume with intensity profile $V(x, y, z)$, and aperture mask q with $OTF_z^{(q)}(k_x, k_y)$, the corresponding image on the sensor $I^{(q)}(x, y)$ is described as

$$I^{(q)}(x, y) = \int_z \mathcal{F}^{-1}_{k_x,k_y}\{OTF_z^{(q)}(k_x, k_y)\mathcal{F}_{x,y}\{V(x, y, z)\}\}dz \tag{16}$$

For computing purposes, we work with a model in which the sample volume is discretized into $n_v = n_x \times n_y \times n_z$ voxels, and the camera sensor is (naturally) discretized into $n_p = n_x \times n_y$ pixels. Our goal is to reconstruct a 3D volume from a sequence of Q images corresponding to each of the Q aperture masks with the stochastic forward image formation model:

$$\mathbf{I} \sim \text{Pois}(A\mathbf{x} + \mathbf{b}), \tag{17}$$

where $\mathbf{I} \in \mathbb{R}^{n_p Q}$ is the vectorized stack of sensor images, $\mathbf{x} \in \mathbb{R}^{n_v}$ is the vectorized discrete sample volume, $\mathbf{b} \in \mathbb{R}^{n_p Q}$ is the vectorized background signal, and $A \in \mathbb{R}^{n_p Q \times n_v}$ is the measurement matrix of the system, such that each entry $A_{i+(q-1)n_p,j}$ is the intensity contribution of voxel j onto pixel i of aperture q in the sequence of Q apertures. This model assumes Poisson-distributed sensor measurements, which is appropriate for scientific cameras. In accordance with Eq. 16, A is composed by

$$A = P_z F_b^{-1} K F_x, \tag{18}$$

where $F_x \in \mathbb{C}^{n_v \times n_v}$ takes the 2D discrete Fourier transform (DFT) of each z-slice of the volume, $K \in \mathbb{C}^{n_v Q \times n_v}$ is a stack of diagonal matrices that perform element-wise multiplication with the incoherent OTF for each of the Q aperture masks, $F_b^{-1} \in \mathbb{C}^{n_v Q \times n_v Q}$ is a block diagonal DFT matrix that takes the inverse 2D DFT along k_x and k_x, and $P_z \in \mathbb{R}^{n_p Q \times n_v Q}$ applies a projection along z of the resulting PSF-convolved volume for each aperture mask setting. This composition, with noise added, produces the final real-valued vector $\mathbf{I}$ of the Q 2D sensor images.

Solution of inverse problem A variety of non-blind deconvolution algorithms can be used to recover an optimal

estimate of **x**. We use the common Richardson-Lucy iterative algorithm, which finds the maximum likelihood estimation of the volume **x** given the Poisson-noised measurements **I**, background signal **b**, and known matrix A [1]. We maximize the log-likelihood of **x** given **I** and estimate **b**, a convex optimization problem, via multiplicative gradient descent updates:

$$\mathbf{x}^{(q+1)} = \text{diag}(A^H\mathbf{1})^{-1}\text{diag}[A^H\text{diag}(A\mathbf{x}^{(q)} + \mathbf{b})^{-1}\mathbf{I}]\mathbf{x}^{(q)} \tag{19}$$

where A^H is the adjoint of A. We implement the updates with matrix-free operations of the forward image formation and its adjoint, as described in the next section. The number of iterations we run depends on the size of **x** and **I**.

Adjoint operation for deconvolution The adjoint operation A^H equals the conjugate transpose of matrix A (Eqn. 18):

$$A^H = F_x^{-1}K^H F_b P_z^\intercal \tag{20}$$

where we have used the fact that the conjugate transpose of the DFT matrix is the inverse DFT matrix. $P_z^\intercal \in \mathbb{R}^{n_v Q \times n_p Q}$ can be understood as a back-projection of the sensor image onto all z-slices in the volume, and K^H performs element-wise multiplication with the incoherent OTFs and then sums the contributions to a given voxel from each aperture. This intuitive understanding of A^H is useful because, rather than explicitly compute each entry of A, a prohibitively large matrix, we can apply linear operators that perform the functions equivalent to matrix multiplication by A and A^H.

Monte Carlo simulations of noise

We used the Monte Carlo method to generate multiple instances of Poisson noise for various volume patterns, densities, and photon counts. We found that the noise spectrum was flat with RMS value equal to the square root of the mean image intensity. This equivalence could also be derived theoretically by approximating Poisson noise as signal-dependent, additive, zero-mean Gaussian noise. This finding allowed us to set a uniform noise floor in our frequency domain SNR analysis. Fig. 3 shows the noise spectra along $k_y = 0$ for simulated and averaged noisy sensor images using a ring-shaped mask of a randomly generated volume of point sources.

