Supplemental Document

OPTICA

Confocal 3D reflectance imaging through multimode fiber without wavefront shaping: supplement

SZU-YU LEE,[1,2] VICENTE J. PAROT,[1,3] BRETT E. BOUMA,[1,2,4] AND MARTIN VILLIGER[1,*]

[1] *Harvard Medical School and Massachusetts General Hospital, Wellman Center for Photomedicine, Boston, Massachusetts 02114, USA*
[2] *Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3] *Institute for Biological and Medical Engineering, Pontificia Universidad Católica de Chile, Santiago 7820244, Chile*
[4] *Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[*] *Corresponding author: mvilliger@mgh.harvard.edu*

Supplement DOI: https://doi.org/10.6084/m9.figshare.17152670

Parent Article DOI: https://doi.org/10.1364/OPTICA.446178

Confocal 3D reflectance through multimode fiber without wavefront shaping

1. DETAILED DESCRIPTION OF EXPERIMENTAL SETUP

All experiments used a 1-m-long step-index MMF with 105 µm core diameter and a NA of 0.22 (FG105LCA, Thorlabs) that theoretically supports $\sim$ 550 guided modes per polarization. The fiber was coiled with a minimum radius of curvature of $\sim$ 50 mm. The monochromatic calibration matrix **T** was measured by sequentially probing the MMF input channels, while holographic detection was used for all output channels concurrently, as depicted in Fig. S1(a). Each input and output channel included two orthogonal polarization states: horizontal (H) and vertical (V). To alternate the illumination polarization between H and V, a laser beam (λ = 1550 nm and linewidth < 100 kHz) was linearly polarized and passed through a fiber-based electro-optical phase retarder (PR, Boston Applied Technologies). The laser was steered by a two-axis galvanometer scanning stage (GM, GVSM002-US, Thorlabs), and then focused by an objective lens (Plan Apo NIR Infinity Corrected, Mitutoyo) with a NA of 0.4 into a 2.5 µm full-width at half maximum (FWHM) spot on the proximal facet of the MMF. The angular spectrum of the spot exceeded the NA of the MMF to ensure efficient population of all modes. The focal spot position on the proximal input side was indexed by u and the speckle pattern exiting on the distal side was imaged with another identical objective lens and a tube lens ($f = 30cm$) onto an InGaAs camera (OW1.7-VS-CL-LP-640, Raptor Photonics) with exposure time of 20 µs at 120 frames per second. The distal channel in real-space was indexed by ν. The object plane of the distal imaging system determined the calibration plane, which was approximately 100 µm away from the distal facet. We define d as the distance of the OP away from the MMF distal facet (at $d = 0$). A beam displacer (BD40, Thorlabs) was used in front of the camera to spatially separate the output into H and V polarization states. An angled plane reference wave polarized at 45° independently interfered with the two speckle patterns on the camera to record the speckle field amplitude and phase through off-axis holography in both detection polarization states simultaneously. Images of the two polarization states were demodulated, spatially registered, and flattened into a column vector of **T** directly in the Fourier domain, with output channels at (k_x, k_y) indexed by ν_F. To uniformly probe all the MMF's guided modes, transmission was recorded for an oversampled grid of input spot positions u within the core region, typically $\sim$700 points for each input polarization state, sequentially generated by driving the GM and PR. The total acquisition time was 20 seconds. The input and output spatial channels of **T** have been ordered first by spatial coordinate, then by polarization. Applying singular value decomposition on the measured **T**, we determined the number of MMF supported modes to be 550 per polarization state, consistent with the theoretical model.

