CONTENTS

I. MEASUREMENT DETAILS

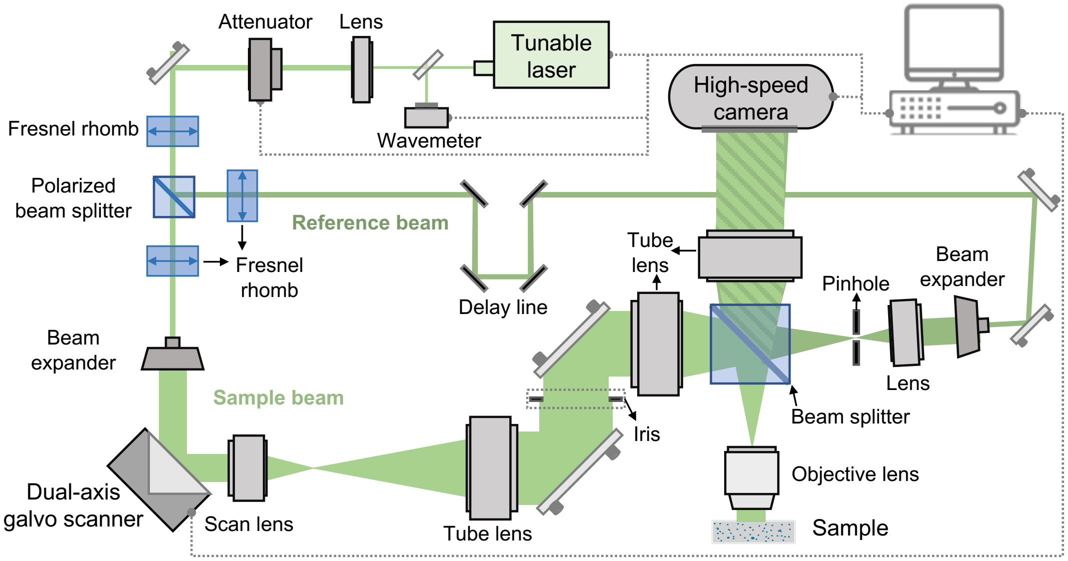
A. Optical setup

In this section, we describe the detailed experimental setup for measuring the hyperspectral reflection matrix (**Supplementary Fig. 1**). We use off-axis holography in the Mach–Zehnder configuration [53, 79] to measure the complex field information including amplitude and phase. The light source is generated by a tunable Ti:Sapphire laser (M Squared SolsTiS 1600) and is polarized, the frequency of which can be scanned and is monitored by a wavemeter (HighFinesse Wavemeter WS6-600). To balance the power at different frequencies, we use a motorized attenuator (Newport Motorized Variable Beam Splitter VA-CB-2-CONEX) to produce a frequency-dependent attenuation.

The laser beam is split into a sample beam and a reference beam through a polarized beam splitter (Newport Polarizing Cube Beamsplitter 05FC16PB.5). The sample beam is focused onto the back focal plane of an objective lens (Mitutoyo M Plan Apo NIR 100X) to create a plane-wave-like directional illumination onto the sample, with the reference plane being the focal plane of the objective lens. The angle of illumination is scanned by a dual-axis galvo scanner (ScannerMAX Saturn 5B). After the galvo scanner, we use a telecentric system composed of one scan lens (Thorlabs Scan Lens SL50-2P2) and two tube lenses (Thorlabs Tube Lens TTL200MP) to create a sharp focus on the back focal plane of the objective lens. Between the two tube lenses, there is a plane conjugate to the focal plane of the objective lens, where we place an iris diaphragm to control the illumination area on the sample. The scattered light from the sample is measured by a microscope imaging system composed of the objective lens, another tube lens (Thorlabs Tube Lens TTL200-S8), and a high-speed camera (Photron FASTCAM NOVA S6). The camera sensor is located on a plane that is also conjugate to the objective focal plane. When the objective lens focuses light on a mirror or a sample surface, the iris in the incidence path can be clearly imaged on the camera. The numerical aperture of the objective lens is NA = 0.5. With a magnification of 100X and a camera pixel size of 20 µm, the detection system provides a pixel resolution of $\delta x_{\text{pixel}} = 0.2$ µm; this pixel resolution is smaller than the diffraction limit in order to resolve the interference fringes in off-axis holography (Supplementary Sect. I C). The camera sensor has a pixel count of 1024×1024. In this work, we use a region of interest of 256×256 pixels, corresponding to a field of view (FOV) of 51.2×51.2 µm^2.

By using plane-wave-like (unfocused) illumination in momentum space $\mathbf{k}_{\parallel}^{\text{in}}$, we avoid nonlinear response or photo-damage of the sample. By using a real-space detection $\mathbf{r}_{\parallel}^{\text{out}}$, we avoid localized saturation of the camera, allowing us to tune the power to optimize the signal-to-noise ratio.

The sample is mounted on a 3-axis (XYZ) translation stage, with a motorized actuator (Newport Motorized Actuator TRA6CC) for controlling and monitoring the z position of the sample. A separate camera (Edmund EO-0514) images the sample from the other side to facilitate selecting the region of interest in the sample. We use a switchable sample holder (Thorlabs Multi-Position Sliders ELL9K) to switch between samples and/or a mirror. For an accurate calibration of the incident angles (Supplementary Sect. III A), the mirror is placed perpendicular to the optical axis



Supplementary Fig. 1. **Detailed experimental setup.**

of the system. To this end, we first adjust the galvo scanner so that the part of the incident beam reflected by the back end of the objective lens follows the optical axis and reaches the center of the camera sensor. This indicates that the incident beam aligns with the optical axis and is focused at the center of the objective lens. Another portion of the beam passes through the objective lens and is reflected by the mirror. The camera captures the resulting interference between reflections from the objective lens and the mirror. When the mirror's reflection deviates from the optical axis, interference manifests as a striped pattern. Adjustments are made to the mirror's tilting angle to maximize interference fringe spacing, aligning the mirror reflection with the optical axis. When the mirror plane is perpendicular to the optical axis, the interference pattern transitions into concentric circles, characterized by a maximum fringe spacing.

The scattered light from the sample is combined with the reference beam at an angle to create interference patterns which are recorded by the camera (Supplementary Sect. I C). The reference beam arrives at the camera as a collimated beam, the size of which is matched with the sample beam size. To produce a clean and spatially uniform reference beam profile, we place a pinhole on the focal plane of the tube lens.

A delay line in the reference arm is used to match the optical path lengths of the two interfering beams. To set the delay, we first use a ruler to do a rough measurement of the optical path length difference ΔL between the two beams with an accuracy of a few centimeters. For finer tuning, we examine the reflection phase of a mirror measured by off-axis holography (Supplementary Sect. I C). The residual path length difference ΔL produces a linear frequency-dependent phase $e^{i(\omega/c)\Delta L}$ where c is the speed of light, which we measure with a fine frequency spacing (with the etalon scan mode of the laser).

The reflected sample beam and the reference beam are combined via a non-polarized beam splitter. Some incident light in the sample beam is internally reflected on the surfaces of the beam splitter and may enter the camera without arriving at the sample. To remove such unwanted signal, we slightly rotate the beam splitter by about 4 degrees to create a misalignment [80] so that internal reflections from its surfaces do not enter the camera sensor.