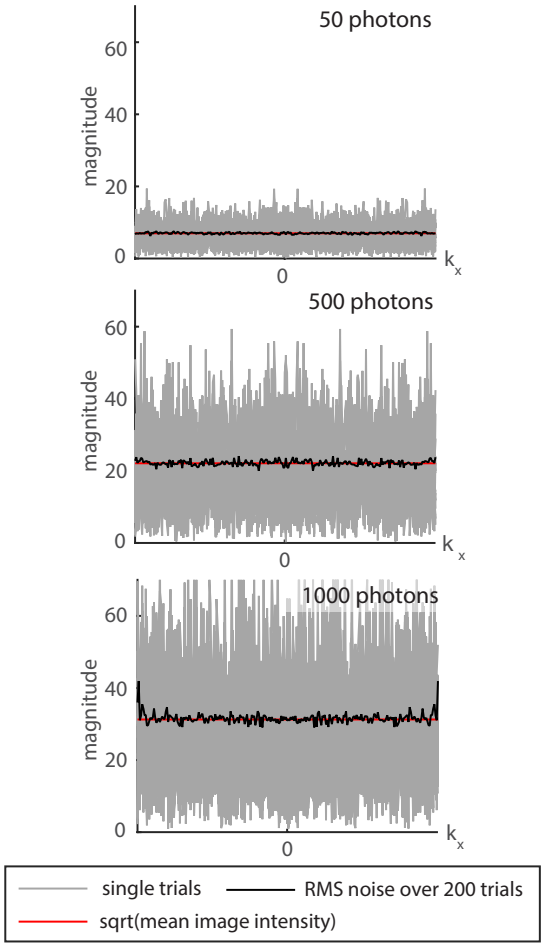

Figure 3. Poisson noise spectra along $k_y = 0$ for different mean image intensities through a ring-shaped aperture mask.

Implementation

Prototype implementation details Here we describe in greater detail the equipment used in the prototype implementation of ALF. We used a commercial Olympus BX-63 microscope with a 20x 0.5 NA Olympus Plan Fluorite objective for all the experiments. We used a 4f-system extension ($f_1 = 150$ mm, $f_2 = 300$ mm) for easy access to a conjugate pupil plane and to add an additional 2x magnification to more clearly see the diffraction-limited resolution. For epi-fluorescence imaging we used a mercury lamp (Olympus U-HGLGPS) with filter cube (Olympus U-FBNA) with excitation filter BP 470-495, dichroic mirror DM 505, and emission filter BP 510-550. Images were taken with the Hamamatsu Photonics ORCA-Flash4.0 V2 digital CMOS camera.

Aperture masks, sized according to the back aperture of the objective, were printed by Fineline printing at 50,800 DPI, mounted on magnetic quick release plates (Thorlabs CP90F), see Fig. 4, and placed at the pupil plane on a motorized rotation mount (Thorlabs PRM1Z8). Ring aperture masks had a ring width of 0.23 mm. This design required use of only a few different printed masks for reproducibly placing the aperture mask at desired grid points. For the resolution chart sample, non-centered masks were rotated in 60 degree increments to 6 different center locations, and an image was taken through each aperture position. For the

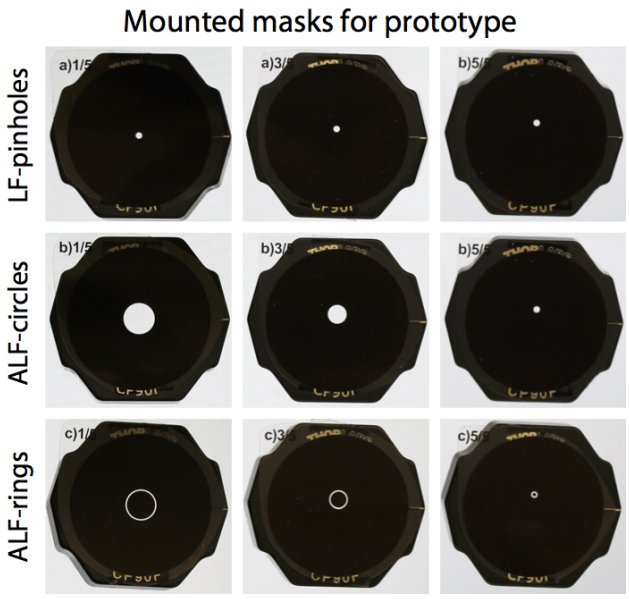

Figure 4. Printed amplitude masks on magnetic mounts used in prototype.

fluorescent bead and pollen samples, non-centered masks were rotated in 30 degree increments to 12 different center locations.