In imaging experiments, as illustrated in Fig. S1(b), a sample was placed in front of the MMF distal tip (b-1) and the round-trip **M** was measured from the proximal side (b-2). We again sequentially coupled light into the MMF through the same set of proximal input states. Light with a power of 0.5 mW exited the distal facet and propagated towards the sample, where part of the light backscattered and coupled back into the same MMF. On the proximal side, we recorded the round-trip light transmission by decoupling its path from the illumination with a non-polarizing beam splitter and directing it to the same off-axis holography setup. The exposure time was set in the range 200-1000 µs depending on the sample. A complete round-trip sample measurement was acquired in 20 s. To preserve the symmetry between the illumination and the detection configurations and to obtain a square matrix **M**, we sampled the recorded complex output fields at the ordered positions identical to the set of input states. The matrix **M** was then constructed with the same procedure that was used to find **T**. Numerical corrections to compensate for the physical misalignment were then applied to the output channels of the measured **M** to accurately match the input channels and recover the underlying transpose symmetry as previously described [1].

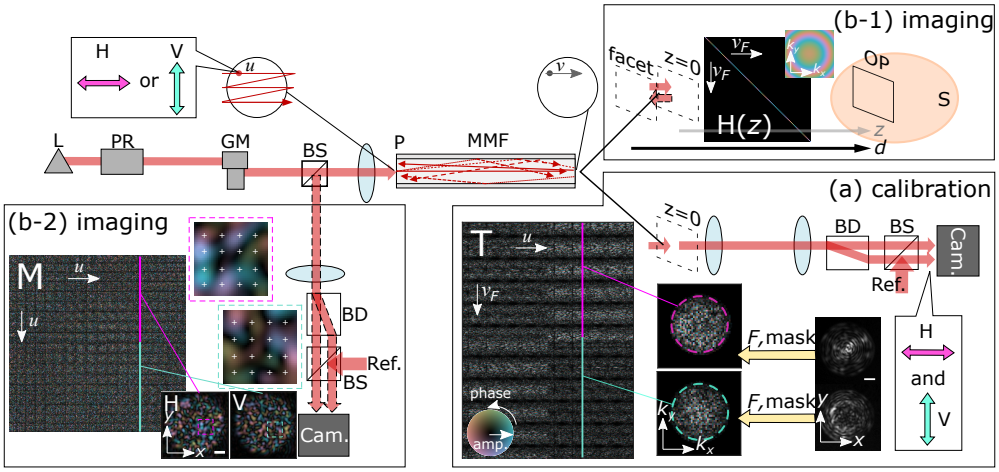

Fig. S1. Measurements of the MMF TMs. The fiber, although drawn as if it were straight, was in fact coiled in experiments. The red arrows correspond to light pathways, and the dashed ones indicate reflected light traveling from the sample through the MMF in the reverse direction to the proximal detection. BD: beam displacer, BS: non-polarization beam splitter, Ref.:reference wave, Cam.: camera, S: sample. A focused spot was scanned with the GM across positions indexed by u distributed over the MMF proximal input facet and alternating between H and V polarizations by means of the PR. The output field was split into two orthogonal polarization states by the BD, and interfered with matching reference waves for simultaneous recording. (a) In the calibration phase, the camera records the transmitted speckle pattern interfering with the reference wave in spatial coordinates (x, y) (rightmost insets). In the Fourier domain, we isolated the demodulated complex-valued signals in momentum coordinates (k_x, k_y) confined to a frequency band imposed by the fiber NA and rearranged them into a column vector of $\mathbf{T}$, as indicated by the solid vertical line, color-coded in magenta and cyan for the H and V polarizations, respectively. The forward transmission $\mathbf{T}$ has rows and columns indexed by ν_F and u, respectively, and was ordered first by the spatial modes, and then by polarization states. Only a subset of $\mathbf{T}$ is shown here. The color map encodes complex values. (b-1) In the imaging phase, light backscattered from distal OPs at varying d from the fiber facet. Free-space propagation, modeled by $\mathbf{H}$, is a diagonal matrix in ν_F where it defines a quadratic phase in the Fourier domain. (b-2) The detected images were demodulated into complex-valued images of the proximal output speckle in spatial coordinates (x, y), then down-sampled at the positions of the input foci (shown as white markers in the dashed magenta and cyan boxes) following the same ordering as for illumination, and flattened into column vectors of the square matrix $\mathbf{M}$. $\mathbf{M}$ thus has rows and columns indexed both by u. Only a subset of $\mathbf{M}$ is shown here. The scale bars in the insets are 20 µm.