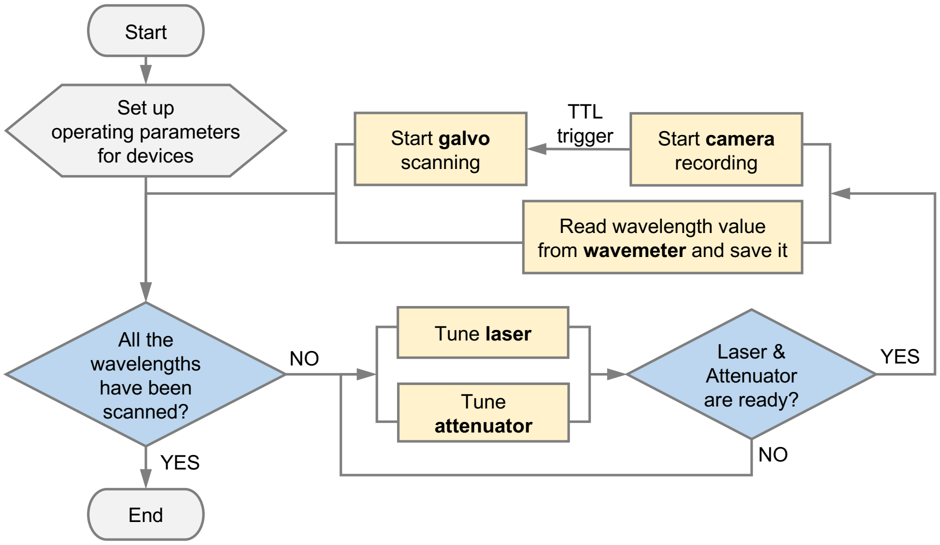
To balance the power between the two interfering beams, we adjust the Fresnel rhomb before the polarized beam splitter so that the reference beam intensity is comparable to the scattered light intensity from a piece of white paper.

The two Fresnel rhombs after the polarized beam splitter are used to control the polarization of the reflection matrix: the one on the sample beam adjusts the incident polarization, and the one on the reference beam adjusts the detected polarization. In this work, both polarization states are set to horizontal.

To improve the phase stability (characterized in Supplementary Sect. II B), we enclose the setup to reduce air flow and to block acoustic noises. We control the setup remotely, which also avoids vibrations caused by onsite operations.

B. Control and automation

We automate the measurement with a LabVIEW program that controls the laser, attenuator, wavemeter, galvo scanner, and camera. The flow chart of this process is shown in **Supplementary Fig. 2**.



Supplementary Fig. 2. Flow chart of the automated measurement.

A while loop traverses all frequencies. In each iteration of the loop, we first tune the laser and the attenuator simultaneously. Currently, we scan the laser frequency by mechanically tuning a birefringent filter in a scan-and-stop mode. The filter position is set based on a lookup table, and we use the wavemeter to verify the actual output frequency (with an accuracy of 600 MHz). It is also possible to scan the birefringent filter continuously without stopping, in which case the laser output will jump between discrete frequencies with 0.286-THz spacing at a scanning rate of about 20 msec per frequency. Note that SMT adopts a non-uniform discrete Fourier transform, so it is not necessary to have the frequencies equally spaced.

The recorded image intensity can change substantially across frequencies due to frequency dependencies of the laser power and of the camera sensor efficiency. To maximize the signal-to-noise ratio, it is best to balance the measured intensity across frequencies such that the camera photon count is close to uniformly high but does not saturate the sensors. Therefore, we use an attenuator (Newport Motorized Variable Beam Splitter VA-CB-2-CONEX) to provide a frequency-dependent attenuation by rotating a half wave plate.

We operate the camera with a frame rate of 64,000 Hz and a region of interest of 256×256 pixels (corresponding to a FOV of 51.2×51.2 µm^2 on the sample). To avoid saturating the camera, we use an exposure time of 1/300,000 second, smaller than the inverse frame rate. At each frequency, the camera records 5030 frames (corresponding to 5030 incident angles). The camera sends a TTL signal to trigger the galvo and start the angular scan, with a delay of less than 0.2 msec. To operate the dual-axis galvo scanner, we use a multi-function data acquisition device (DAQ) (Measurement Computing DAQ USB-1608GX-2AO). One analog input of the DAQ receives the trigger signal from the camera, and two analog outputs of the DAQ controls the operating voltages for the x and y scanner heads with a sample rate of 400,000 points per second. We use a triangular waveform for the fast x-axis scan and a step-shaped waveform with 80 steps for the slower y-axis scan, with a bidirectional raster pattern over a circular region in the angular space that is slightly larger than the back focal plane of the objective lens.

C. Off-axis holography

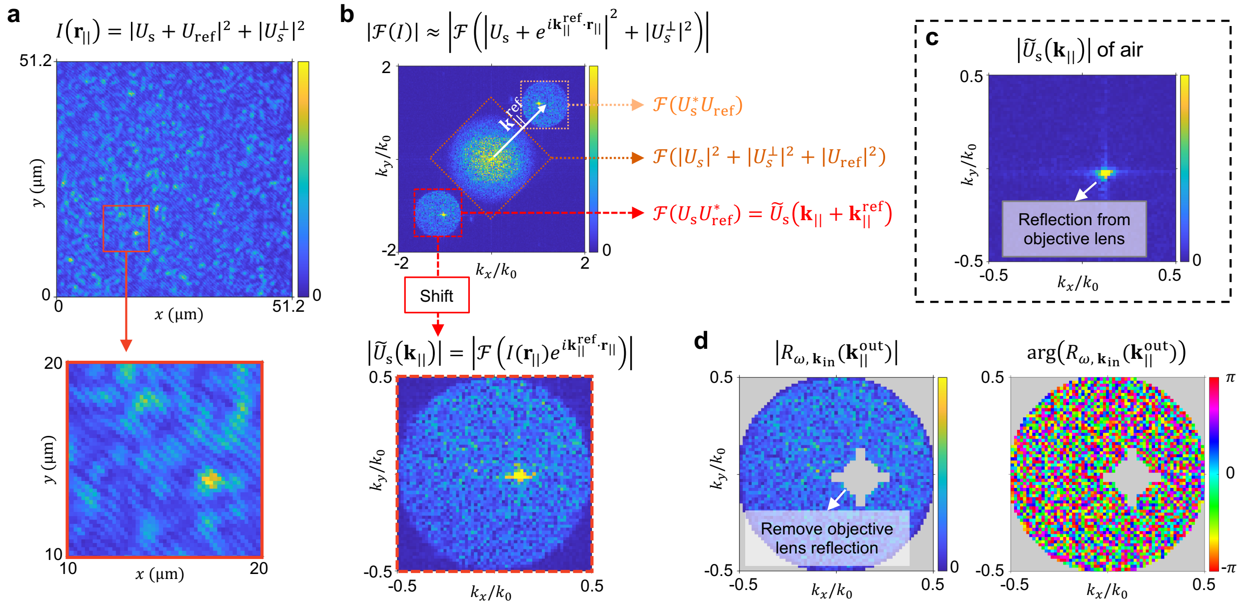
We use digital off-axis holography [53] to retrieve the complex-valued reflected fields of the sample beam, containing both amplitude and phase information. The interference pattern $I(\mathbf{r}_{\|})$ ($\mathbf{r}_{\|}$ is the transverse position) recorded by the camera is a superposition of the reference beam $U_{\rm ref}(\mathbf{r}_{\|})\hat{e}_{\rm ref}$ in $\hat{e}_{\rm ref}$ polarization and the sample beam $U_{\rm s}(\mathbf{r}_{\|})\hat{e}_{\rm ref} + U_{\rm s}^{\perp}(\mathbf{r}_{\|})\hat{e}_{\perp}$ which may contain both $\hat{e}_{\rm ref}$ and its orthogonal $\hat{e}_{\perp}$ polarization. The reference beam is close to a plane wave propagating along $\mathbf{k}_{\rm ref}$ direction, so that $I(\mathbf{r}_{\|}) \approx \left|U_{\rm s}(\mathbf{r}_{\|}) + e^{i\mathbf{k}_{\|}^{\rm ref}\cdot\mathbf{r}_{\|}}\right|^2 + \left|U_{\rm s}^{\perp}(\mathbf{r}_{\|})\right|^2$. **Supplementary Fig. 3**a shows an example of $I(\mathbf{r}_{\|})$ with the sample being a piece of white paper; the interference fringes show up as the anti-diagonal stripes in the zoom-in. We take the 2D Fourier transform of $I(\mathbf{r}_{\|})$ with respect to $\mathbf{r}_{\|} = (x, y)$,