PSF calibration In an ideal setup, we could use the printed aperture mask templates directly in our PSF generation model to find the PSF for each aperture mask. In practice, uncertainties in exact placement of aperture masks, rotation angle of the rotating mount, and alignment of other optical components cause disparities between the simulated and empirical PSF, requiring further calibration of the simulated PSFs. For calibration, we collected a focal stack across +/- 100 μm of a 1 μm green fluorescent bead with each aperture mask positioned at desired rotation angles. First, for each aperture mask, we determined the angle of rotation from the x-axis (perpendicular to the optical axis) by fitting a line to the centroids of the PSFs across defocus depth. Next, we fit the exact radius from the optical axis to the center of mass of the aperture opening based on the rate at which the PSF translated with defocus. Further, we fit a small spherical aberration coefficient d for the optical setup (caused mostly by the lenses of our 4f system and an inability to perfectly set the distances between lenses), which could easily be incorporated at the pupil plane of the forward model. Primary spherical aberration with parameter d was added as a radially symmetric phase mask in the PSF generation model according to:

$$\rho = \sqrt{k_x^2 + k_y^2} \tag{21}$$

$$\phi(d, \rho) = d\sqrt{5}\left(6\rho^4 - 6\rho^2 + 1\right) \tag{22}$$

$$W(k_x, k_y) = \exp(i2\pi\phi(d, \sqrt{k_x^2 + k_y^2})) \tag{23}$$

$$\bar{ATF}_z(k_x, k_y) = ATF_z(k_x, k_y)W(k_x, k_y) \tag{24}$$

where ATF_z is the ATF described in Eq. 4 and $\bar{ATF}_z$ is the ATF modeled with spherical aberration. Finally, we fit the exact size of each aperture mask opening by comparing the empirically measured and simulated PSFs.

Experimental results

Fluorescent bead sample reconstructions In Fig. 5, we show a reconstruction of a 1 μm fluorescent bead at the focal plane and a pair of fluorescent beads 35 μm from the focal plane. With ALF-rings it is slightly clearer that there are actually two beads shown at z = 35 μm. We also include the maximum intensity projections of the entire captured bead volume (approximately 160x160 μm laterally and 100 μm axially) to demonstrate larger FOV reconstruction capabilities of ALF. We encourage the electronic reader to zoom in to examine the reconstruction differences with the different aperture mask schemes of beads near the focal plane and farther from the focal plane.

Pollen grain reconstruction cross-sections In Fig. 6 we include three z-cross-sections of the reconstructed pollen grain volume, for each of LF-pinholes, ALF-circles, and ALF-rings. We also show a full aperture focal stack, with and without deconvolution. The distance between consecutive displayed slices is 9 μm. Since in this case the focal stack is not undersampled, the deconvolved focal stack is expected to have the best performance. Nonetheless, the performance of ALF-circles is closely comparable.

References

[1] M. Bertero, P. Boccacci, G. Desiderà, and G. Vicidomini. Image deblurring with poisson data: from cells to galaxies. *Inverse Problems*, 25(12):123006, 2009.

[2] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy. Wave optics theory and 3-d deconvolution for the light field microscope. *Optics express*, 21(21):25418–25439, 2013.

[3] J. W. Goodman. *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.

[4] J. Mertz. *Introduction to optical microscopy*. Roberts, 2010.

[5] D. Nishimura. *Principles of Magnetic Resonance Imaging*. Stanford University, 2010.

