

## **Supplementary Material for:**

# **Fluorescence Interference Structured Illumination Microscopy for 3D morphology imaging with high axial resolution**

**Yile Sun,<sup>a,h</sup> Hongfei Zhu,<sup>a,h</sup> Lu Yin,<sup>a,e</sup> Hanmeng Wu,<sup>a</sup> Mingxuan Cai,<sup>a</sup> Weiyun Sun,<sup>f</sup> Yueshu Xu,<sup>a,c</sup> Xinxun Yang,<sup>a</sup> Jiaxiao Han,<sup>a</sup> Wenjie Liu,<sup>a</sup> Yubing Han,<sup>a,g</sup> Xiang Hao,<sup>a</sup> Renjie Zhou,<sup>b,\*</sup> Cuifang Kuang,<sup>a,c,d,\*</sup> Xu Liu<sup>a,c,d</sup>**

<sup>a</sup>State Key Laboratory of Extreme Photonics and Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China

<sup>b</sup>Department of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong, China

<sup>c</sup>ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, 311200, China

<sup>d</sup>Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China

<sup>e</sup>College of Optical and Electronic Technology, China Jiliang University, Hangzhou, Zhejiang 310018, China

<sup>f</sup>Institute of Pharmacology, College of Pharmaceutical Sciences, Zhejiang University of Technology, Hangzhou 310014, China

<sup>g</sup>Britton Chance Center for Biomedical Photonics-MoE Key Laboratory for Biomedical Photonics, Advanced Biomedical Imaging Facility-Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

<sup>h</sup>These authors contributed equally to this work

\*Renjie Zhou, E-mail: [rjzhou@cuhk.edu.hk](mailto:rjzhou@cuhk.edu.hk); Cuifang Kuang, E-mail: [cikuang@zju.edu.cn](mailto:cikuang@zju.edu.cn);

### **Note S1. 4Pi configuration and structured illumination instrumentation**

To maintain the stability of our instrument and minimize external mechanical vibrations, a 900 mm × 580 mm × 80 mm upright marble base plate was used, and all the optical elements were mounted on two 900 mm × 580 mm marble planes of the base plate. Two lasers were used for illumination and excitation: a 1500-mW 560-nm laser (MPB Communications, 2RU-VFL-P-1500-660-B1R) and a 1500-mW 647-nm laser (MPB Communications, 2RU-VFL-P-1500-647-B1R). The wavelength selection and power of the lasers was controlled by an Acousto-Optical Tunable Filter (AOTF) (AA Quanta Tech, AOTFnC-400.650-TN). After the AOTF, the linearly polarized beam was coupled into a single-mode polarization maintaining optical fiber (Schäfter + Kirchhoff GmbH, PMC-E-400RGB-3.5-NA011-3-APC.EC-350-P) using a beam coupler (Schäfter + Kirchhoff GmbH, 60SMS-1-4-RGBV11-47) and then collimated by a beam collimator (Schäfter + Kirchhoff GmbH, 60FC-T-4-M25-01). After the collimator, as shown in Figure S2, the beam was divided into three paths with orthogonal polarization and equal intensity using a combination of Polarization Beam Splitters (PBS) (Thorlabs, CCM1-PBS251/M) and half-wave plates (HWP) (Thorlabs, AHWP10M-600). The central beam of the three beams in 3D SIM was changed to circular polarization by a Quarter Wave Plate (QWP) (Thorlabs,

AQWP10M-580) in order to maintain the pattern modulation when changing the directions. By applying a piezo stage (Physik Instrumente, S-303) and mounting a mirror on it, the central beam could cause a phase shift during the 3D-SIM process. Then, other two beams were directed into two identical modules, which include a galvanometer set (Cambridge Technology, CTI 8310k) to change the orientation of the illumination pattern by altering the incident position of the excitation beam on the back focal plane of the objective lens. In one of the modules, a fixed reflecting mirror was replaced with a mirror mounted on the other piezo stage (Physik Instrumente, S-303) to shift the phase of patterns. Subsequently, these two beams passed through a customized pizza-type HWP (PHWP) (Union Optics), which was composed of six HWPs with different directions of the fast axis and adjusted the polarization direction to be parallel to the tangential direction (Figure S11). The three beams would converge on the back focal plane of the lower objective lens by a beam splitter (BS) (Thorlabs, BS031), 4f configuration and 200-mm tube lens (Thorlabs, TTL200-A) to form an interference pattern in the objective space. After exciting the sample, fluorescence was collected using two 100 $\times$  oil immersion objectives (Olympus, UPLXAPO100XO). After the objective lenses, both upper and lower detection paths 4f configuration and a deformable mirror (Boston Micromachines Corporation, Multi-3.5L-DM) respectively to correct both system-induced aberrations and sample-induced aberrations (see Note S10). In the upper detection path, the beam passed through a QWP (Thorlabs, AQWP10M-580) with a fast axis at 45° to make the s component equal to the p component (see Note S3). In the lower counterpart, the beam passed through the same QWP (Thorlabs, AQWP10M-580) as in the upper path and an additional QWP (Thorlabs, AQWP10M-580) with a fast axis at 0°. The latter was used to generate a phase difference between s and p (see Note S3). Then, the two parts would interfere at a BS (Thorlabs, BS031) and pass through two identical and symmetrical paths, respectively. Each consists of a 4f configuration and square aperture stop (SS) (Daheng Optics, GCM-5711M) to confine the FOV. After a PBS (Thorlabs, CCM1-PBS251/M), the two s components and two p components from the lower and upper paths, respectively, were collected using an sCMOS (Hamamatsu, ORCA-FusionBT C15440-20UP) through Knife-Edge Prism Mirrors (Thorlabs, MRAK25-P01) to form four sub-images that denote four different interference phases (Figure S7).

The total magnification of the imaging process was  $70\times$ , which denoted a pixel size of 92 nm. To make the phase of fluorescence interference uniform in each sub-image, the SS should be adjusted to make the FOV within 30  $\mu\text{m}$ . In the experiments, the actual FOV was 25  $\mu\text{m}$  for each sub-image. The total size of the four images in the sCMOS was  $800\times 800$  pixels, which enabled a maximum frame rate of up to 200 frames per second. For fluorescence interference, the 4Pi configuration can be moved slowly along the z-axis using two vertically placed linear translation stages (Applied Scientific Imaging, LS-50).

## Note S2. Phase retardancy of beam splitter

When using a beam splitter (Figure S14), the Jones Matrix can be written as:

$$\begin{bmatrix} E_3 \\ E_4 \end{bmatrix} = \begin{bmatrix} R & T \\ T' & R' \end{bmatrix} \cdot \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} \quad (\text{S1})$$

Where  $R$ ,  $R'$ ,  $T$ , and  $T'$  are the reflectance and transmittance of the beam splitter on the two surfaces, respectively, which are also denoted as complex values.

For simplicity, we assume a symmetrical beam splitter, that is,  $R' = R$  and  $T' = T$ . Thus,

$$\begin{bmatrix} E_3 \\ E_4 \end{bmatrix} = \begin{bmatrix} R & T \\ T & R \end{bmatrix} \cdot \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} \quad (\text{S2})$$

The energy conservation requires  $|E_3|^2 + |E_4|^2 = |E_1|^2 + |E_2|^2$ . Therefore, the equation can be transformed as:

$$\begin{cases} |R|^2 + |T|^2 = 1 \\ RT^* + TR^* = 0 \end{cases} \quad (\text{S3})$$

Since  $R$  and  $T$  are complex values, we can assume

$$\begin{cases} R = |r| e^{i\varphi_R} \\ T = |t| e^{i\varphi_T} \end{cases} \quad (\text{S4})$$

where the reflectance and transmittance carry the real number  $|r|$  and  $|t|$  with relative phase shift  $\varphi_R$  and  $\varphi_T$ .

Therefore, it can be represented as

$$|R||T|e^{i(\varphi_R - \varphi_T)} + |T||R|e^{i(\varphi_T - \varphi_R)} = 0 \quad (\text{S5})$$

$$2\cos(\varphi_R - \varphi_T) = 0 \quad (\text{S6})$$

Thus,

$$\varphi_R - \varphi_T = \frac{\pi}{2} \quad (\text{S7})$$

## Note S3. Phase difference of sub-images

In generally, we believe that the fluorescence polarization is random. To ensure that the four collected sub-images can reflect the phase information of fluorescence interference, the amplitudes of the s polarization and p polarization in the detection path should be equal before the interference occurs. Therefore, the linear polarization intensity obtained by the sCMOS after fluorescence interference (BS3 in Figure S2) and PBS splitting (PBS3 in Figure S2) should only be affected by the phase and not by its initial polarization state.

Here, we chose to use two quarter-wave plates (QWP4 and QWP2 in Figure S2) in the upper and lower arms of the interference cavity in the detection path, respectively. In particular, the fast axis of both QWPs was set at 45° to the horizontal direction to make the s- and p-component equal in amplitude. It can be proved by the Jones Matrix that when the fast axis of a QWP is at 45° to the horizontal direction, its Jones Matrix can be written as

$$M = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \quad (\text{S8})$$

If the original polarization state of a beam is  $\begin{bmatrix} A_1 \\ B_1 \end{bmatrix}$  and after propagating through the QWP, the polarization state

will be changed to  $\begin{bmatrix} A_2 \\ B_2 \end{bmatrix}$ . The following equation exists:

$$\begin{bmatrix} A_2 \\ B_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} A_1 + iB_1 \\ iA_1 + B_1 \end{bmatrix} \quad (\text{S9})$$

It is clear that  $|A_2| = |B_2|$ , which denotes that after QWP, the amplitude of the s component equals the p component.

In the lower arm of the detection path, we added another QWP (QWP3 in Figure S2) between the QWP in aforementioned. In contrast, the fast axis of QWP3 was set parallel to the horizontal direction. The purpose of this was to combine QWP3 and BS3 into a phase-delay generator, to divide the fluorescence received by the sCMOS into four sub-images with different phases. The phase difference  $\Delta\varphi = \Delta\varphi_p - \Delta\varphi_s$  should be as large

as possible (close to  $\frac{\pi}{2}$ ); thus, the phase of each channel would be  $-\frac{\pi}{2}$ ,  $\Delta\varphi - \frac{\pi}{2}$ ,  $\frac{\pi}{2}$ , and  $\Delta\varphi + \frac{\pi}{2}$ , which

correspond to s1, p1, s2, and p2, respectively. Theoretically, the phase retardancy generated by QWP3 between the p- and s-components is  $\frac{\pi}{2}$  (90°) for the p-component and 0° for the s-component. Therefore, the phases of

the final sub-images will be (including the phase retardancy of  $\frac{\pi}{2}$  from BS, as mentioned in Note S2) 0,  $\frac{\pi}{2}$ ,  $\pi$ ,

and  $\frac{3\pi}{2}$  ( $-\frac{\pi}{2}$ ). However, in practice, the phases are not accurate. According to our measurement and

calibration, the average phases of each sub-image were -17.7°, 90°, 72.8°, and 270° (Figure S7).