2. RESOLVING AXIAL INFORMATION WITH NUMERICAL REFOCUSING

For a sample with volumetric structures, under weakly scattering regime and the Born approximation, we can express the total light reflection counting from the calibration plane ($z = 0$) as a summation of backscattering fields contributed from N individual OPs at varying axial positions,

$$\sum_{i=1}^{N} \mathbf{H}^{\mathrm{T}}(z_i)\mathbf{R}(z_i)\mathbf{H}(z_i). \tag{S1}$$

$\mathbf{H}$ is a unitary TM modeling the loss-less free-space propagation from the calibration plane to the OP. Due to the unitary matrix properties,

$$\mathbf{H}^{-1} = \mathbf{H}^{\dagger} \quad \text{and} \quad \mathbf{H}^{-\mathrm{T}} = \mathbf{H}^{\star}, \tag{S2}$$

where the superscript –1, –T, †, and $\star$ indicate true inverse, true inverse of transpose, Hermitian transpose, and conjugate, respectively. Note that $\mathbf{H}$ simply reduces to an identity matrix when $z = 0$. According to Fresnel diffraction theory under paraxial approximation, the transfer function

of a free-space propagation is a convolution kernel in real space, or a multiplicative quadratic phase term in the Fourier domain. Depending on the distance, z, of a selected OP, we can compute the Fourier phase term accounting for the propagation process

$$F(k_x, k_y, z) = \exp(\frac{-iz(k_x^2 + k_y^2)}{k_n}), \tag{S3}$$

where k_n is the wavenumber in the given medium and k_x and k_y are the coordinates in the in-plane momentum domain. The matrix $\mathbf{H}(z)$ in Fourier domain is then a diagonal matrix and incorporating it into $\mathbf{T}$ through left-multiplication extends the output of $\mathbf{T}$ to the OP at z. Note that $z/k_n = zn/k_0$, where k_0 is the wavenumber in $n = 1$, encodes the optical path length, which is the physical thickness of the medium multiplied with its refractive index n, and $\mathbf{H}$ is parameterized only by z and independent of fiber shape and $\mathbf{T}$. Plugging Eq. S1 into Eq. 1 and then into Eq. 2, and substituting respectively $\mathbf{T}^{-1(\text{tik})}$ with $\mathbf{T}^{-1(\text{tik})}\mathbf{H}^{\dagger}(z)$ and $\mathbf{T}^{-\text{T}(\text{tik})}$ with $\mathbf{H}^{\star}(z)\mathbf{T}^{-\text{T}(\text{tik})}$ using Eq. S2, we have a new transpose-symmetric reflection matrix $\mathbf{R}$ with input and output channels shifted to z

$$\tilde{\mathbf{R}}(z) = \mathbf{H}^{-\text{T}}(z)\mathbf{T}^{-\text{T}(\text{tik})}\mathbf{M}\mathbf{T}^{-1(\text{tik})}\mathbf{H}^{-1}(z) \approx \sum_{i=1}^{N} \mathbf{H}^{-\text{T}}(z)\mathbf{H}^{\text{T}}(z_i)\mathbf{R}(z_i)\mathbf{H}(z_i)\mathbf{H}^{-1}(z), \tag{S4}$$

which is the same as Eq. 2. If we set $z = z_j$, Eq. S4 becomes

$$\tilde{\mathbf{R}}(z_j) \approx \mathbf{R}(z_j) + \sum_{i \neq j}^{N} \mathbf{H}^{-\text{T}}(z_j)\mathbf{H}^{\text{T}}(z_i)\mathbf{R}(z_i)\mathbf{H}(z_i)\mathbf{H}^{-1}(z_j), \tag{S5}$$