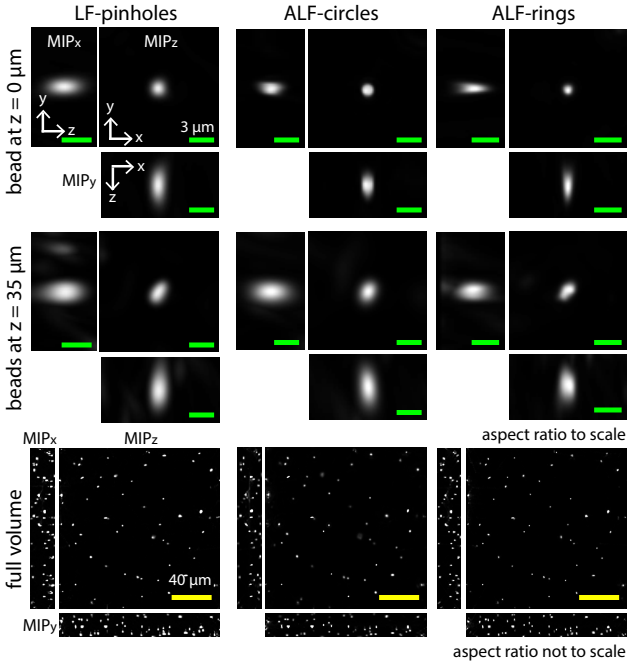

Figure 5. Experimental results from captured data of a specimen composed of 1 µm fluorescent beads in agarose with LF-pinholes, ALF-circles, and ALF-rings. (top) Maximum intensity projections (MIPs) of one or two beads at the focal plane and 35 µm from the focal plane. Green scale bars measure 3 µm. (bottom) MIPs of the full reconstructed volume, measuring 160x160 µm across with 100 µm thickness, limited only by the actual thickness of the sample. Yellow scale bars measure 40 µm. All techniques used a budget of 25 captured images.

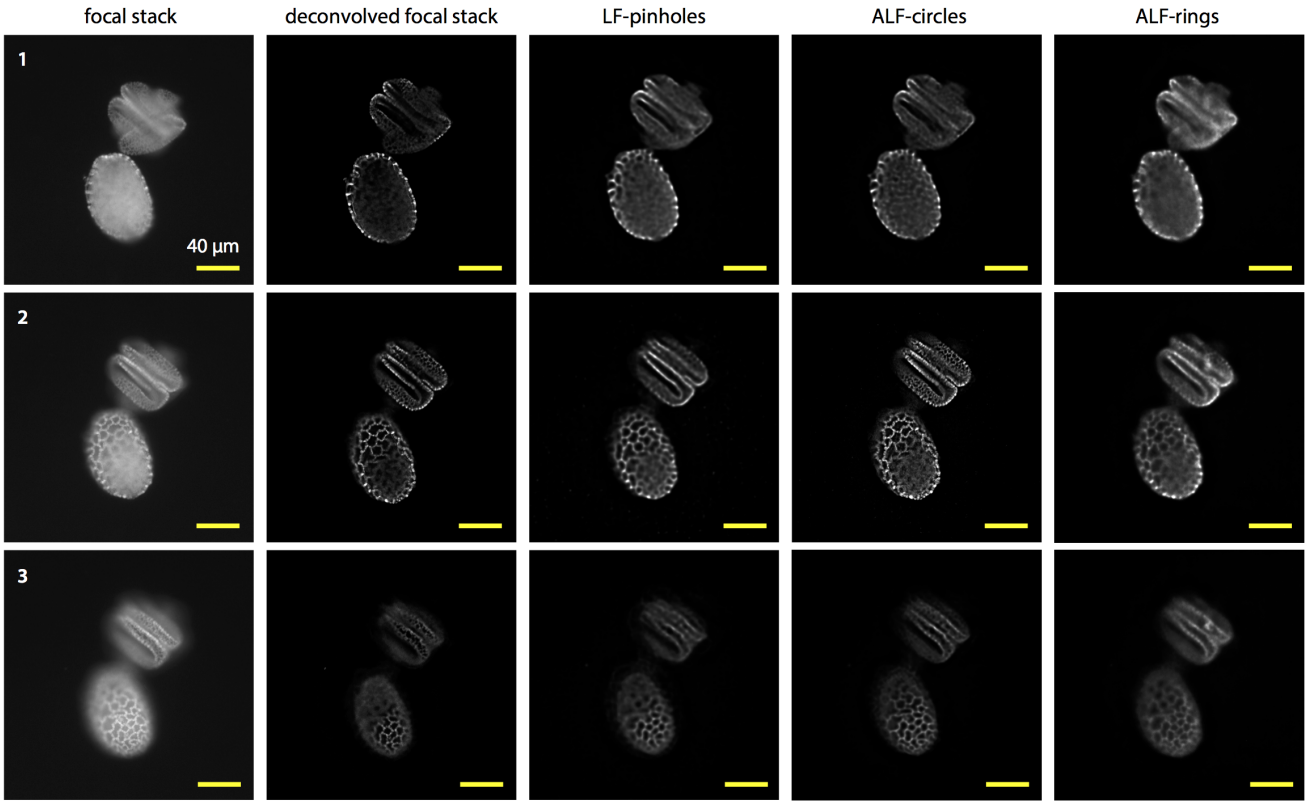

Figure 6. Cross-sections of a reconstructed volume of pollen grains using a focal stack, LF-pinholes, ALF-circles, and ALF-rings. Slices of the focal stack without deconvolution are also shown. The distance between consecutive displayed slices (1, 2, 3) is 9 μm. Yellow scale bars measure 40 µm. All techniques again used a budget of 25 captured images.