#### Note S4. Phase unwrapping

When processing the axial reconstruction, the key is to unwrap the phase from four intensities obtained from four ROI from the same position in the corresponding four sub-images.

Assume that the four intensities are from two channels of  $s$  components  $s_1$  and  $s_2$  and two channels of  $p$  components  $p_1$ ,  $p_2$  respectively:

$$\begin{cases} s_1 = 1 + m \cos(\varphi + \Delta\varphi_s) \\ s_2 = 1 - m \cos(\varphi + \Delta\varphi_s) \\ p_1 = 1 + m \cos(\varphi + \Delta\varphi_p) \\ p_2 = 1 - m \cos(\varphi + \Delta\varphi_p) \end{cases} \quad (\text{S10})$$

where the phase retardancy from the BS is  $\frac{\pi}{2}$  (Note S2), the phase retardancies from the QWP are  $\Delta\varphi_s$  (for the  $s$  component) and  $\Delta\varphi_p$  (for the  $p$  component);  $\varphi$  is the real phase denoting the axial position.  $m$  is the parameter obtained from deriving intensity from amplitude, which equals to  $\frac{|E_1 E_2|}{E_1 E_1^* + E_2 E_2^*}$ , here  $E_1$  and  $E_2$  denote the amplitude of fluorescence from upper arm and lower arm of detection paths.

Then the following two variables can be obtained:

$$rms = \frac{s_1 - s_2}{s_1 + s_2} = \cos(\varphi + \Delta\varphi_s) \quad (\text{S11})$$

$$rmp = \frac{p_1 - p_2}{p_1 + p_2} = \cos(\varphi + \Delta\varphi_p) \quad (\text{S12})$$

Phase  $\varphi$ , which is related to the axial position, can be derived from the above two equations (known  $\Delta\varphi_s$  and  $\Delta\varphi_p$ ). In our method, the  $\Delta\varphi_p$  was approximately  $72.8^\circ$  ( $\Delta\varphi_s$  was 0) after calibration (Figure S7).

## Note S5. Mathematical derivation of axial reconstruction accuracy and simulation verification

For each channel of  $(s_1, s_2, p_1, p_2)$ , we chose the average intensity of an ROI (a region of  $(2M+1) \times (2M+1)$  pixels) as the estimated intensity of the central pixel. In these cases, the fluorophores in adjacent pixels of the central pixel affect the estimated intensity, thereby affecting the accuracy of the axial reconstruction. Furthermore, the random Gaussian noise caused by the confluence of multiple factors (such as shot noise and dark current) influences the precision of axial reconstruction. Here we derive the accuracy under a comprehensive consideration of the influence of adjacent pixels; then, the exact values of accuracy can be obtained numerically.

We denote  $PSF_i$  ( $i=1,2,3,4$ ) as the point spread function for channel  $(s_1, s_2, p_1, p_2)$  and PSF as the normalized accumulation of four channels,  $PSF = \sum_{i=1}^4 PSF_i$  and  $\iiint PSF(x, y, z) dx dy dz = 1$ .

To calculate the axial reconstruction accuracy, we need to define the ground truth axial coordinate  $z_0$  first. We stipulate that  $z_0$  of each pixel can be defined as follows:

Each horizontal pixel corresponds to a unique 3D local biological structure (this local structure can be part of a cylinder, sphere, or other shape). We denote  $f(x, y, z)$  as a distribution function of the fluorophores within the structure.  $f(x, y, z)$  can be further represented as  $f(x, y, z) = \sum_{i=1}^k \delta(x - x_i, y - y_i, z - z_i)$ , which means total  $k$  fluorophores with coordinate  $(x_i, y_i, z_i)$  are uniformly distributed within the local structure. And  $z_0$  is the average value of the axial coordinates of all the fluorophores inside the local structure:

$$z_0 = \frac{\sum_{i=1}^k z_i}{k} \quad (\text{S13})$$

Then, we derive the estimated intensity of the central pixel (neglecting the influence of the fluorophores beyond the ROI range) and use the phase unwrap method to determine the estimated axial coordinate  $z_1$  and regard the difference between  $z_0$  and  $z_1$  as the accuracy for the central pixel. It is worth noting that the accuracy is highly related to the local anatomy of the structure (regarding the fluorophores are uniformly distributed inside the local structure), that is, different structures lead to different accuracies of axial reconstruction.

As the area of the ROI is relatively small and the sectioning thickness is approximately the coherent cycle, we can consider arbitrary  $(x, y, z)$ :

$$\begin{aligned} & PSF_1(x, y, z) : PSF_2(x, y, z) : PSF_3(x, y, z) : PSF_4(x, y, z) \\ & = [1 + m \cos(\varphi + \Delta\varphi_s)] : [1 - m \cos(\varphi + \Delta\varphi_s)] : [1 + m \cos(\varphi + \Delta\varphi_p)] : [1 - m \cos(\varphi + \Delta\varphi_p)] \end{aligned} \quad (\text{S14})$$

Here  $\varphi = 2kz$ , where  $k = \frac{2\pi}{\lambda}$ ;  $\Delta\varphi_s = 0$  and  $\Delta\varphi_p = \frac{\pi}{2}$  theoretically.

Thus,

$$\begin{aligned} & PSF_1(x, y, z) : PSF_2(x, y, z) : PSF_3(x, y, z) : PSF_4(x, y, z) \\ & = [1 + m \cos(2kz)] : [1 - m \cos(2kz)] : \left[1 + m \cos\left(2kz + \frac{\pi}{2}\right)\right] : \left[1 - m \cos\left(2kz + \frac{\pi}{2}\right)\right] \end{aligned} \quad (\text{S15})$$

We can get the follow relation:

$$PSF_1(x, y, z) = \frac{1}{4} [1 + m \cos(2kz)] \times PSF = g_1 \times PSF \quad (\text{S16})$$

$$PSF_2(x, y, z) = \frac{1}{4} [1 - m \cos(2kz)] \times PSF = g_2 \times PSF \quad (\text{S17})$$

$$PSF_3(x, y, z) = \frac{1}{4} \left[1 + m \cos\left(2kz + \frac{\pi}{2}\right)\right] \times PSF = g_3 \times PSF \quad (\text{S18})$$

$$PSF_4(x, y, z) = \frac{1}{4} \left[1 - m \cos\left(2kz + \frac{\pi}{2}\right)\right] \times PSF = g_4 \times PSF \quad (\text{S19})$$

We denote  $(x_j, y_k)$  as the coordinates of each pixel of the ROI using  $j, k = -M, -M+1, \dots, M-1, M$  and  $x_j = j \times \text{pixelsize}$ ,  $y_k = k \times \text{pixelsize}$ .  $f_{jk}$  is the fluorophore distribution function of the corresponding local structure of pixel  $(x_j, y_k)$ .

For pixel  $(x_l, y_n)$  we can also obtain the corresponding ROI range centered on this pixel. Similarly,  $(x_j^{ln}, y_k^{ln}) = (x_l + j \times \text{pixelsize}, y_n + k \times \text{pixelsize})$  with  $j, k = -M, -M+1, \dots, M-1, M$ .  $f_{jk}^{ln}$  is the fluorophore distribution function of the corresponding local structure of pixel  $(x_j^{ln}, y_k^{ln})$ .

We first considered the captured intensity  $I_{00}^i$  of the central pixel  $(x_0, y_0)$  in each channel ( $i$  represents different channels).

$$\begin{aligned} I_{00}^i &= \sum_{jk} \iiint f_{jk}(x', y', z') \times PSF_i(0 - x', 0 - y', 0 - z') dx' dy' dz' \\ &= \sum_{jk} \iiint f_{jk}(x', y', z') g_i(-z') \times PSF_i(x', y', z') dx' dy' dz' \end{aligned} \quad (\text{S20})$$

Similarly, we can get  $I_{ln}^i$

$$\begin{aligned} I_{ln}^i &= \sum_{jk} \iiint f_{jk}^{ln}(x', y', z') \times PSF_i(x_l - x', y_n - y', 0 - z') dx' dy' dz' \\ &= \sum_{jk} \iiint f_{jk}^{ln}(x', y', z') \times g_i(-z') \times PSF_i(x_l - x', y_n - y', 0 - z') dx' dy' dz' \end{aligned} \quad (\text{S21})$$

Thus, the captured intensity of channel  $i$  is as follows,

$$I^i = \sum_{l,n} I_{ln}^i = \sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times g_i(-z') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz' \quad (\text{S22})$$

We have:

$$\begin{cases} s_1 = I^1 \propto 1 + m \cos(2kz_1) \\ s_2 = I^2 \propto 1 - m \cos(2kz_1) \\ p_1 = I^3 \propto 1 + m \cos(2kz_1 + \frac{\pi}{2}) \\ p_2 = I^4 \propto 1 - m \cos(2kz_1 + \frac{\pi}{2}) \end{cases} \quad (\text{S23})$$

$$\begin{aligned} rms &= \frac{s_1 - s_2}{s_1 + s_2} = \cos(2kz_1) \\ &= \frac{m \sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times \cos(-2kz') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz'}{\sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz'} \end{aligned} \quad (\text{S24})$$

$$\begin{aligned} rmp &= \frac{p_1 - p_2}{p_1 + p_2} = -\sin(2kz_1) \\ &= \frac{-m \sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times \sin(2kz') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz'}{\sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz'} \end{aligned} \quad (\text{S25})$$

Thus, the estimated axial coordinate  $z_1$  is in the form of

$$z_1 = \frac{1}{2k} \cos^{-1} \left\{ \frac{m \sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times \cos(-2kz') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz'}{\sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz'} \right\} \\ \times \text{sgn} \left[ \frac{m \sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times \sin(2kz') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz'}{\sum_{l,n} \sum_{j,k} \iiint f_{jk}^{ln}(x', y', z') \times PSF(x' - x_l, y' - y_n, z') dx' dy' dz'} \right] \quad (S26)$$

Accuracy can be regarded as the absolute error between  $z_0$  and  $z_1$ , which equals  $|z_1 - z_0|$ .

Several simulation experiments on axial reconstruction accuracy and precision have been carried out (Figure S5 and S6). It shows that the reconstruction error using wide-field 4Pi detection is less than 10 nm, indicating a <20 nm axial resolution of FI-SIM. Compared with 4Pi single molecule localization microscopy, FI-SIM possesses better noise robustness. FI-SIM calculates the average depth of a local structure filled of thousands of fluorophores. Different from the sparse blinking of fluorescent molecules in 4Pi-SMLM, simultaneous luminescence of a large number of fluorophores provides a higher signal-to-noise ratio and better noise robustness (<2 nm when applying 10% gaussian noise to the raw data).