where we isolate the in-focus from the out-of-focus matrices. Assuming the out-of-focus reflective planes are separated from the in-focus plane by much more than a depth of focus, and the total background energy is uniformly distributed over all spatial channels, we can approximate the summation of out-of-focus terms in Eq. S5 as a complex matrix with random phases but a constant amplitude. Collecting the on-diagonal elements of $\tilde{\mathbf{R}}(z_j)$ hence leads to signal predominance by the *en face* reflectivity at $z = z_j$ and suppression of out-of-focus signals, or background rejection. In short, after measuring $\mathbf{M}$, by obtaining $\tilde{\mathbf{R}}$ at intended axial position following Eq. S4 and then applying Eq.3 , we can digitally shift to the j$^{\text{th}}$ OP and image the *en face* reflectivity at $z = z_j$ without repeated measurements.

3. DIGITAL RESAMPLING OF IMAGE PHYSICAL AND DIGITAL DIMENSIONS

The light transport through a MMF and interaction with a distal sample can be well modeled with measured TMs, which contain full complex propagation information of wave-vectors within the NA of the MMF. While the experimental $\mathbf{T}$ has output channels stored in Fourier domain, with an one-time measured $\mathbf{M}$ in the imaging phase, arbitrary resampling of 2D image dimensions and also digital adjustment of image size on any selected OPs can be readily configured based on Fourier relations. This offers flexible trade-off between image processing speed and accuracy in a pragmatic circumstance: a lower resampling rate or smaller physical dimension reduces the computational burden, which is suitable for a faster image preview, whereas a higher resampling rate produces a detailed and smooth image at the expense of longer processing duration. Here, we quantify the trade-off by timing the image processing on a personal computer with a 3.4 GHz Intel Core i7 CPU and 16 GB RAM using MATLAB.

For an arbitrary setting of image physical and digital dimensions, we upsampled the output spatial channels of $\mathbf{T}$ in the Fourier domain by interpolation, and pre-computed an inverse discrete Fourier Transform (iDFT) matrix for converting the distal channels to resampled real-space coordinates during the 2D real-space image reconstruction. We focus on the upsampling that corresponds to a valid augmentation to the initial pupil size on the calibration plane ($\sim$105 µm in diameter). Note that the interpolation of $\mathbf{T}$ output channels and the calculation of iDFT matrices are performed prior to actual image formation processes. The necessary computation of images on an OP involves application of phase terms to $\mathbf{T}$ outputs for intended numerical refocusing, distal spatial channels conversion into real-space coordinates with the prepared iDFT matrix, left and right multiplication of $\mathbf{M}$ with regularized inversion of extended backward and forward TMs following Eq. S4 to retrieve $\tilde{\mathbf{R}}$, and reshaping back to a 2D image using Eq.3.

In the experiment, the initial $\mathbf{T}$ had output channels accounting for 247 $\times$ 247 square area of camera recording pixels conjugating a physical size of 123 $\times$ 123 µm^2, and the resolution chart