$$\begin{aligned}\mathcal{F}(I) &= \mathcal{F}\left(|U_{\rm s}|^2 + \left|U_{\rm s}^{\perp}\right|^2 + |U_{\rm ref}|^2\right) + \mathcal{F}\left(U_{\rm s}^* U_{\rm ref}\right) + \mathcal{F}\left(U_{\rm s} U_{\rm ref}^*\right) \\ &\approx \mathcal{F}\left(|U_{\rm s}|^2 + \left|U_{\rm s}^{\perp}\right|^2 + |U_{\rm ref}|^2\right) + \tilde{U}_{\rm s}^*(\mathbf{k}_{\|} - k_{\|}^{\rm ref}) + \tilde{U}_{\rm s}(\mathbf{k}_{\|} + \mathbf{k}_{\|}^{\rm ref}),\end{aligned} \tag{S1}$$

where $\mathbf{k}_{\|} = (k_x, k_y) = \mathbf{k}_{\|}^{\rm out}$ is the in-plane components of the outgoing momentum. The first term of Eq. (S1) forms the zeroth-order signal. The latter two terms (which contain the complex field of the sample beam $\tilde{U}_{\rm s}$ in momentum space) form the $\pm 1^{\rm st}$-order signal and are shifted by $\pm\mathbf{k}_{\|}^{\rm ref}$ in the $\mathbf{k}_{\|}$ space. When $|\mathbf{k}_{\|}^{\rm ref}|$ is sufficiently large, the $\pm 1^{\rm st}$-order signals will be separated from the zeroth-order signal in $\mathbf{k}_{\|}$ space (upper figure of **Supplementary Fig. 3**b), and we can obtain $\tilde{U}_{\rm s}$ by truncating the $-1^{\rm st}$-order signal and shifting it back to the center of the Fourier space.

An opaque medium generally scatters light toward all possible outgoing directions, so the sample field $\mathcal{F}(U_{\rm s})$ covers a disk with radius $k_0 \cdot {\rm NA}$ in the momentum space (where $k_0 = 2\pi/\lambda$ and λ is the vacuum wavelength), limited only by the NA of the objective lens. Similarly, $\mathcal{F}\left(|U_{\rm s}|^2 + \left|U_{\rm s}^{\perp}\right|^2\right)$ covers a disk with radius $k_0 \cdot 2{\rm NA}$ in the momentum space. To separte the zeroth-order and the $\pm 1^{\rm st}$-order terms, the angle of the reference beam must be large enough that $|\mathbf{k}_{\|}^{\rm ref}| > k_0 \cdot 3{\rm NA}$. An angle much larger than this threshold is not desirable since it makes the interference fringes narrower and harder to resolve. Specifically, the maximal $|k_x|$ and $|k_y|$ components of $\mathcal{F}(I)$ must be smaller than $\pi/\delta x_{\rm pixel}$ to be resolved by pixel resolution $\delta x_{\rm pixel}$. Here, we set $|\mathbf{k}_{\|}^{\rm ref}| \approx 1.65 k_0$ (corresponding to a reference beam angle of about 1° prior to the 100X magnification), while $3{\rm NA} = 1.5$. To reduce the maximal $|k_x|$ and $|k_y|$, we use an azimuthal angle of $\angle\mathbf{k}_{\|}^{\rm ref} \approx 45°$. In this case, ${\rm max}|k_x| = k_0{\rm NA} + |\mathbf{k}_{\|}^{\rm ref}|/\sqrt{2}$.

From $\pi/\delta x_{\rm pixel} > {\rm max}|k_x|$ and $|\mathbf{k}_{\|}^{\rm ref}| > k_0 \cdot 3{\rm NA}$, we get the largest pixel resolution one may use for off-axis holography is $\delta x_{\rm pixel} < \lambda/[2(1 + \frac{3}{\sqrt{2}}){\rm NA}] \approx \lambda/(6.24{\rm NA})$.



Supplementary Fig. 3. Complex field retrieval with off-axis holography. a. The interference pattern $I(\mathbf{r}_{\|})$ formed by the scattered light from a piece of white paper $U_{\mathrm{s}}\hat{e}_{\mathrm{ref}} + U_{\mathrm{s}}^{\perp}\hat{e}_{\perp}$ and the reference beam $U_{\mathrm{ref}}\hat{e}_{\mathrm{ref}}$. **b.** 2D Fourier spectrum $\mathcal{F}(I)$. There are three circular bright regions: the middle one contains the zeroth order and the other two are the $\pm 1^{\mathrm{st}}$ orders, $\mathcal{F}(U_{\mathrm{s}}^* U_{\mathrm{ref}})$ and $\mathcal{F}(U_{\mathrm{s}} U_{\mathrm{ref}}^*)$. With a plane-wave-like reference beam $U_{\mathrm{ref}}(\mathbf{r}_{\|}) \approx e^{i\mathbf{k}_{\|}^{\mathrm{ref}}\cdot\mathbf{r}_{\|}}$, the minus one order is approximately the sample field $\tilde{U}_{\mathrm{s}}$ shifted by $-\mathbf{k}_{\|}^{\mathrm{ref}}$. Thus, $\tilde{U}_{\mathrm{s}}$ can be obtained by shifting the minus one order back to the center of Fourier space. **c.** The retrieved field when the sample is removed (*i.e.*, measuring the reflection of air) shows a peak that comes from reflection by the objective lens. **d.** The retrieved complex scattered field from the sample. Data outside the NA range and near the objective lens reflection are excluded (gray regions).

A white-paper sample produces sufficient large-angle scattered light that the perimeter and the center of the $\pm 1^{\mathrm{st}}$-order terms can be accurately identified, from which we obtain $\mathbf{k}_{\|}^{\mathrm{ref}}$. When using discrete Fourier transforms, the pixel size in momentum space is $\delta k = 2\pi/\mathrm{FOV}$. To obtain a $\tilde{U}_{\mathrm{s}}$ centered at $\mathbf{k}_{\|}^{\mathrm{out}} = 0$ without restriction of the discrete δk, we multiply $I(\mathbf{r})$ with $e^{i\mathbf{k}_{\|}^{\mathrm{ref}}\cdot\mathbf{r}_{\|}}$ before taking the 2D discrete Fourier transform.

For the reflection matrix, we only keep data points of $\tilde{U}_{\mathrm{s}}(\mathbf{k}_{\|}^{\mathrm{out}})$ with $|\mathbf{k}_{\|}^{\mathrm{out}}| < k_0 \cdot \mathrm{NA}$. Within the NA range, the sample beam also includes unwanted reflected light from the objective lens. For an incident plane wave at direction $\mathbf{k}_{\mathrm{in}}$, such objective lens reflection shows up mainly near $\mathbf{k}_{\|}^{\mathrm{out}} = -\mathbf{k}_{\|}^{\mathrm{in}}$ in Fourier space; we can clearly see it when we remove the sample (*i.e.*, measure air only) as shown in **Supplementary Fig. 3**c. For the reflection matrix data, we exclude $\mathbf{k}_{\|}^{\mathrm{out}}$ near such objective lens reflection angles. We obtain the location of the objective lens reflection by searching for the maximum in the region near $\mathbf{k}_{\|}^{\mathrm{out}} = -\mathbf{k}_{\|}^{\mathrm{in}}$. Such objective lens reflection signal spreads more in the Fourier space when the incident angle is closer to normal incidence, so we also exclude more $\mathbf{k}_{\|}^{\mathrm{out}}$ there. Note that such exclusion does not fully remove the reflection from objective lens; the residual objective lens reflection is currently the main limiting factor on the sensitivity of our detection (Supplementary Sect. II C). **Supplementary Fig. 3**d shows the amplitude and phase of $\tilde{U}_{\mathrm{s}}(\mathbf{k}_{\|}^{\mathrm{out}})$, with the excluded data points shown in gray.