## Note S6. Image acquisition

The galvanometer, piezo stage, AOTF, sCMOS were all controlled by a 32-channel analog\_out card (National Instruments, NI PXIe-6738). According to generating analog voltage trigger signals, devices can be controlled synchronously. The control software was self-developed based on LabVIEW. The phase shift, period and orientation of pattern, exposure time and ROI selection of sCMOS, etc. can be adjusted conveniently. The image acquisition would be processed once all the parameters were set. Figure S13 illustrated a typical example of waveform for device control and image collection.

Galvo 1x, 1y and Galvo 2x, 2y controlled the positions where two edge beams in 3D SIM focused on the back focal plane of the lower objective; thus, the orientations of the illumination patterns at 0°, 60°, and -60° can be generated respectively.

The piezo stages were moved 5 times to make phase shifts in each orientation of the pattern; that is, phase shifts of 15 times were required in each group of 3D SIM images. It is worth mentioning that both the central beam and one of the edge beams should involve phase shift process, and the amount of phase shift of the latter is twice as much as the former.

The exposure imaging of our camera required a certain time sequence to coordinate with the control scheme of hardware such as galvo-mirrors, piezo stage and AOTF; therefore, we used an external trigger mode to control the camera. Figure S13 shows the time sequence of the external trigger signal. When the external trigger signal arrives, the camera requires a delay (owing to the duration of jitters and opening each line of the detection array

of the sCMOS) to initialize the camera and move the hardware to the corresponding positions. Subsequently, the AOTF controlled the laser to turn on, and the camera began to expose and start imaging.

### Note S7. Data processing and image display

Data pre- and post-processing were performed both in ImageJ and MATLAB. Owing to the nuances of magnification and detection paths of the four sub-images in the process of fluorescence propagation, it was necessary to register the four sub-images, including adjusting rotation, scaling, and displacement between different sub-images. Here, we used affine transformation to obtain a transformation matrix before other data processing<sup>33</sup>. First, we took one of the sub-images as the reference and used the Fourier–Merlin transform (using ‘fmmatch’ function from the MATLAB dipimage toolbox) of the other three sub-images with the reference sub-image to obtain an initial estimation of the transform. Then, the matrix of affine transformation was obtained by the previous result (using ‘find\_affine\_trans’ function from the MATLAB dipimage toolbox). Finally, the four sub-images were transformed by affine transformation (using ‘affine\_trans’ function from the MATLAB dipimage toolbox) using the affine matrix.

For the 3D SIM reconstruction, we used a state-of-the-art algorithm, HiFi-SIM<sup>34</sup>, to recover the super-resolution image. Owing to its optimization in the OTF, there were fewer artifacts in the reconstructed image. The background can lead to poor image quality and become an obstacle in axial reconstruction. Especially in the 4Pi configuration, more out-of-focus information is generated making image segmentation difficult. Here, we used an Alternating Direction Method of Multipliers (ADMM) based denoising algorithm (Note S8) to reduce the background and artifacts of SIM images as much as possible.

To obtain a reliable binary mask from SIM images, we applied a plugin from ImageJ called Trainable Weka Segmentation<sup>48</sup>, which can learn the features of both microtubules and mitochondria and segment them from the background. This approach is much more robust than intensity thresholding, because in SIM, the intensities of the image do not reflect the real fluorescence signals, and it is easy to regard the background and artifacts as signals.

In the axial reconstruction process, the mask filtered the four wide-field sub-images maintained the lateral resolution of SIM and excluded the parts beyond the period of fluorescence interference by optical sectioning of the 3D SIM and 4Pi PSF. Then, axial reconstruction and phase unwrapping were processed in each ROI of the filtered sub-images by calculating the corresponding phase through the intensity distribution of the four sub-images (Figure S3 and Note S4, S5). Depending on the sample type, the ROI can be selected to be slightly larger (e.g., mitochondria) or smaller (e.g., microtubules) than the diffraction limit (Note S5).

A depth-coded image was obtained from the intensity distributions of the four sub-images to show the axial distribution of the sample and illustrate the 3D volume in the 2D image. The color bar from warm to cold reflects the relative axial positions of different ROIs. For the volumetric display, we used ImageJ and commercial software named Vutara SRX Viewer from Bruker. The 3D distribution of the sample was obtained by importing the 3D position data of the reconstructed volume.

## Note S8. Image denoising algorithm based on ADMM

In microscopic imaging, the mathematical model of the observed Poisson image can be expressed as

$$f \sim P(Ax + d) \quad (\text{S27})$$

Where  $x \in \mathbb{R}_+^{n_x n_y}$  is the imaged sample,  $f \in \mathbb{R}_+^{n_x n_y}$  is the observed image degraded by the point spread function (PSF) convolution and Poisson noise,  $d \in \mathbb{R}_+^{n_x n_y}$  is the background noise,  $A$  is the PSF of the optical system, and  $P$  is the Poisson noise process.

The probability of photon counting at a particular image pixel  $i$  can be given as:

$$p(f_i | Ax, d) = \frac{(Ax + d)_i^{f_i} e^{-(Ax+d)}}{f_i!} \quad (\text{S28})$$

To solve the deconvolution problem in Equation S27, we derive a maximum a posteriori (MAP) estimator<sup>49</sup> for the inverse problem that minimizes the negative log-likelihood of Equation S28 and introduces the nonnegativity constraints  $I_{\mathbb{R}_+}(x)$  and other regularization term  $\Gamma(x)$  on the volume as:

$$\min_x -\log p(f_i | Ax, d) + I_{\mathbb{R}_+}(x) + \Gamma(x) \quad (\text{S29})$$

$I_{\mathbb{R}_+}(x)$  represents the regularization term that enforces the nonnegativity constraints:

$$I_{\mathbb{R}_+}(x) = \begin{cases} 0 & \dots x \in \mathbb{R}_+ \\ +\infty & \dots x \notin \mathbb{R}_+ \end{cases} \quad (\text{S30})$$

Where  $\mathbb{R}_+$  stands for the closed convex set representing real-valued nonnegative numbers.

In this study, we utilized the alternating direction method of multipliers<sup>50</sup> (ADMM) to solve Equation S29:

$$\begin{aligned} & \min_x -\underbrace{\log p(f_i | z_1, d)}_{g_1(z_1)} + \underbrace{I_{\mathbb{R}_+}(z_2)}_{g_2(z_2)} + \underbrace{\Gamma(z_3)}_{g_3(z_3)} \\ & \text{subject to } \underbrace{\begin{bmatrix} A \\ I \\ I \end{bmatrix}}_K x - \underbrace{\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}}_z = 0 \end{aligned} \quad (\text{S31})$$

The augmented Lagrangian of the objective in Equation S31 is formulated as:

$$L(x, z, y) = \sum_{i=1}^3 g_i(z_i) + y^T(Kx - z) + \frac{\rho}{2} \|Kx - z\|_2^2 \quad (\text{S32})$$

For convenience, we utilize the scaled form of ADMM by applying the scaled dual variable  $u = \frac{1}{\rho}y$ , leading to:

$$L(x, z, y) = \sum_{i=1}^3 g_i(z_i) + \frac{\rho}{2} \|Kx + u - z\|_2^2 + \text{cons} \quad (\text{S33})$$

The iterative updates for the ADMM with scaled dual variable are given as:

$$z_1^{k+1} := \arg \min_{z_1} g_1(z_1) + \frac{\rho}{2} \|Ax^k + u^k - z_1\|_2^2 \quad (\text{S34})$$

$$z_2^{k+1} := \arg \min_{z_2} g_2(z_2) + \frac{\rho}{2} \|x^k + u_2^k - z_2\|_2^2 \quad (\text{S35})$$

$$z_3^{k+1} := \arg \min_{z_3} g_3(z_3) + \frac{\rho}{2} \|x^k + u_3^k - z_3\|_2^2 \quad (\text{S36})$$

$$u^{k+1} := u^k + Kx^{k+1} - z^{k+1}, \dots, u^k = \begin{bmatrix} u_1^k \\ u_2^k \\ u_3^k \end{bmatrix} \quad (\text{S37})$$

$$x^{k+1} := \arg \min_x \frac{\rho}{2} \|Kx + u^k - z^k\|_2^2 \quad (\text{S38})$$

To solve the proximal subproblem in Equation S34 for the Poisson term, the proximal operator can be expressed as:

$$\arg \min P(z_1) = \arg \min_{z_1} -\log p(f_i | z_1, d) + \frac{\rho}{2} \|Ax + u_1 - z_1\|_2^2 \quad (\text{S39})$$

The closed form solution of the proximal operator is given as:

$$\arg \min P(z_1) = -\frac{1 + \rho d - \rho(Ax + u_1)}{2\rho} + \sqrt{\left(\frac{1 + \rho d - \rho(Ax + u_1)}{2\rho}\right)^2 + \left(\frac{f + \rho d(Ax + u_1) - d}{\rho}\right)} \quad (\text{S40})$$

To solve the proximal subproblem in Equation S35 for the nonnegative term, the proximal operator can be directly expressed as

$$\arg \min P(z_2) = \arg \min_{z_2 \in \mathbb{R}_+} \frac{\rho}{2} \|x + u_2 - z_2\|_2^2 = Q_{\mathbb{R}_+}(x, u_2) \quad (\text{S41})$$

where  $Q_{\mathbb{R}_+}$  represents the projection operator onto the convex set  $\mathbb{R}_+$ :

$$Q_{\mathbb{R}_+}(x, u_2) = \begin{cases} 0 & \dots x < u_2 \\ x + u_2 & \dots x \geq u_2 \end{cases} \quad (\text{S42})$$

To solve the proximal subproblem in Equation S36 for other regularization terms,  $\Gamma(x)$ , the proximal operator is given as

$$\arg \min P(z_3) = \arg \min_{z_3} \Gamma(z_3) + \frac{\rho}{2} \|x + u_3 - z_3\|_2^2 \quad (\text{S43})$$

Various regularization strategies have been proposed to solve the SIM deconvolution problem. Among these, the sparsity prior, total variation (TV), and continuity prior (Hessian matrix regularization)<sup>51</sup> are the most effective priors for denoising the observed noisy images in SIM. For the choice of L1 norm sparsity and TV prior, solving Equation S43 results in a soft-thresholding operator, which is commonly used in L1 norm regularized optimization. It is noteworthy that the choice of the TV prior will additionally result in the constraint  $z_3 = D_1 x$  with  $D_1$ , which means the first-order differential operator. For the choice of the continuity prior, similar to the TV prior, solving Equation S43 with Hessian regularization requires the addition of a differential operator in the constraints  $z_{3,x} = D_{xx} x$ ,  $z_{3,y} = D_{yy} x$  where  $D_{xx}$  and  $D_{yy}$  are second-order differential operators in the  $x$ - and  $y$ -directions, respectively.