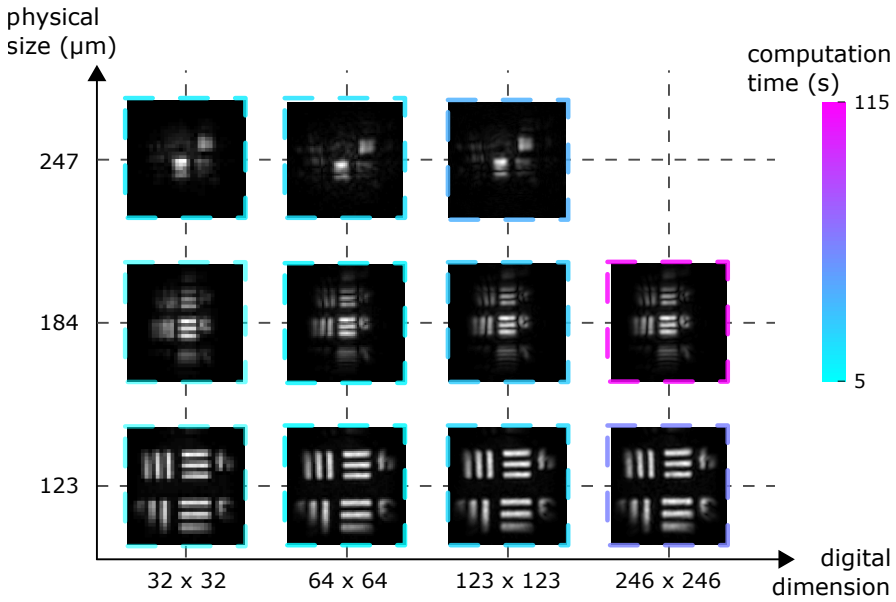

Fig. S2. Computation time and quality of confocal images as a function of physical and digital image size. We only show intensity images in a single polarization state since the sample is binary and isotropic. The time in second is color coded.

as sample was placed on an OP at $d = 10$, 600, 1200 μm away from the facet. For each imaging setting, we timed only the necessary computation. As shown in Fig. S2, the computation for co-polarization 2D confocal images at $d = 10$ μm with original dimensions and size takes ∼58.2 sec. To reduce computation complexity and complete image formation in a shorter time, we can down-sample the image dimensions to 32×32 in the same physical extent, resulting in pixelated images on OPs at $d = 10$ μm calculated within ∼5.2 sec. For images on an OP at $d = 1200$ μm from the distal MMF facet, illuminating light diverges, and a larger configured image physical dimension is needed to avoid image clipping. For instance, the computation time of 32×32 confocal intensity images covering 247×247 μm^2 on the OP at $d = 1200$ μm is $\sim$ 11.2 sec. Table S1 summarizes the computation time of individual settings.

computation time (sec)				
physical size (μm)\digital dimension	32	64	123	247
123	5.2	7.9	17.5	58.2
184	7.4	13.7	28.5	114.81
247	11.2	20.3	39.9	

Table S1. Computation time of confocal images considering different image configuration settings.

4. 3D RESOLUTION AND FIELD OF VIEW

To calculate the theoretical lateral and axial resolution, we need to first compute the effective NA, NA_{eff}, specific to an OP at an axial position. While the effective NA may also be dependent of the lateral displacement from the optical axis, we only consider an on-axis point object on the OP for

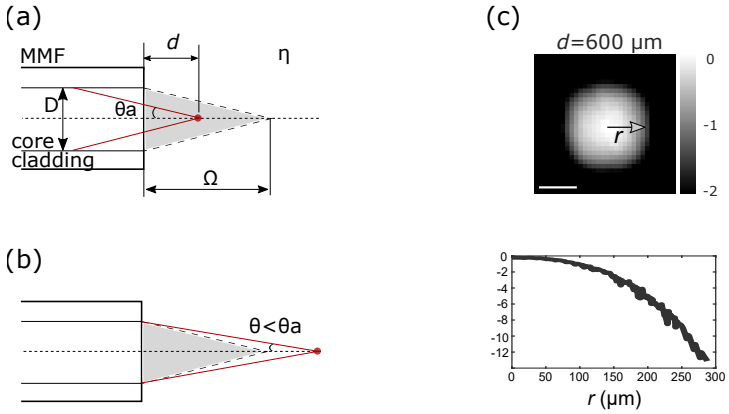

Fig. S3. Illustration of effective NA for an on-axis point object (red dot) on an OP (a) within and (b) beyond the MMF focal length. The dotted line indicates optical axis. (c) The simulated beam divergence at $d = 600$ μm with experimental $\mathbf{T}$, and the circular blob diameter associates with the imaging FOV. The plotted radius-wise mean intensity in logarithmic scale with defined threshold determines the expected FOV on the OP at d. The scale bar is 100 μm.