This $\tilde{U}_{\mathrm{s}}(\mathbf{k}_{\|}^{\mathrm{out}}) = R(\mathbf{k}_{\|}^{\mathrm{out}}, \mathbf{k}_{\|}^{\mathrm{in}}, \omega)$ forms one column of the hyperspectral reflection matrix. Note that $|R(\mathbf{k}_{\|}^{\mathrm{out}}, \mathbf{k}_{\|}^{\mathrm{in}}, \omega)|^2$ is proportional to the detected power but not the reflected power since the camera sensor efficiency varies with the wavelength. The incident power also varies with both the wavelength and the incident angle. We do not normalize $R(\mathbf{k}_{\|}^{\mathrm{out}}, \mathbf{k}_{\|}^{\mathrm{in}}, \omega)$ by the incident power and sensor efficiency since doing so increases both noise and acquisition time.

Because we scan incident angles across a range that is slightly larger than the NA of the objective lens (Supplementary Sect. I B), some incident angles do not go through the objective lens. To exclude such incident angles, we exclude inputs where the mirror reflection does not show a strong peak. We also exclude the first 10 inputs at each wavelength where the galvo scanning has not yet started.

After the above exclusions, the number of input angles averages to around 3900 (ranging between 3830 to 3900 at different frequencies), and the number of output angles averages to around 2900 (ranging between 2210 to 3590). Note that the incident angles are slightly over-sampled, with spacing finer than $2\pi/\mathrm{FOV}$.

D. Data acquisition time

It takes 180 seconds to complete the data acquisition across the 250 frequencies × 5030 input angles for the USAF target under tissue sample, with data captured at a resolution of 256×256 pixels for each incident angle. This speed is sufficient since the phase drift in the system (Supplementary Sect. II B) is only $\pi/8$ radian per minute and can be removed during dispersion compensation. The angular scan (running at the camera frame rate of 64 kHz; Supplementary Sect. I B) took 79 msec per frequency, totaling 20 seconds for the 250 frequencies. The remaining 160 seconds are used for the laser frequency tuning. To simplify the control, we currently separate the frequency scan and the angular scan: the camera recording starts only after the laser and the attenuator report back as having completed their tuning, and the next frequency tuning starts only after the previous angular scan completes. The birefringent filter used for the frequency tuning needs to accelerate and decelerate for every frequency requested, which contributes to most of the overall measurement time.

As mentioned in Supplementary Sect. II B, one can also continuously scan the birefringent filter without stopping, which results in the laser output jumping between discrete frequencies with 0.286-THz spacing at a scan rate of 20 msec per frequency. If one has a camera that operates at 250 kHz frame rate (*e.g.*, Phantom TMX-5010 with 640×320 pixels [81]) and scans the camera, galvo, and laser concurrently, the data acquisition across the 250 frequencies × 5030 input angles can complete in 5 seconds.

II. CHARACTERIZATION OF RAW DATA

A. Spectrum

From the measured hyperspectral reflection matrix $R(\mathbf{k}_\parallel^{\text{out}}, \mathbf{k}_\parallel^{\text{in}}, \omega)$, we calculate the angle-averaged detected power as a function of the wavelength,

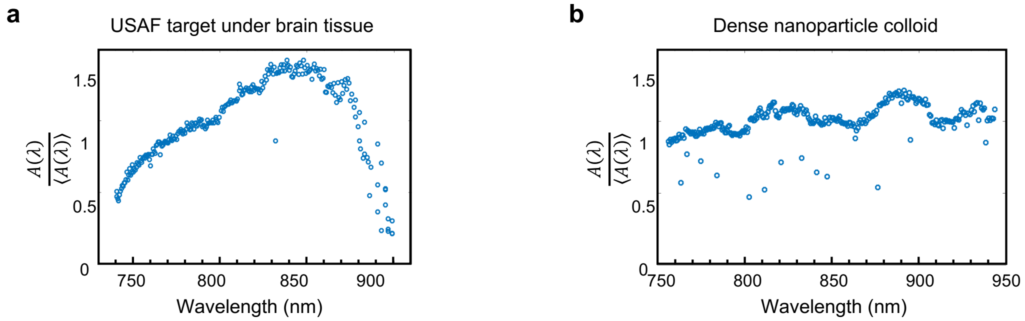
$$A(\lambda) = \sqrt{\frac{1}{N_{\text{in}} N_{\text{out}}} \sum_{\mathbf{k}_\parallel^{\text{in}}, \mathbf{k}_\parallel^{\text{out}}} \left| R\left(\mathbf{k}_\parallel^{\text{out}}, \mathbf{k}_\parallel^{\text{in}}, \omega = \frac{2\pi}{\lambda}\right)\right|^2}. \tag{S2}$$

Supplementary Fig. 4 shows the normalized $A(\lambda)$ for the USAF-target-under-tissue sample and the dense nanoparticle colloid studied in this work. The former was measured at 247 frequencies with λ from 741 nm to 899 nm, and the latter was measured at 275 frequencies from 756 nm to 943 nm. Both were measured with a frequency spacing of $\delta f = 0.286$ THz. The two spectra differ because the laser output changes between the two measurements.

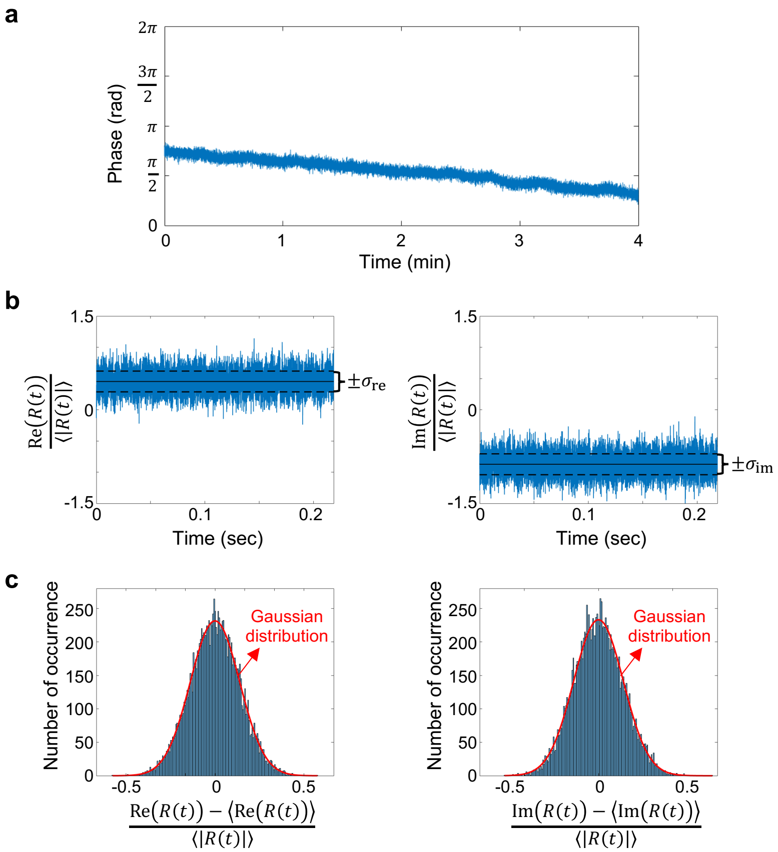
The power incident on the sample, measured with a power meter at the sample position, ranges from 0.02 mW and 0.2 mW depending on the wavelength.

B. Phase stability and noise

To evaluate the phase stability, we monitor how the measured reflection phase of a mirror changes over time (**Supplementary Fig. 5**a). In the short term, the phase fluctuates with a standard deviation of 0.09 radian. In the long term, the phase drifts slowly at a pace of about $\pi/8$ radian per minute, likely due to temperature change.



Supplementary Fig. 4. Angle-averaged reflection matrix spectra given by Eq. (S2).