### **Note S9. Sample preparation**

25 mm diameter round precision glass coverslips (Thorlabs) were immersed in ultrapure water and vibrated in an ultrasonic cleaner for 15 min. Coverslips were sterilized using 95% ethanol. BSC were purchased from the American Type Culture Collection and cultured in Minimum Essential Media (MEM; Thermo Fisher Scientific, Inc.) supplemented with 10% (v/v) fetal bovine serum (FBS; Thermo Fisher Scientific, Inc.). U2OS were cultured in McCoy's 5A Media (Thermo Fisher Scientific, Inc.) supplemented with 10% FBS. Cultures were maintained at 37 °C in a humidified environment containing 5% CO<sub>2</sub>.

For bead samples, we used red fluorescent FluoSpheres beads (F8801) for fluorescent bead samples, which were maximally excited at 589 nm. Dilute the beads in ultrapure water at a ratio of 1:10, and vibrate in an ultrasonic cleaner for 5 min. Next, 200 µl of the diluted beads were placed on a coverslip and allowed to stand for 10 min. The absorbent tissue was then used to remove excess water and Prolong Glass Antifade (Thermo Fisher Scientific, Inc.) before sealing the coverslip.

For the immunolabeling of microtubules, BSC cells were seeded into the prepared coverslips. After overnight incubation, the cells were washed thrice with phosphate-buffered saline (PBS; Thermo Fisher Scientific, Inc.), fixed with 3% (m/v) paraformaldehyde (Electron Microscopy Sciences) and 0.1% (v/v) glutaraldehyde (Sigma-Aldrich Co., LLC) for 13 min at 37°C, and quenched with NaHB4 for 7 min at room temperature (RT). The cells were incubated with 0.2% (v/v) Triton X-100 (Sigma-Aldrich Co., LLC) and 5% goat serum (Thermo Fisher Scientific, Inc.) for 1 h at RT. Microtubules were stained with mouse alpha and beta tubulin antibodies (WA31679510 and TG2597441, Invitrogen) at 1:1:200 dilution in PBS overnight at 4°C and goat anti-mouse STAR Orange (Abberior GmbH) at 1:200 dilution in PBS for 1 h. Samples were washed three times with PBS before imaging, with two washes for 3 min and the last wash for 15 min.

To label the mitochondria in live U2OS cells, the cells were seeded onto the prepared coverslips. After overnight incubation, the cells were washed thrice with PBS and incubated with 0.25µM Mito-Tracker Deep Red (M22426; Thermo Fisher Scientific, Inc.) in a 5% CO<sub>2</sub> atmosphere at 37 °C for 30 min. The supernatant was discarded, and the cells were washed twice gently with PBS before imaging; For labeling the microtubule in live U2OS cells, the cells were seeded onto the prepared coverslips. After overnight incubation, the cells were washed thrice with PBS and incubated with 1000 times diluted Tubulin-Tracker Deep Red (1000X, C2215S; Beyotime, Inc.) in a 5% CO<sub>2</sub> atmosphere at 37°C for 30 min. The supernatant was discarded, and the cells were washed twice gently with PBS before imaging. To label the migrasomes in live L929 cells, the cells were seeded onto the prepared coverslips. After overnight incubation, the cells were washed thrice with PBS incubated with 10uM DID (C1995S: Bevotime, Inc.) in a 5% CO<sub>2</sub> atmosphere at 37°C for 30 min.

### **Note S10. Aberration correction by deformable mirror**

In super-resolution microscopy, optical aberrations distort the point spread function (PSF) of the imaging systems in three dimensions, thereby reducing the image quality and resolution. The major aberrations are induced by both optical systems and biological samples. System-induced aberrations come from the fact that optical components are not ideal and suffer from manufacturing defects, while sample-induced aberrations arise from the refractive index mismatch between the objective immersion medium and the samples, primarily appearing as spherical aberrations. When applying 4Pi configuration and oil immersion objective lenses to fluorescence microscopy, it is inevitable to consider the influence of aberrations, as significant aberrations will reduce the signal-to-noise ratio and affect the uniformity of fluorescence interference.

To correct for both system-induced aberrations and sample-induced aberrations, image-quality-based sensor-less adaptive optics (AO) was used. This method relies on two deformable mirrors (DM) and wide-field images obtained from oil immersion objective lenses with high numerical aperture (NA) as an input for the algorithm that estimates the optical aberrations which are characterized by Zernike polynomials. The implementation of our method is depicted in the flowchart in Figure S9. For each Zernike mode, a series of frames are acquired with different biases (amplitude of the Zernike mode). Specific image quality metric is selected and computed for each image. Next, a metric curve is fitted to the measured extreme points and the optimal amplitude of each aberration mode is determined by the estimated peak of the curve. This metric curve fitting was implemented by least-squares fitting with a Gaussian function with an offset (4 fit parameters). This procedure is repeated iteratively on all modes to be corrected to get the optimal strength for each mode separately. Then all modes are combined linearly to form the final corrected wave fronts. If needed, the above process can be repeated a set number of times with smaller step size or until the image no longer improves. This metric-based approach is conceptually compatible with any type of microscope and can be used for correcting sample-induced aberrations. For the example presented in our paper we were sequentially correcting preselected Zernike modes to reach the optimal correction. It should be noted that the two deformable mirrors will be controlled separately to correct aberrations from each detection path, rather than using two DMs simultaneously after merging both images from the upper and lower detection paths.

The key for image-quality-based AO is to find a useful image metric, a quality measure computed from the acquisitions. In our method, we use the sharpness metric M reported by Mlodzianoski<sup>52</sup>. This metric is robust to the effect of image noises and is sensitive to the slight change of optical aberrations<sup>53</sup>. The metric has the profile

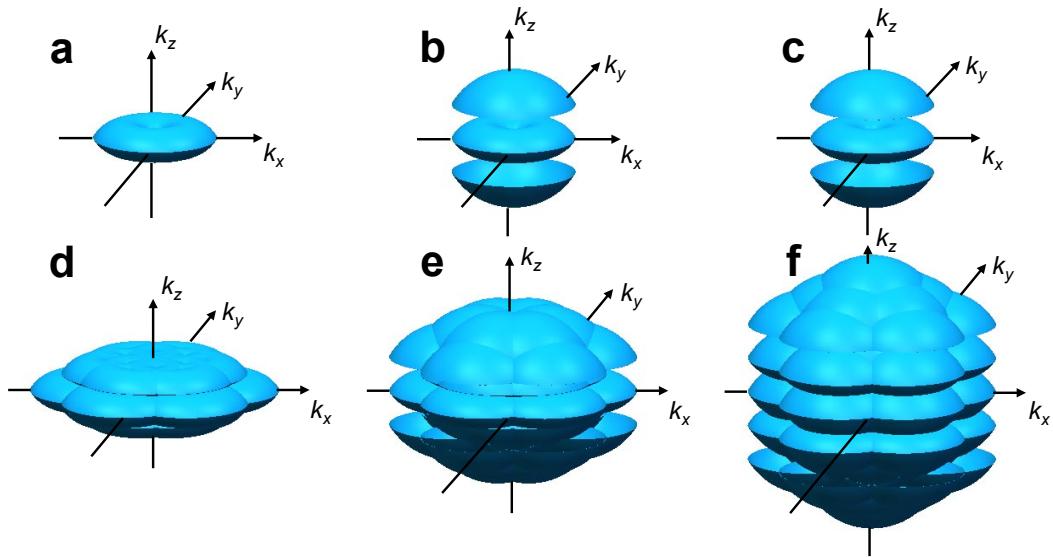
$$M = \frac{\sum_{n,m} \hat{I}_{n,m} (1 - \text{OTF}_{n,m})^\alpha \text{OTF}_{n,m} \mu_{n,m}}{\sum_{n,m} \hat{I}_{n,m} \mu_{n,m}} \quad (\text{S44})$$

Where  $n$  and  $m$  are incremented by 1,  $n, m \in [-\frac{L}{2}, \frac{L}{2} - 1]$  ( $L$  is the number of pixels along a single dimension),

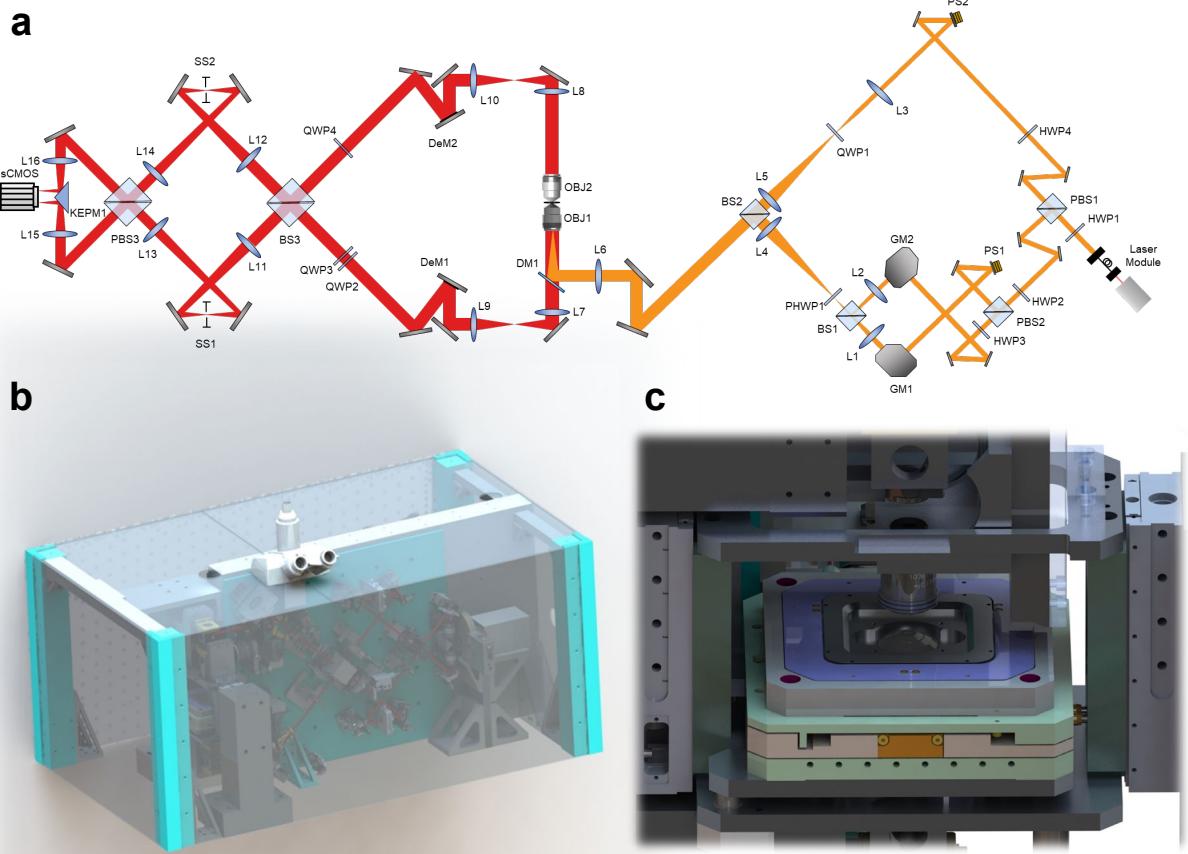
$\hat{I}_{n,m}$  is the discrete Fourier transform magnitude of the image,  $\text{OTF}_{n,m}$  is the discrete Fourier transform of an ideal PSF for our system,  $\alpha$  is an empirically determined value ( $\alpha = 1.3$  here),  $\mu_{n,m}$  is a circular mask,

$$\mu_{n,m} = \begin{cases} 1, & \sqrt{n^2 + m^2} \leq w \\ 0, & \sqrt{n^2 + m^2} > w \end{cases} \quad (\text{S45})$$