convenience. Given the object at a distance d away from the MMF facet, the effective NA can be calculated from the maximal angle formed with the point as the vertex and marginal rays within the MMF acceptance angle as sides, as illustrated in Fig. S3(a) and (b). When d is within the focal length of the MMF, $\Omega \sim \eta D/2NA$, a full NA can be obtained, which is determined during MMF fabrication. Here, η is the medium refractive index, and θ_a is the fiber acceptance angle. Once d is larger than this range, only a partial NA can be achieved due to the limited MMF diameter. The value of effective NA is summarized as

$$NA_{eff} = \begin{cases} NA, & \text{if } d < \frac{\eta D}{2NA}. \\ \sim \frac{D}{2d}, & \text{otherwise.} \end{cases} \tag{S6}$$

Given the effective NA, we can then compute the expected lateral and axial resolution as in confocal microscopy [2]

$$\delta x = \frac{0.4\,\lambda}{NA_{eff}} \tag{S7a}$$

$$\delta z = \frac{1.4\,\eta\,\lambda}{NA_{eff}^2}, \tag{S7b}$$

where we see that the axial resolution has a strong dependence on the system NA.

With the circular symmetry of the fiber core shape, we can define the FOV on an OP as the diameter of a circular area with circumference from furthest off-axis points having normalized confocally detected intensity dropped below 1% threshold. Using the measured $\mathbf{T}$ of the MMF, we can free-space propagate each output light field per input to an OP and incoherently sum all output light intensity over each input realization. This results in a circular blob on the OP indicating the average illumination power at each spatial channel. Taking the spatial-channel-wise intensity square of the blob informs confocally detectable power, as shown in Fig. S3(c), where the OP is 600 μm away from the MMF distal facet. The low light coupling efficiency at FOV peripheral causes the vignetting effect on reconstructed images, and the quantified FOV has ∅∼260 μm by applying the threshold to plotted radius-wise mean intensity.

5. RECONSTRUCTING CONFOCAL IMAGES FROM PARTIAL TM MEASUREMENT

In the imaging phase, we can reconstruct confocal images from the round-trip measurement of $\mathbf{M}$ by obtaining the reflection matrix, $\tilde{\mathbf{R}}$, processing its elements, and reshaping into 2D real-space coordinates at a selected OP. While the measurement of a full $\mathbf{M}$ by sequentially coupling light

into all MMF proximal channels delivers maximal information of the distal sample bounded by the MMF throughput, intermediate confocal images for preview can also be reconstructed from a round-trip measurement with partial set of input realizations, $\ddot{\mathbf{M}}$, which is a subset of $\mathbf{M}$ containing constituent column vectors, leading to a rectangular matrix. As illustrated in Figure S4(a), with $\ddot{\mathbf{M}}$, we can reconstruct a speckled image on an OP from a computed reflection matrix, $\ddot{\mathbf{R}}$, by respectively left and right multiplying $\ddot{\mathbf{M}}$ with full $\mathbf{T}^{\text{-T(tik)}}$ and $\ddot{\mathbf{T}}^{\text{-1(tik)}}$, which is the regularized inverse of a subset of $\mathbf{T}$ with constituent column vectors at input channels corresponding to $\ddot{\mathbf{M}}$. Physically speaking, the image derived from the partial measurement corresponds to the distal sample under statistically non-uniform illumination. Using confocal intensity images $\mathbf{I}$ for demonstration here, we define the completeness of an intermediate image as the normalized intensity correlation, C, with the final image reconstructed from full $\mathbf{M}$ measurement,