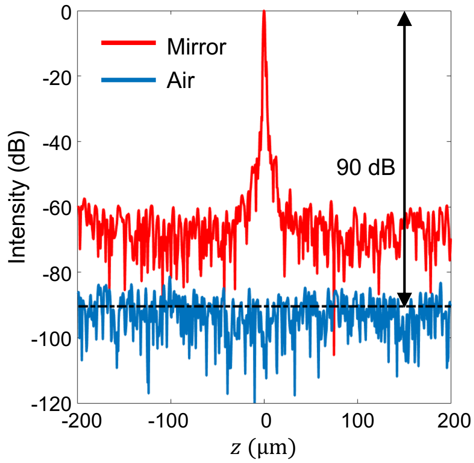
Supplementary Fig. 5. **Phase stability and noise.** **a.** Long-term evolution of the measured reflection phase of a static mirror, with $x = y = 25$ µm. **b.** Measured reflected field from a static white-paper sample at $\mathbf{k}_{\parallel} = (-0.28k_0, -0.24k_0)$. **c.** Histogram of the measured field in **b**, showing a Gaussian distribution.

To quantify the noise, we examine the short-term fluctuations of the reflected fields from a white-paper sample (**Supplementary Fig. 5**b). These short-term fluctuations come from a combination of vibrations, air flow, dither noise in the galvo scanner, dark current and read noise and photon shot noise at the camera sensor, and phase and amplitude noise from the laser. Both the real part and the imaginary part follow a Gaussian distribution (**Supplementary Fig. 5**c). The averaged standard deviation, $\frac{1}{N_{\mathrm{out}}}\sum_{\mathrm{out}}\frac{1}{2}(\sigma_{\mathrm{real}}+\sigma_{\mathrm{imaginary}})$, is around 1% of the averaged amplitude $\sqrt{\frac{1}{N_{\mathrm{out}}}\sum_{\mathrm{out}}\langle|R|^2\rangle}$. Note that this is the noise per element of the reflection matrix. The SMT image is built by summing over many (around two billion in our experiment) matrix elements, after which the signal-to-noise ratio (SNR) is boosted significantly. We quantify the SNR of the SMT image in the next section.

C. Sensitivity of SMT

The sensitivity is defined as one over the smallest reflectivity at which SNR = 1. It is common to assume the SNR to be proportional to the sample reflectivity, under which the sensitivity equals the SNR when the sample is a mirror with unity reflectivity; we use this approach here. We take three hyperspectral reflection matrix measurements under the same condition: the first two are performed for two different mirror samples and the third one is performed for air with no sample. We use the first mirror measurement to correct for the dispersion and aberrations in the optical system (Supplementary Sect. III B) and perform SMT reconstructions using each of the three data sets to obtain a self-corrected mirror image $I_{\mathrm{SMT}}^{\mathrm{mirror,self}}$, a cross-corrected mirror image $I_{\mathrm{SMT}}^{\mathrm{mirror,cross}}$, and an air image $I_{\mathrm{SMT}}^{\mathrm{air}}$. Comparing the peak of $I_{\mathrm{SMT}}^{\mathrm{mirror,cross}}$ and the average of $I_{\mathrm{SMT}}^{\mathrm{air}}$, we find the sensitivity to be 90 dB (**Supplementary Fig. 6**).

The SMT sensitivity may be limited by both measurement noises (as characterized in the previous section) and the residual reflection from the objective lens. In the self-corrected mirror image, most of the phase noises are removed during the correction stage to yield a constant phase, so $I_{\mathrm{SMT}}^{\mathrm{mirror,self}}$ provides an estimate of the would-be peak intensity



Supplementary Fig. 6. **Sensitivity of SMT.** Normalized axial intensity of SMT images of a mirror with cross-correction $I_{\mathrm{SMT}}^{\mathrm{mirror,cross}}$ and of air $I_{\mathrm{SMT}}^{\mathrm{air}}$. The intensities are evaluated at the center of the FOV, at $x = y = 25$ μm. The peak of $I_{\mathrm{SMT}}^{\mathrm{mirror,cross}}$ is 90 dB higher than the averaged air intensity (which comes primarily from the residual reflection of the objective lens).

in the absence of noises. We find the peak of $I_{\mathrm{SMT}}^{\mathrm{mirror,self}}$ to be 2.5 dB higher than that of $I_{\mathrm{SMT}}^{\mathrm{mirror,cross}}$, indicating that noises misalign the phases to reduce the SMT signal by roughly a factor of 2.

If reflection matrix data with $\mathbf{k}_{\parallel}$ near the peak of the objective lens reflection signal is not excluded (Supplementary Sect. I C), $I_{\mathrm{SMT}}^{\mathrm{air}}$ would be 12 dB higher, showing the importance of such exclusion. However, such exclusion does not fully remove the objective's reflection because the signal from the objective's reflection also exists at other $\mathbf{k}_{\parallel}$ points; several factors contribute: the incident beam does not form a point focus on the back focal plane of the objective lens due to aberrations of the optical system and the laser beam profile, and reflection from the objective lens can come from multiple interfaces beyond the back focal plane. In the absence of noise and the objective's reflection, $I_{\mathrm{SMT}}^{\mathrm{air}}$ would be zero. Therefore, to estimate the contribution of the residual objective lens reflection, we perform a fourth reflection matrix measurement with the laser turned off, which fully removes the objective's reflection while preserving the camera's dark current and read noise. The resulting laser-off SMT image is 9 dB lower than $I_{\mathrm{SMT}}^{\mathrm{air}}$. This indicates that the residual reflection from the objective lens contributes more to our SMT sensitivity than noises do. In our current measurement, the objective lens reflection shows up as a coherent background because the coherence length of our tunable CW laser is far greater than the path difference between the objective lens and the sample. If a sensitivity much larger than 90 dB is necessary, one may reduce the coherence length of the laser and/or clean the incident beam to make the objective-lens-reflection exclusion more complete.

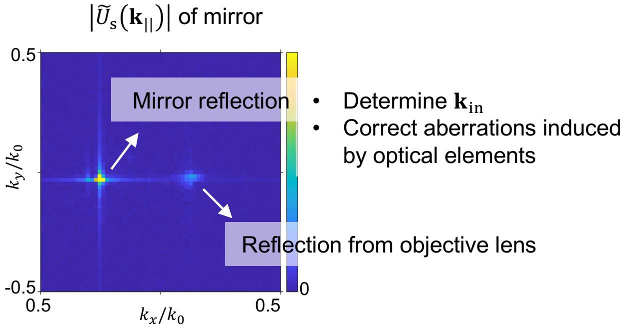
III. DIGITAL CORRECTIONS AND CALIBRATIONS

A. Incident angle calibration

In addition to the sample of interest, we also measure the hyperspectral reflection matrix of a mirror (Thorlabs Protected Silver Mirror PF10-03-P01). The specular reflection from the mirror produces a strong peak in Fourier space (**Supplementary Fig. 7**) located at $\mathbf{k}_{\parallel}^{\mathrm{out}} = \mathbf{k}_{\parallel}^{\mathrm{in}}$, which we use to calibrate the incident angle $\mathbf{k}_{\parallel}^{\mathrm{in}}$.

We also use the mirror data to determine the precise location of the reflection from the objective lens in the momentum space (the weaker peak in **Supplementary Fig. 7**) so we can exclude data points at those $\mathbf{k}_{\parallel}^{\mathrm{out}}$ from our sample reflection matrix as described in Supplementary Sect. I C.

After such calibration, the input momentum and the output momentum of the reflection matrix both lie on the same grid points with spacing $\delta k = 2\pi/\mathrm{FOV}$ (Supplementary Sect. I C), for all frequencies. But the number of inputs and the number of outputs both depend on the frequency. Recall that the incident angles are slightly over-sampled (Supplementary Sect. I C), so some incident angles share the same mirror-reflection peak location. For the convenience of further processing, we average over data points with the same $\mathbf{k}_{\parallel}^{\mathrm{in}}$ and pad zeros at the excluded $\mathbf{k}_{\parallel}^{\mathrm{out}}$ and at large $\mathbf{k}_{\parallel}^{\mathrm{in/out}}$ such that the size of matrix $R(\mathbf{k}_{\parallel}^{\mathrm{out}}, \mathbf{k}_{\parallel}^{\mathrm{in}}, \omega)$ is 3761 x 3761 at all frequencies and that all frequencies share the same list of $\mathbf{k}_{\parallel}^{\mathrm{in}}$ (which equals the list of $\mathbf{k}_{\parallel}^{\mathrm{out}}$).