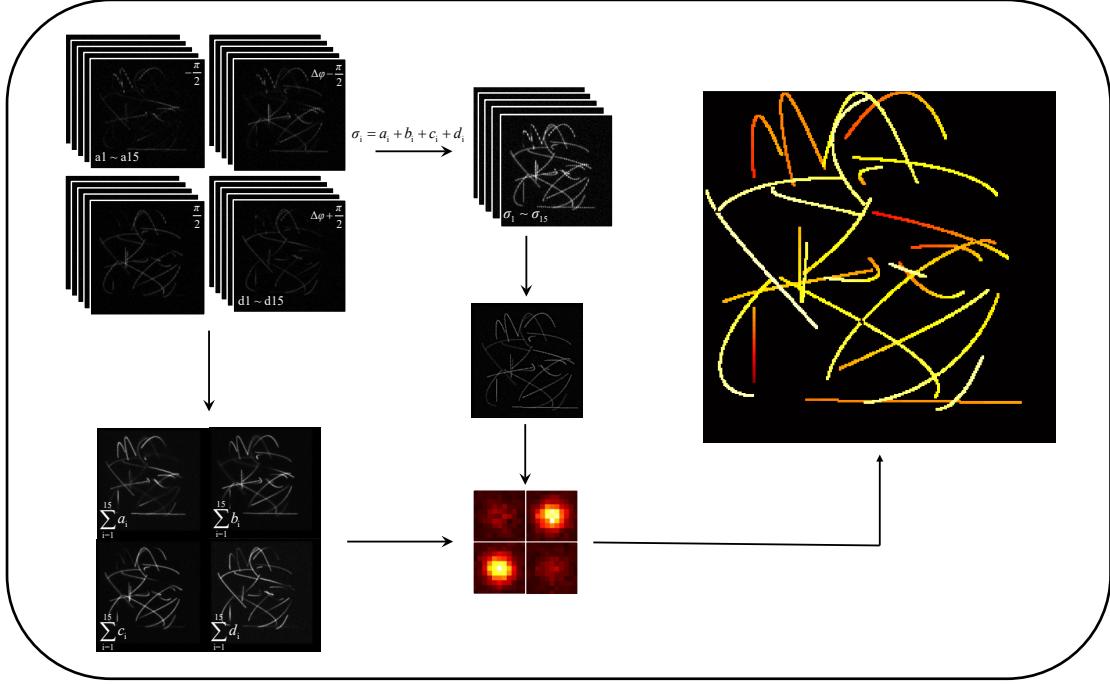
$n' = \frac{n}{L} \Delta x$ ,  $m' = \frac{m}{L} \Delta y$ ,  $\Delta x$  and  $\Delta y$  represent the pixel size of the image in two dimensions (generally,  $\Delta x = \Delta y$ ), and  $w = \frac{\text{NA}}{\lambda}$ , where  $\lambda$  is the emission wavelength and NA is the numerical aperture of the objective lens. The highest weighted region of this metric is located where the deviation between the Fourier transform  $\hat{I}_{n,m}$  and the ideal  $\text{OTF}_{n,m}$  is most pronounced.



**Figure S1.** Optical Transfer Function (OTF) support of SIM in 3D rendering. (a) OTF support domain of a conventional optical microscope. (b), (c) The OTF support of the 4Pi configuration (detected from both objective lenses), compared with the OTF using a single objective lens, is stretched in the  $k_z$  axis. (d) Effective OTF support produced by three-beam illumination in conventional 3D-SIM. (e) Effective OTF support for the proposed method. (f) Effective OTF support by illumination with six beams and detection using two opposing objective lenses.



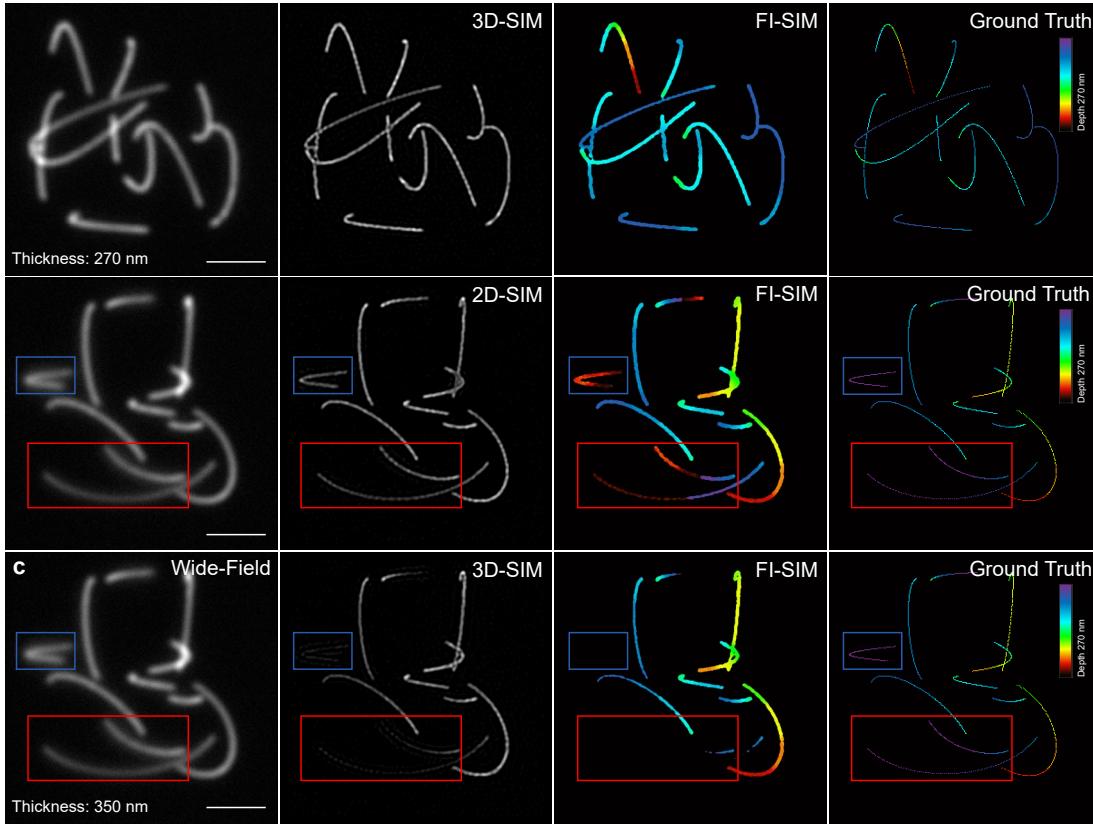
**Figure S2.** Optical setup. (a) Light path of FI-SIM system. (b) Mechanical assembly drawing of FI-SIM. (c) 4Pi structure overview. HWP, half-wave plate; PBS, polarized beam splitter; PS, piezo stage; GM, galvo-mirror; L, lens; PHWP, pizza-typed half wave plate; QWP, quarter-wave plate; BS, beam splitter; OBJ, objective lens; DM, dichroic mirror; DeM, deformable mirror; SS, square aperture stop; KEPM, knife-edge prism mirror.



**Figure S3.** In the image acquisition process, there are four sub-images for each exposure, which are denoted as

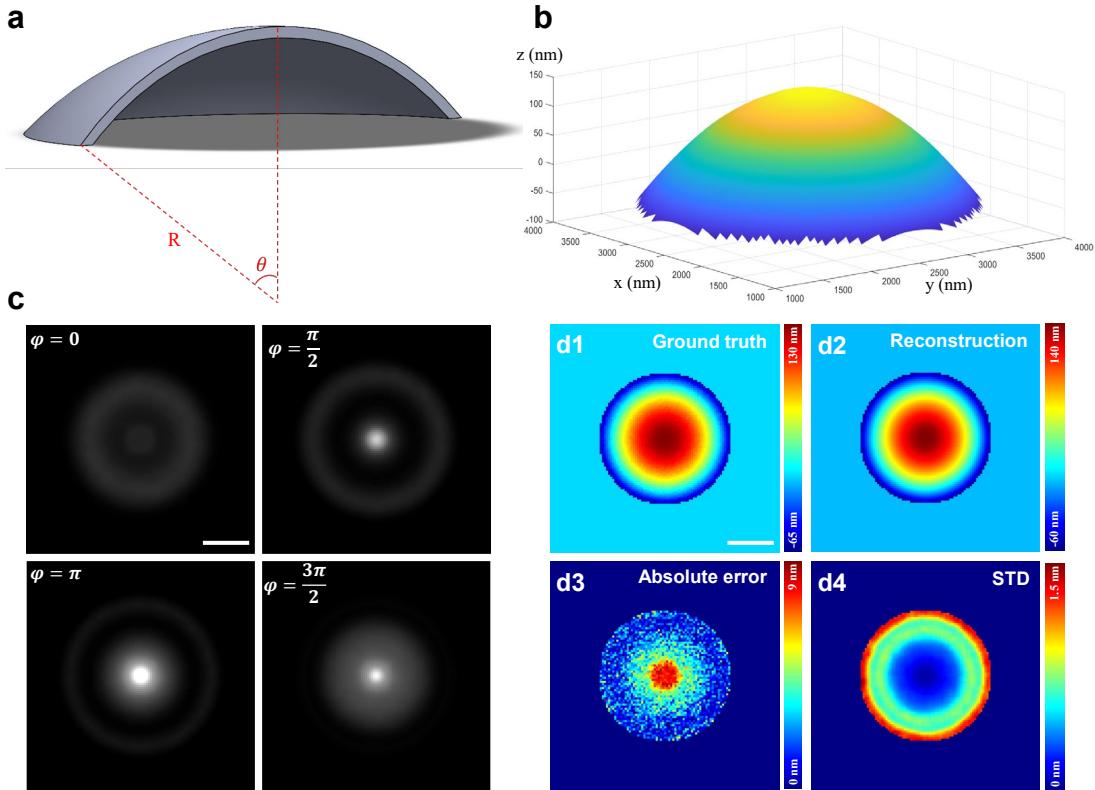
four different phases ( $-\frac{\pi}{2}$ ,  $\Delta\varphi - \frac{\pi}{2}$ ,  $\frac{\pi}{2}$ , and  $\Delta\varphi + \frac{\pi}{2}$ ) of fluorescence interference. To reconstruct the 3D-SIM

image, 15 images (three orientations and five phases for each orientation) were obtained for each reconstruction process. First, four sub-images of each acquisition are summed to form 15 raw images of SIM; this group of images is applied to reconstruct a super-resolution image of 3D-SIM using HiFi-SIM. Then, the Alternating Direction Method of Multipliers (ADMM) based denoising algorithm can denoise the reconstructed image and suppress artifacts, after that, the SIM image is segmented to form a binary spatial mask through Trainable Weka Segmentation. Meanwhile, the original 15 images were summed in each channel to form four wide-field images with different phases. By combining the spatial mask with four wide-field images with different phases, phase unwrapping and axial reconstruction can be conducted.