$$C = \frac{\sum_{x,y} \mathbf{I}_i(x,y)\mathbf{I}_f(x,y)}{\sum_{x,y} \mathbf{I}_i(x,y) \sum_{x,y} \mathbf{I}_f(x,y)}, \tag{S8}$$

where $\mathbf{I}_i$ and $\mathbf{I}_f$ are intermediate and final images, respectively. The completeness arrives at $C = 1$ when $\mathbf{I}_i = \mathbf{I}_f$. Figure S4(b) shows examples of intermediate images with their quantified completeness. Here, the sample is a resolution chart, and the full $\mathbf{M}$ is a 1354-by-1354 square matrix. We assume that the 1354 proximal input spots uniformly couple to the 1100 MMF guided modes and define the compression ratio as $1 - m/1354$, where m is the number of input realizations. We tested two input channel sampling orders: the original proximal scanning spot basis order (blue curve) and a random sampling order (orange curve). From the plot, we can see that the completeness quickly improves with the number of input realizations and achieves 0.9 with $\sim$200 and $\sim$80 input realizations, which are only $\sim$15% and $\sim$5.9% of the total number of realizations (85.2% and 94.1%compression ratio) in the two ordering conditions, respectively. The random sampling order has a steeper completeness compared to the original since any input channel is less correlated with the next. The intermediate images start from speckled pattern and evolve to clean and high-contrast final confocal intensity image.

6. 3D CONFOCAL IMAGES WITH VARIOUS CONTRASTS

To demonstrate 3D imaging of biological samples through the MMF based on numerical refocusing, a sample with multiple layers was prepared following similar volumetric reconstruction experiments performed by others [3–5]. A proximal reflectance measurement of $\mathbf{M}$ through the MMF included reflectance from multiple layers of a sample, shown in Fig. S5(top), including buccal epithelial cells deposited on both surfaces of a glass coverslip with thickness of $\sim$200 µm, placed at $d = 120$ µm in air. From this single $\mathbf{M}$, 3D volumetric imaging was computed by numerical refocusing and image reconstruction. The depth-dependent γ plot was consistent with the physical location of each reflective interface, considering the refractive indices of each layer (1.44 in glass). High-resolution confocal images with intensity, phase, and scattering contrasts were computed at the two individual coverslip surfaces ($d = 120$ and 320 µm). Note that in complex samples, because optical phase accumulates as light is reflected from further into the sample, the phase of shallower cells is overlaid on deeper-lying cells.

REFERENCES

1. S.-Y. Lee, V. J. Parot, B. E. Bouma, and M. Villiger, "Reciprocity-induced symmetry in the round-trip transmission through complex systems," APL Photonics **5**, 106104 (2020).
2. T. J. F. Kenneth R. Spring and M. W. Davidson, "Resolution and contrast in confocal microscopy," Olymp. - Life Sci. Solutions .
3. V. Tsvirkun, S. Sivankutty, G. Bouwmans, O. Katz, E. R. Andresen, and H. Rigneault, "Wide-field lensless endoscopy with a multicore fiber," Opt. Lett. **41**, 4771–4774 (2016).
4. J. Shin, D. N. Tran, J. R. Stroud, S. Chin, T. D. Tran, and M. A. Foster, "A minimally invasive lens-free computational microendoscope," Sci. Adv. **5** (2019).
5. W. Choi, M. Kang, J. H. Hong, O. Katz, Y. Choi, and W. Choi, "Fourier holographic endoscopy for label-free imaging through a narrow and curved passage," arXiv 2010.11776 (2020).