Supplementary Fig. 7. **Mirror reflection data.** The specular mirror reflection gives a strong peak in Fourier space located at $\mathbf{k}_{\|} = \mathbf{k}_{\|}^{\mathbf{in}}$. Subtracting the peak phase from the sample reflection matrix corrects the dispersion and most of the input aberrations in the optical system. The reflection from the objective lens shows up as a much weaker peak than the mirror peak.

B. Reference plane and corrections for input aberrations and dispersion in the optical system

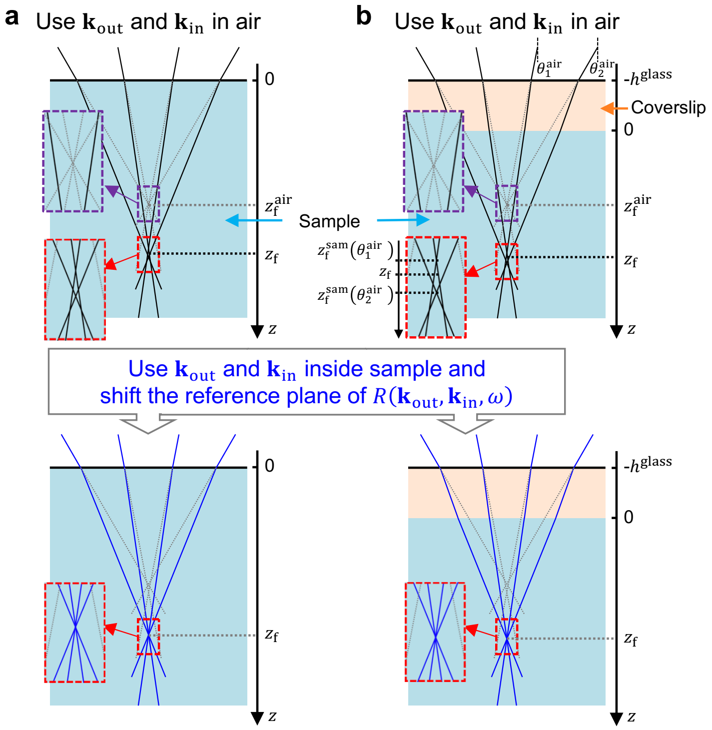
With the off-axis holography (Supplementary Sect. I C), the reflection matrix is associated with outgoing plane waves in air, $e^{i\mathbf{k}_{\text{out}}^{\text{air}}\cdot(\mathbf{r}-\mathbf{r}_{\text{ref}}^{\text{air}})}$, such that different $\mathbf{k}_{\text{out}}^{\text{air}}$ share the same reference point $\mathbf{r}_{\text{ref}}^{\text{air}} = (0, 0, z_{\text{ref}}^{\text{air}})$ with $z_{\text{ref}}^{\text{air}}$ being the focal plane of the objective lens in air. It will be convenient if the incident plane waves also have the same reference plane, $e^{i\mathbf{k}_{\text{in}}^{\text{air}}\cdot(\mathbf{r}-\mathbf{r}_{\text{ref}}^{\text{air}})}$. However, the galvo scanning introduces an additional propagation phase shift $\Delta\phi_{\text{scan}}(\mathbf{k}_{\|}^{\text{in}}, \omega)$ that varies with the incident angle $\mathbf{k}_{\|}^{\text{in}}$ and frequency ω. Since the reflection matrix of the sample and that of the mirror share the same $\Delta\phi_{\text{scan}}(\mathbf{k}_{\|}^{\text{in}}, \omega)$, we record the phase $\phi_{\text{mirror}}(\mathbf{k}_{\|}^{\text{in}}, \omega) \equiv \arg\left[R_{\text{mirror}}(\mathbf{k}_{\|}^{\text{out}} = \mathbf{k}_{\|}^{\text{in}}, \mathbf{k}_{\|}^{\text{in}}, \omega)\right]$ at the specular reflection peak $\mathbf{k}_{\|}^{\text{out}} = \mathbf{k}_{\|}^{\text{in}}$ of the mirror reflection matrix, and subtract $\phi_{\text{mirror}}(\mathbf{k}_{\|}^{\text{in}}, \omega)$ from the phase of the sample reflection matrix for all $\mathbf{k}_{\|}^{\text{in}}$ and all ω. Hereafter, we use $R'(\mathbf{k}_{\|}^{\text{out}}, \mathbf{k}_{\|}^{\text{in}}, \omega)$ to denote the sample reflection matrix after this subtraction. This $R'(\mathbf{k}_{\|}^{\text{out}}, \mathbf{k}_{\|}^{\text{in}}, \omega)$ has its input reference plane and output reference plane both at the focal plane of the objective lens, which equals the z position of the calibration mirror, $z_{\text{ref}}^{\text{air}} = z_{\text{mirror}}$.

In addition to the propagation phase shift $\Delta\phi_{\text{scan}}(\mathbf{k}_{\|}^{\text{in}}, \omega)$ from galvo scanning, the mirror reflection matrix also contains the aberrations and dispersion in the optical system in the input path $\Delta\phi_{\text{in}}^{\text{sys}}(\mathbf{k}_{\|}^{\text{in}}, \omega)$ and those in the output path $\Delta\phi_{\text{out}}^{\text{sys}}(\mathbf{k}_{\|}^{\text{out}}, \omega)$. Therefore, $\phi_{\text{mirror}}(\mathbf{k}_{\|}^{\text{in}}, \omega) = \Delta\phi_{\text{scan}}(\mathbf{k}_{\|}^{\text{in}}, \omega) + \Delta\phi_{\text{in}}^{\text{sys}}(\mathbf{k}_{\|}^{\text{in}}, \omega) + \Delta\phi_{\text{out}}^{\text{sys}}(\mathbf{k}_{\|}^{\text{out}} = \mathbf{k}_{\|}^{\text{in}}, \omega)$. When we subtract $\phi_{\text{mirror}}(\mathbf{k}_{\|}^{\text{in}}, \omega)$ from the phase of the sample reflection matrix, we not only remove $\Delta\phi_{\text{scan}}(\mathbf{k}_{\|}^{\text{in}}, \omega)$; we also correct for the aberrations and dispersion in the input path $\Delta\phi_{\text{in}}^{\text{sys}}(\mathbf{k}_{\|}^{\text{in}}, \omega)$. However, $\Delta\phi_{\text{out}}^{\text{sys}}(\mathbf{k}_{\|}^{\text{out}}, \omega)$ got subtracted along the input angles $\mathbf{k}_{\|}^{\text{in}}$ instead of along the output angles $\mathbf{k}_{\|}^{\text{out}}$. Therefore, the phase of the calibrated reflection matrix $\arg\left[R'(\mathbf{k}_{\|}^{\text{out}}, \mathbf{k}_{\|}^{\text{in}}, \omega)\right]$ still contains aberrations and dispersion in the output path in the form of $\Delta\phi_{\text{out}}^{\text{sys}}(\mathbf{k}_{\|}^{\text{out}}, \omega) - \Delta\phi_{\text{out}}^{\text{sys}}(\mathbf{k}_{\|}^{\text{out}} = \mathbf{k}_{\|}^{\text{in}}, \omega)$.

C. Correction for refractive index mismatch

The objective lens provides a sharp focus only in the designed background medium. Dry lenses designed to operate in air without an immersion liquid are the most commonly available and the most convenient to use. But because of refraction, dry lenses focus poorly inside media with other refractive indices such as water. As illustrated in **Supplementary Fig. 8**a, a perfect focus to $(x_{\text{f}}, y_{\text{f}}, z_{\text{f}}^{\text{air}})$ in air gets diffracted to an imperfect focus at $(x_{\text{f}}, y_{\text{f}}, z_{\text{f}})$. The focus becomes poorer at deeper depths, and the situation is even more severe when a coverslip is used (**Supplementary Fig. 8**b). Correction for such degradation typically requires expensive specially designed objective lenses or additional hardware like spatial light modulators. In SMT, we correct for it digitally.