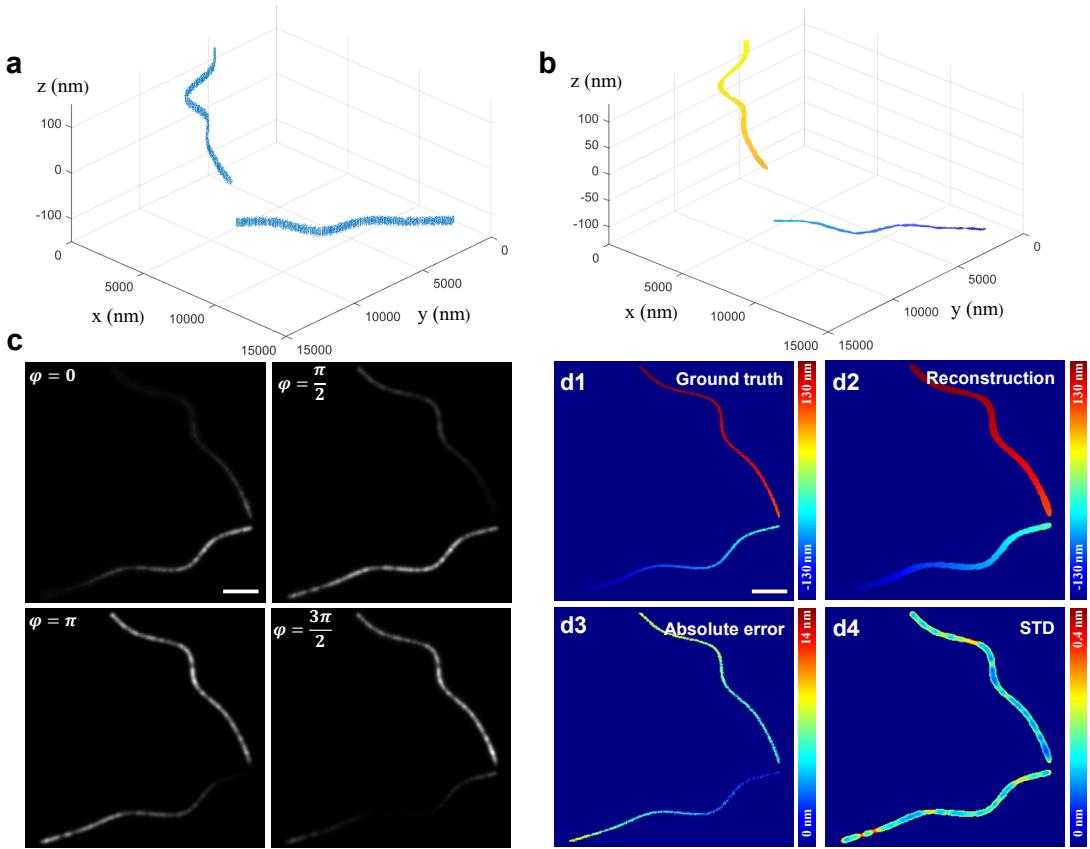


**Figure S4.** Influence of sample thickness and SIM mode. (a) Simulated 270 nm thick sample reconstructed using 3D-SIM and axial reconstruction. Leftmost: wide-field image; middle-left: 3D-SIM image; middle-right: result of axial reconstruction; rightmost: ground truth. When the sample thickness is less than one period of fluorescence interference, that is, less than half of the emission wavelength, the unwrapped phase in each ROI has a unique corresponding axial position, and there is no axial reconstruction error. (b) Simulated 350 nm thick sample reconstructed using 2D-SIM and axial reconstruction. Leftmost: wide-field image; middle-left: 2D-SIM image; middle-right: result of axial reconstruction; rightmost: ground truth. When the sample thickness is larger than one period of fluorescence interference, that is, more than half of the emission wavelength, the corresponding axial positions of some unwrapped phases are not unique, which causes positioning errors (shown in the red and blue outlined regions). (c) Simulated 350 nm thick sample reconstructed using 3D-SIM and axial reconstruction. Leftmost: wide-field image; middle-left: 3D-SIM image; middle-right: result of axial reconstruction; rightmost: ground truth. When the sample thickness is larger than one period of fluorescence interference, that is, more than half of the emission wavelength, the regions beyond the fluorescence interference

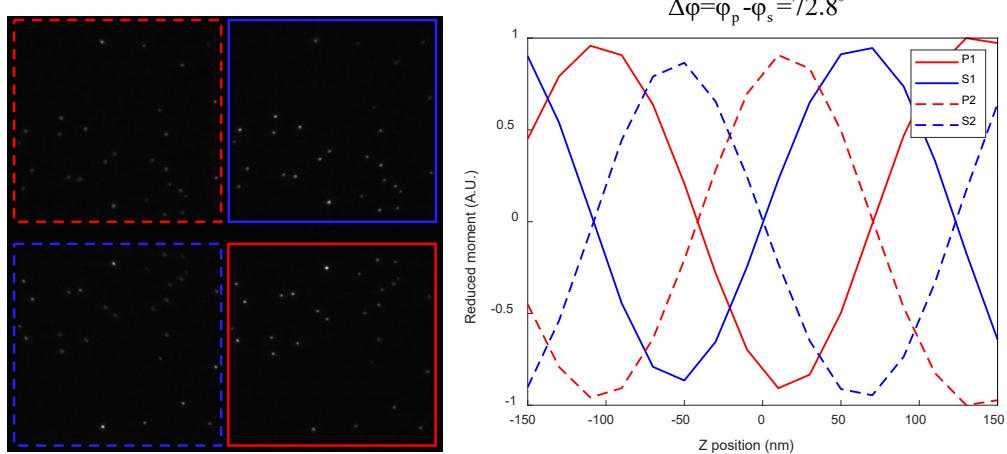
period will be excluded owing to optical sectioning of 3D-SIM and 4Pi-PSF (Figure S1). Thus, positioning error does not exist. Scale bar, 2  $\mu$ m.



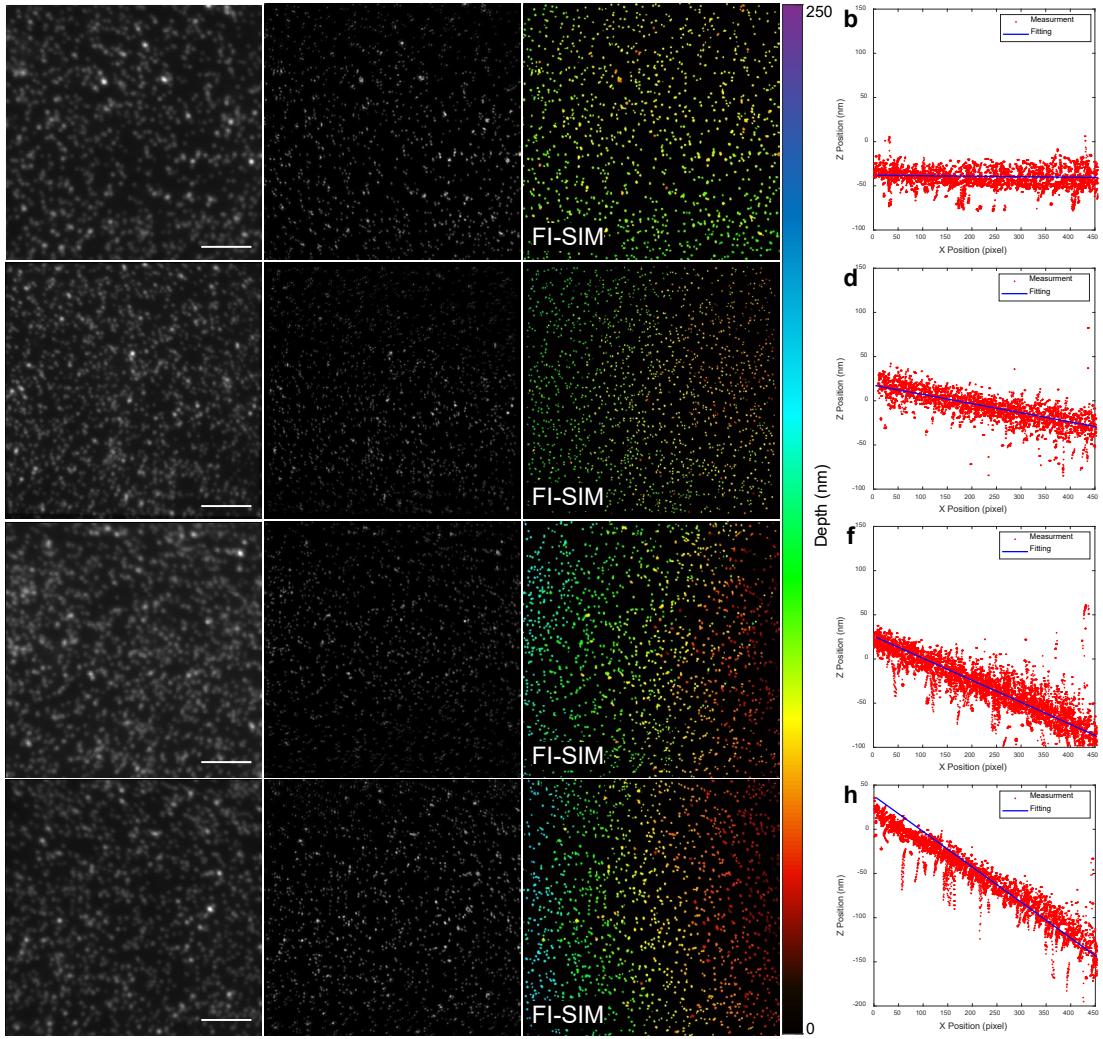
**Figure S5.** Axial reconstruction accuracy and precision of a partial spherical shell structure. (a) 3D spherical shell structure used in simulation. The radius  $R$  of the inner sphere is 5.4  $\mu\text{m}$  and the thickness of the shell is 35 nm. The field angle  $\theta$  is set as 18 degrees. Total 200000 fluorophores are randomly distributed within the shell with 8000 emitted photons of each fluorophore in each 4Pi channel. (b) 3D reconstruction image obtained by phase unwrapping using ROI with size of 300 nm. (c) Wide-field image of 4Pi channels with different interference phases under Poisson noise. (d1), (d2) Depth-coded image of the ground truth and reconstruction results. (d3) Absolute error between reconstruction result and ground truth. (d4) Reconstruction precision with 10% Gaussian noise (500 repeated experiments). Scale bar: 2  $\mu\text{m}$ .