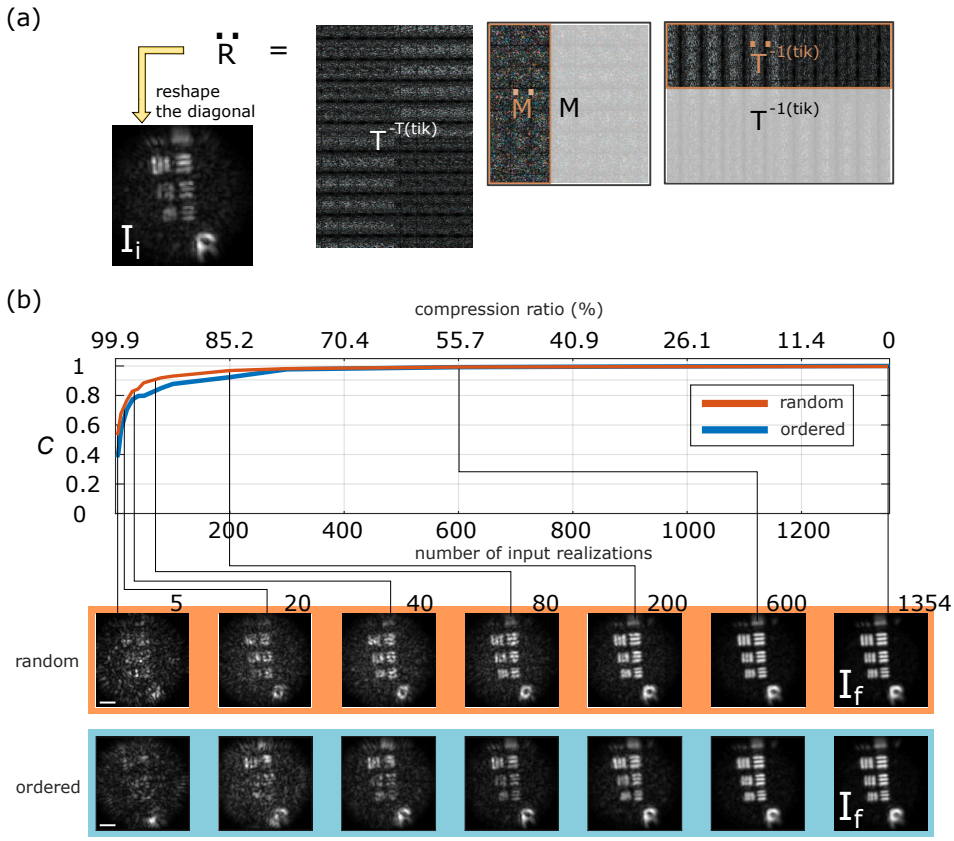

Fig. S4. Confocal intensity image reconstruction with partial measurement (a) Illustration of computational image reconstruction with the MMF forward TM, **T**, and a partially measured round-trip reflection matrix, $\ddot{\mathbf{M}}$ (b) Reconstruction results and image correlation with increasing round-trip measurement completeness in the ordered (blue curve) or a random (orange curve) proximal input channel order. The scale bar is 20 μm.

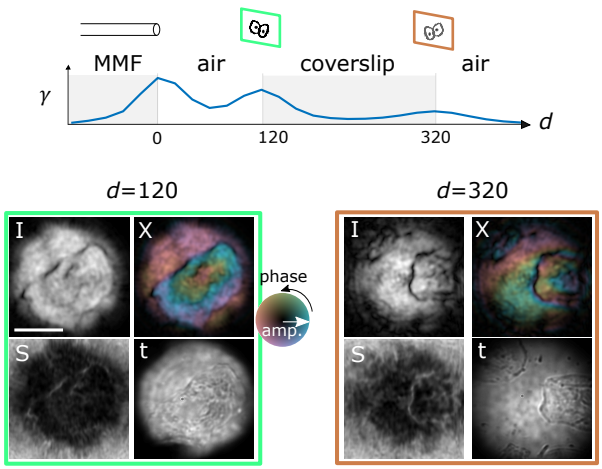

Fig. S5. Label-free 3D computational imaging through the MMF with multiple contrasts. The sample included two layers of buccal epithelial cells deposited on both surfaces of a glass coverslip in air. The d position indicates physical distance and is in the unit of μm. The scale bars are 50 μm.