In the presence of refractive index mismatches, Eqs. (1–2) of the main text continue to hold, but we need to be careful in specifying the reference planes of the reflection matrix, the momenta $\mathbf{k}_{\text{in/out}}$, and the relative coordinates. To find the correct formalism, we start with a generalized framework. The incident and the outgoing waves are still



Supplementary Fig. 8. **Correction for refractive index mismatches. a.** A system with a single air-sample interface. When a dry objective lens is used, incident light from different angles can focus in air but do not focus well inside the medium due to refraction at the air-sample interface. **b.** A sample covered by a coverslip with thickness h^{glass}, where there is an air-glass interface and another glass-sample interface. When light is focused to $z_{\text{f}}^{\text{air}}$ in air, the focal plane in the sample $z_{\text{f}}^{\text{sam}}$ varies with the incident angle θ^{air}, leading to an elongated focus at z_{f}. SMT corrects the refractive index mismatches by converting the input/output momenta from those in the air to those in the sample medium and shifting the reference plane of the reflection matrix to the sample surface at $z = 0$, as in Eqs. (S13), (S18).

written as superpositions of plane waves

$$E_{\text{in}}(\mathbf{r}, t) = \sum_{\mathbf{k}_\parallel^{\text{in}}, \omega} \tilde{E}_{\text{in}}(\mathbf{k}_\parallel^{\text{in}}, \omega) u_{\text{in}}(\mathbf{r}; \mathbf{k}_\parallel^{\text{in}}, \omega) e^{-i\omega t}, \tag{S3}$$

$$E_{\text{out}}(\mathbf{r}, t) = \sum_{\mathbf{k}_\parallel^{\text{out}}, \omega} \tilde{E}_{\text{out}}(\mathbf{k}_\parallel^{\text{out}}, \omega) u_{\text{out}}(\mathbf{r}; \mathbf{k}_\parallel^{\text{out}}, \omega) e^{-i\omega t}, \tag{S4}$$

with the reflection matrix R' relating the amplitudes

$$\tilde{E}_{\text{out}}(\mathbf{k}_\parallel^{\text{out}}, \omega) = \sum_{\mathbf{k}_\parallel^{\text{in}}} R'(\mathbf{k}_\parallel^{\text{out}}, \mathbf{k}_\parallel^{\text{in}}, \omega) \tilde{E}_{\text{in}}(\mathbf{k}_\parallel^{\text{in}}, \omega). \tag{S5}$$

The $R'(\mathbf{k}_\parallel^{\text{out}}, \mathbf{k}_\parallel^{\text{in}}, \omega)$ here has its input reference plane and output reference plane both at the z position of the calibration mirror, as described in Supplementary Sect. III B above. The measurements here use a dry objective lens, so these reference planes are with respect to the would-be propagation of the incident and reflected waves in air. We denote such "in-air reference plane" by $z_{\text{ref}}^{\text{air}} = z_{\text{mirror}}$ and the in-air reference point by $\mathbf{r}_{\text{ref}}^{\text{air}} = (0, 0, z_{\text{ref}}^{\text{air}})$. The plane wave profiles $u_{\text{in}}(\mathbf{r}; \mathbf{k}_\parallel^{\text{in}}, \omega)$ and $u_{\text{out}}(\mathbf{r}; \mathbf{k}_\parallel^{\text{out}}, \omega)$ include refraction at interfaces, and their relative phases are chosen such that their in-air reference point is at $\mathbf{r}_{\text{ref}}^{\text{air}}$.

When there is no interface and when $z_{\text{mirror}} = 0$, we have $\mathbf{r}_{\text{ref}}^{\text{air}} = \mathbf{0}$, $u_{\text{in}}(\mathbf{r}; \mathbf{k}_\parallel^{\text{in}}, \omega) = e^{i\mathbf{k}_{\text{in}} \cdot \mathbf{r}}$, $u_{\text{out}}(\mathbf{r}; \mathbf{k}_\parallel^{\text{out}}, \omega) = e^{i\mathbf{k}_{\text{out}} \cdot \mathbf{r}}$, with $\mathbf{k}_{\text{in}} = (\mathbf{k}_\parallel^{\text{in}}, \sqrt{(\omega/c)^2 - |\mathbf{k}_\parallel^{\text{in}}|^2})$ and $\mathbf{k}_{\text{out}} = (\mathbf{k}_\parallel^{\text{out}}, -\sqrt{(\omega/c)^2 - |\mathbf{k}_\parallel^{\text{out}}|^2})$, and we directly obtain Eqs. (1–2) of the main text with $R = R'$.

When there are interfaces, we need more careful definitions for $\mathbf{k}_{\text{in}}$, $\mathbf{k}_{\text{out}}$, and R, as we describe below.

1. Single-interface index mismatch

Consider a single air-sample interface at $z = 0$ (**Supplementary Fig. 8**a) with refractive indices $n_{\mathrm{air}} = 1$ and n_{sam}. Here,

$$u_{\mathrm{in}}(\mathbf{r}; \mathbf{k}_{\parallel}^{\mathrm{in}}, \omega) = \begin{cases} e^{i\mathbf{k}_{\mathrm{in}}^{\mathrm{air}} \cdot (\mathbf{r} - \mathbf{r}_{\mathrm{ref}}^{\mathrm{air}})}, & z \leq 0 \\ e^{i(\mathbf{k}_{\mathrm{in}}^{\mathrm{sam}} \cdot \mathbf{r} - k_{z,\mathrm{in}}^{\mathrm{air}} z_{\mathrm{ref}}^{\mathrm{air}})}, & z > 0 \end{cases} \tag{S6}$$

with $k_{z,\mathrm{in}}^{\mathrm{air}} = \sqrt{(\omega/c)^2 - |\mathbf{k}_{\parallel}^{\mathrm{in}}|^2}$, $k_{z,\mathrm{in}}^{\mathrm{sam}} = \sqrt{(n_{\mathrm{sam}}\omega/c)^2 - |\mathbf{k}_{\parallel}^{\mathrm{in}}|^2}$, $\mathbf{k}_{\mathrm{in}}^{\mathrm{air}} = (\mathbf{k}_{\parallel}^{\mathrm{in}}, k_{z,\mathrm{in}}^{\mathrm{air}})$, and $\mathbf{k}_{\mathrm{in}}^{\mathrm{sam}} = (\mathbf{k}_{\parallel}^{\mathrm{in}}, k_{z,\mathrm{in}}^{\mathrm{sam}})$. Eq. (S6) ensures that the in-air reference point is placed at $\mathbf{r} = \mathbf{r}_{\mathrm{ref}}^{\mathrm{air}}$ and that $u_{\mathrm{in}}(\mathbf{r}; \mathbf{k}_{\parallel}, \omega)$ is continuous across the air-sample interface at $z = 0$. Since we are not interested in imaging the interface, we ignore the interface reflection at $z = 0$ and the resulting amplitude change (*i.e.*, Fresnel coefficients) in u_{in}. Similarly,

$$u_{\mathrm{out}}(\mathbf{r}; \mathbf{k}_{\parallel}^{\mathrm{out}}, \omega) = \begin{cases} e^{i\mathbf{k}_{\mathrm{out}}^{\mathrm{air}} \cdot (\mathbf{r} - \mathbf{r}_{\mathrm{ref}}^{\mathrm{air}})}, & z \leq 0 \\ e^{i(\mathbf{k}_{\mathrm{out}}^{\mathrm{sam}} \cdot \mathbf{r} - k_{z,\mathrm{out}}^{\mathrm{air}} z_{\mathrm{ref}}^{\mathrm{air}})}, & z > 0 \end{cases} \tag{S7}$$

with $k_{z,\mathrm{out}}^{\mathrm{air}} = -\sqrt{(\omega/c)^2 - |\mathbf{k}_{\parallel}^{\mathrm{out}}|^2}$, $k_{z,\mathrm{out}}^{\mathrm{sam}} = -\sqrt{(n_{\mathrm{sam}}\omega/c)^2 - |\mathbf{k}_{\parallel}^{\mathrm{out}}|^2}$, $\mathbf{k}_{\mathrm{out}}^{\mathrm{air}} = (\mathbf{k}_{\parallel}^{\mathrm{out}}, k_{z,\mathrm{out}}^{\mathrm{air}})$, and $\mathbf{k}_{\mathrm{out}}^{\mathrm{sam}} = (\mathbf{k}_{\parallel}^{\mathrm{out}}, k_{z,\mathrm{out}}^{\mathrm{sam}})$.