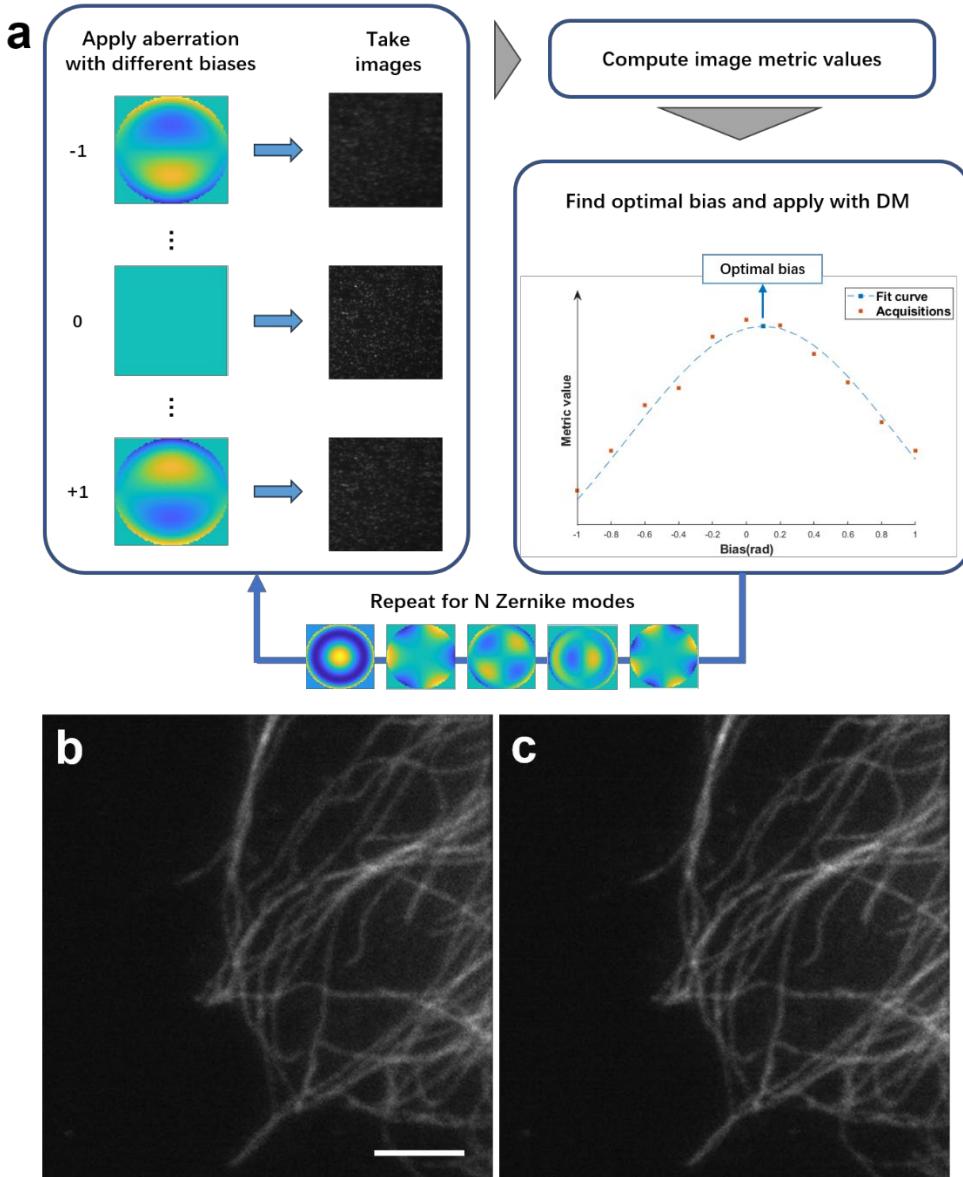
**Figure S6.** Axial reconstruction accuracy and precision of a microtubule structure. (a) 3D microtubule structure used in simulation. The microtubule extends from -130 nm to 130 nm in  $z$  direction. Total 540000 fluorophores are randomly distributed within the microtubule with 8000 emitted photons of each fluorophore in each 4Pi channel. (b) 3D reconstruction image obtained by phase unwrapping using ROI with size of 300 nm. (c) Wide-field image of 4Pi channels with different interference phases under Poisson noise. (d1), (d2) Depth-coded image of the ground truth and reconstruction results. (d3) Absolute error between reconstruction result and ground truth. (d4) Reconstruction precision with 10% Gaussian noise (500 repeated experiments). Scale bar: 2  $\mu$ m.



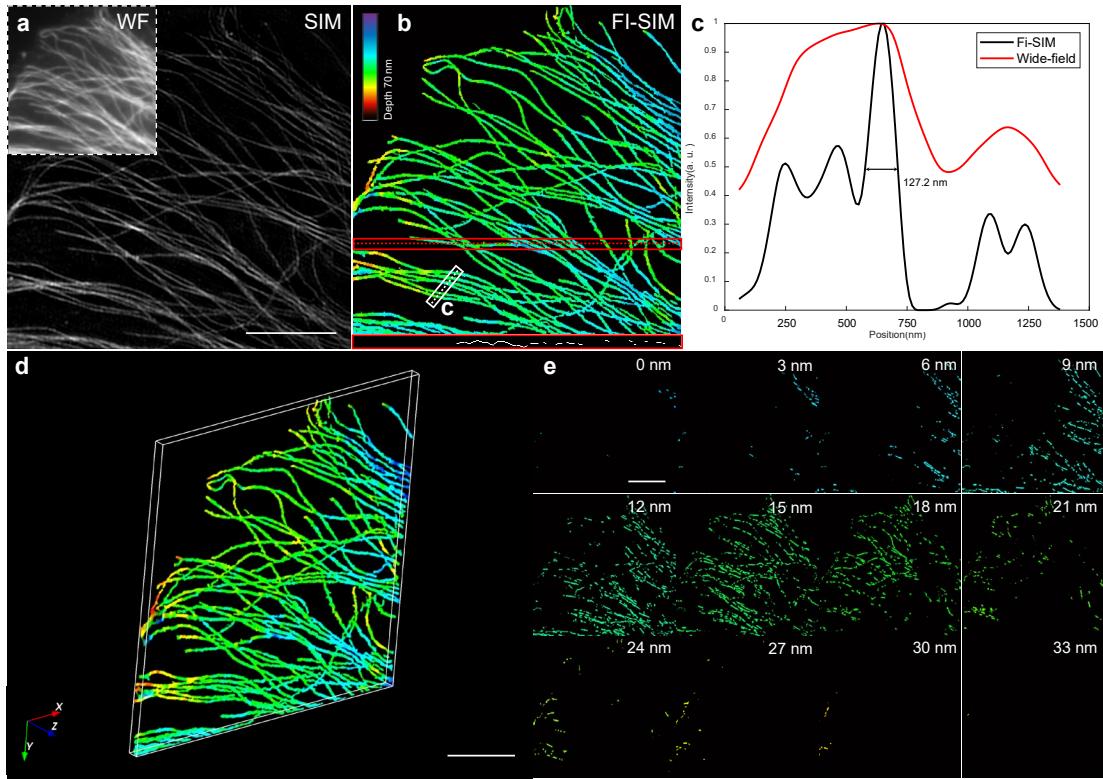
**Figure S7.** Parameters of fluorescence interference calibration. (a) one image in the calibration process and the four outlined regions denote different phases of fluorescence interference:  $-\frac{\pi}{2}$  (s1, blue solid box),  $\Delta\varphi - \frac{\pi}{2}$  (p1, red dotted box),  $\frac{\pi}{2}$  (s2, blue dotted box), and  $\Delta\varphi + \frac{\pi}{2}$  (p2, red solid box), which are also reflected in (b). (b) reduced moment for the four sub-images. Where  $\Delta\varphi = \Delta\varphi_p - \Delta\varphi_s$  ( $\Delta\varphi_p$  is  $72.8^\circ$  and  $\Delta\varphi_s$  is 0 due to the QWP3 in Figure S2). The detailed information can be found from previous article<sup>33</sup>.



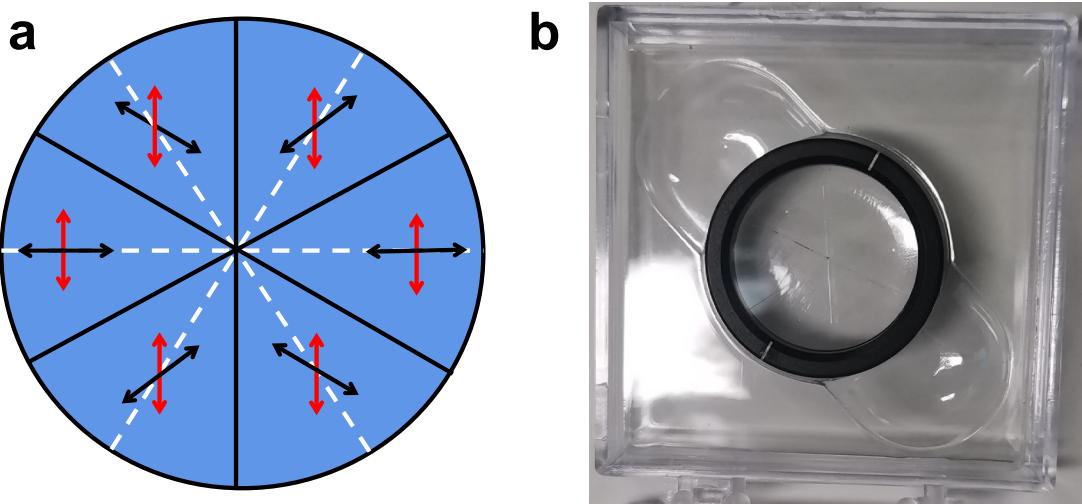
**Figure S8.** More data of calibrations by beads (a,b,e and f are same as Figure 3). (a), (c), (e), (g) WF, SIM and axial reconstruction results when depth difference was 0 nm (sample stage kept level), 50 nm, 100 nm, and 150 nm theoretically from the leftmost to rightmost in FOV, respectively. Left, WF image; middle, SIM image without axial scanning; right, 3D volume super-resolution image. (b), (d), (f), (h) Axial distribution of each pixel as a function of the x positions (red) in (a), (c), (e), (g) respectively; the blue line corresponds to the fitting result, and the fitted axial difference is 2.52 nm, 45 nm, 103.5 nm, 148.5 nm respectively. Scale bar, 5  $\mu\text{m}$ .



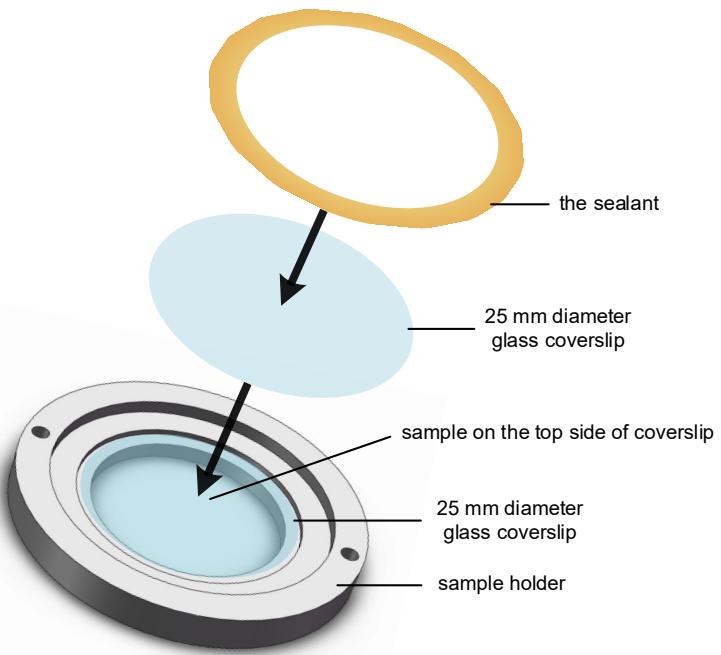
**Figure S9.** Aberration correction by deformable mirrors. (a) Pipeline of aberration correction by each deformable mirror. (b) the wide-field image of microtubules when the images from upper objective lens and lower objective lens merged before aberration correction. (c) the wide-field image of microtubules after aberration correction.



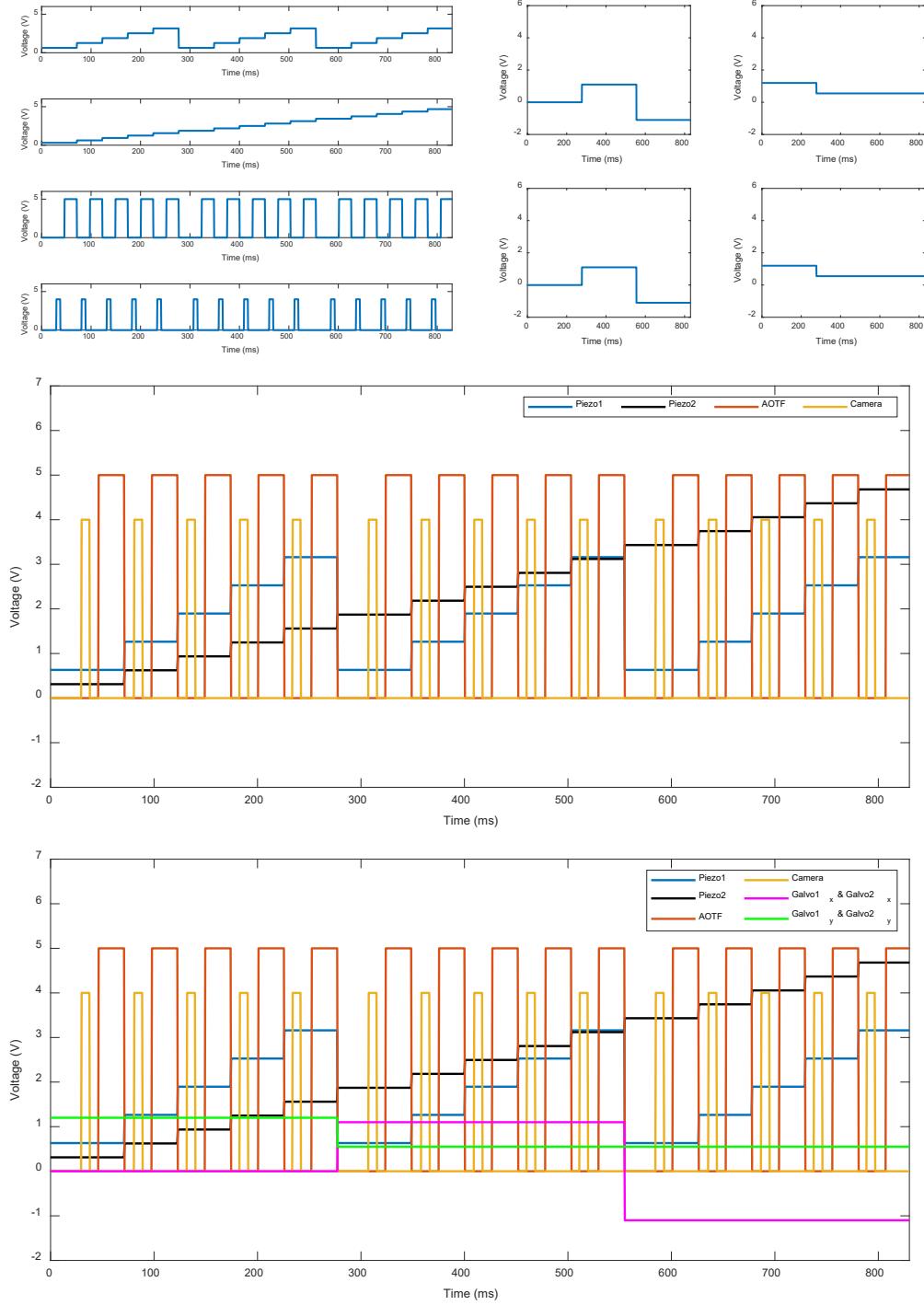
**Figure S10.** More data of fixed microtubules. (a) 3D-SIM image without axial scanning. The top-left of a is corresponding wide-field image. (b) 3D volume super-resolution image of microtubules labeled with STAR Orange in BSC cells; the bottom of (b) shows the vertical cross-section along the red dotted line in the red outlined region. (c) Lateral profile of the white line in b, and the FWHM is 127.2 nm. (d) 3D visualization of the microtubules using Vutara SRX Viewer. e continuous x-y section of the 3D volume from 0 to 33 nm depth with a 3-nm axial interval, illustrating changing morphology of microtubule networks in a narrow axial range. Scale bar, 5  $\mu$ m.



**Figure S11.** Customized pizza-type half-wave plate (PHWP). (a) Principle of customized pizza-type half-wave plates. The red arrows denote the original polarizations of the excitation beams in the excitation path, the black arrows indicate the direction of the fast axis in each fan-shaped wave plate, and the white dotted lines represent the angular bisector of each wave plate. When the two excitation beams pass through the PHWP according to the requirements of SIM (i.e., when the connecting line of the two beams at the convergence points in the back focal plane is  $0^\circ$ ,  $60^\circ$ , and  $120^\circ$ ), the polarization directions of the beams will be changed perpendicular to the connecting line of the two beams. (b) Photograph of the PHWP, with a diameter of 25.4 mm.

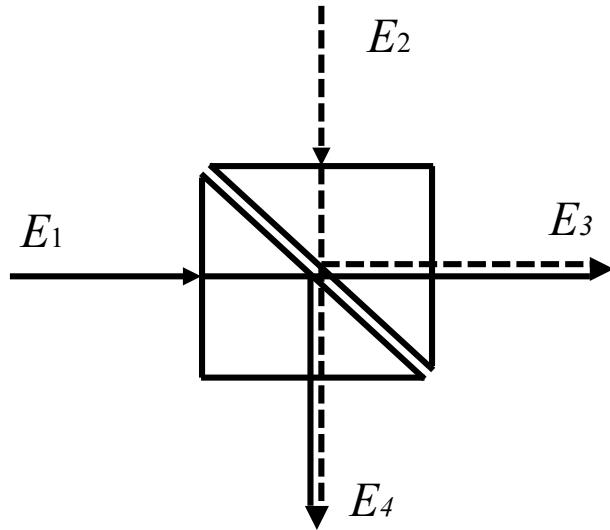


**Figure. S12.** Sample-mounting for 4Pi. A sample coverslip was placed on the sample holder which contained a recess, ensuring that the sample was facing up and that the coverslip was in the recessed area. For the fixed samples, absorbent tissue was used to remove excess water, and Prolong Glass Antifade (Thermo Fisher Scientific, Inc.) was added. For the live-cell sample, 200 uL of phenol red-free DMEM (Thermo Fisher Scientific, Inc.) was placed on the coverslip. The other prepared coverslip was then placed on top of the sample coverslip, taking care to avoid introducing bubbles. Finally, the adhesive was applied to the coverslip and the sample holder to create a seal.



**Figure S13.** Control scheme for synchronous working of devices. (a) Waveforms of control voltages of piezo stages, AOTF, and sCMOS for acquiring a group of images (15 frames, i.e., 60 sub-images). Top: control voltages of piezo stage1 (controls phase shift of edge excitation beam in 3D-SIM); second column: control voltages of piezo stage2 (controls phase shift of middle excitation beam in 3D-SIM); third column: control

voltages of AOTF; bottom: control voltages of sCMOS. (b) Waveforms of control voltages of piezo stages, AOTF, and sCMOS for acquiring a group of images. Upper left: control voltages of galvo-mirror1x. Upper right: control voltages of galvo-mirror1y. Lower left: control voltages of galvo-mirro2x. Lower right: control voltages of galvo-mirror2y. (c) Control scheme of piezo stages, AOTF and sCMOS. (d) Control scheme for all devices that need to be triggered.



**Figure S14.** Sketch map of beam splitter when two orthogonal beams incident and generate phase retardancy