A dry objective lens that focuses to point $\mathbf{r}_{\mathrm{in}}^{\mathrm{air}} = (x_{\mathrm{f}}, y_{\mathrm{f}}, z_{\mathrm{f}}^{\mathrm{air}})$ in air produces an input wavefront

$$\tilde{E}_{\mathrm{in}}^{\mathrm{air}}(\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega) = e^{-i\mathbf{k}_{\mathrm{in}}^{\mathrm{air}} \cdot (\mathbf{r}_{\mathrm{in}}^{\mathrm{air}} - \mathbf{r}_{\mathrm{ref}}^{\mathrm{air}})}, \tag{S8}$$

such that the incident wave is

$$E_{\mathrm{in}}^{\mathrm{air}}(\mathbf{r}, t) = \sum_{\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega} \tilde{E}_{\mathrm{in}}^{\mathrm{air}}(\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega) u_{\mathrm{in}}(\mathbf{r}; \mathbf{k}_{\parallel}^{\mathrm{in}}, \omega) e^{-i\omega t} = \begin{cases} \sum_{\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega} e^{i\mathbf{k}_{\mathrm{in}}^{\mathrm{air}} \cdot (\mathbf{r} - \mathbf{r}_{\mathrm{in}}^{\mathrm{air}}) - i\omega t}, & z \leq 0 \\ \sum_{\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega} e^{i\mathbf{k}_{\mathrm{in}}^{\mathrm{sam}} \cdot \mathbf{r} - i\mathbf{k}_{\mathrm{in}}^{\mathrm{air}} \cdot \mathbf{r}_{\mathrm{in}}^{\mathrm{air}} - i\omega t}. & z > 0 \end{cases} \tag{S9}$$

When $z_{\mathrm{f}}^{\mathrm{air}} \leq 0$, $E_{\mathrm{in}}^{\mathrm{air}}(\mathbf{r}, t)$ forms a perfect spatiotemporal focus at $\mathbf{r} = \mathbf{r}_{\mathrm{in}}^{\mathrm{air}}$, $t = 0$, since all terms in the double summation $\sum_{\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega}$ add up in phase. But when the desired focus is in the medium (with $z_{\mathrm{f}}^{\mathrm{air}} > 0$), such a dry objective lens no longer focuses well. At $\mathbf{r} = \mathbf{r}_{\mathrm{in}}^{\mathrm{air}}$, $t = 0$, the different terms in the double summation have relative phase $e^{i(k_{z,\mathrm{in}}^{\mathrm{sam}} - k_{z,\mathrm{in}}^{\mathrm{air}}) z_{\mathrm{f}}^{\mathrm{air}}}$, no longer in phase. As illustrated geometrically in the top panel of **Supplementary Fig. 8**a, the spatial focus is now shifted to $\mathbf{r}_{\mathrm{in}} = (x_{\mathrm{f}}, y_{\mathrm{f}}, z_{\mathrm{f}})$, where the nearby angular components (but not all angular components) add up in phase. For small NA (where $|\mathbf{k}_{\parallel}^{\mathrm{in}}| \ll \omega/c$), $z_{\mathrm{f}} \approx (n_{\mathrm{sam}}/n_{\mathrm{air}}) z_{\mathrm{f}}^{\mathrm{air}}$, and the pulse arrives at $\mathbf{r}_{\mathrm{in}}$ at time $t = t_{\mathrm{f}} = (n_{\mathrm{sam}} z_{\mathrm{f}} - n_{\mathrm{air}} z_{\mathrm{f}}^{\mathrm{air}})/c$, which compensates for the slower propagation in the medium and the longer distance it needs to propagate (to z_{f} instead of $z_{\mathrm{f}}^{\mathrm{air}}$). So, $E_{\mathrm{in}}^{\mathrm{air}}(\mathbf{r}, t)$ forms an imperfect spatiotemporal focus at $\mathbf{r} = \mathbf{r}_{\mathrm{in}}$, $t = t_{\mathrm{f}}$. Note that to align the temporal gate with the imperfect spatial gate, the necessary pulse arrival time t_{f} depends on the depth z_{f} of the focus.

To form an ideal spatiotemporal focus at $\mathbf{r}_{\mathrm{in}} = (x_{\mathrm{f}}, y_{\mathrm{f}}, z_{\mathrm{f}})$ inside the medium, we digitally adopt a modified input wavefront

$$\tilde{E}_{\mathrm{in}}(\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega) = e^{-i(\mathbf{k}_{\mathrm{in}}^{\mathrm{sam}} \cdot \mathbf{r}_{\mathrm{in}} - k_{z,\mathrm{in}}^{\mathrm{air}} z_{\mathrm{ref}}^{\mathrm{air}})}. \tag{S10}$$

When focusing inside the medium ($z_{\mathrm{f}} > 0$), the resulting incident wave

$$E_{\mathrm{in}}(\mathbf{r}, t) = \sum_{\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega} \tilde{E}_{\mathrm{in}}(\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega) u_{\mathrm{in}}(\mathbf{r}; \mathbf{k}_{\parallel}^{\mathrm{in}}, \omega) e^{-i\omega t} = \sum_{\mathbf{k}_{\parallel}^{\mathrm{in}}, \omega} e^{i\mathbf{k}_{\mathrm{in}}^{\mathrm{sam}} \cdot (\mathbf{r} - \mathbf{r}_{\mathrm{in}}) - i\omega t} \tag{S11}$$

has all of its angular and spectral components add up perfectly in phase at $\mathbf{r} = \mathbf{r}_{\mathrm{in}}$, $t = t_{\mathrm{f}} = 0$. The pulse arrival time $t_{\mathrm{f}} = 0$ here is independent of the focal depth z_{f}.

Combining Eqs. (S4, S5, S7, S10), the outgoing field in the sample region (when $z > 0$) is

$$E_{\mathrm{out}}(\mathbf{r}, t; \mathbf{r}_{\mathrm{in}}) = \sum_{\mathbf{k}_{\parallel}^{\mathrm{out}}, \mathbf{k}_{\parallel}^{\mathrm{in}}, \omega} R'(\mathbf{k}_{\parallel}^{\mathrm{out}}, \mathbf{k}_{\parallel}^{\mathrm{in}}, \omega) e^{i[\mathbf{k}_{\mathrm{out}}^{\mathrm{sam}} \cdot \mathbf{r} - \mathbf{k}_{\mathrm{in}}^{\mathrm{sam}} \cdot \mathbf{r}_{\mathrm{in}} - (k_{z,\mathrm{out}}^{\mathrm{air}} - k_{z,\mathrm{in}}^{\mathrm{air}}) z_{\mathrm{ref}}^{\mathrm{air}} - i\omega t]}. \tag{S12}$$

We align the input spatial gate with the output spatial gate by setting $\mathbf{r}_{\text{in}} = \mathbf{r}$, and then align the temporal gate with the spatial gates by evaluating E_{out} at time $t = t_{\text{f}} = 0$. The resulting index-mismatch-corrected SMT image is

$$\psi_{\text{SMT}}(\mathbf{r}) = E_{\text{out}}(\mathbf{r}, t = 0; \mathbf{r}_{\text{in}} = \mathbf{r}) = \sum_{\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega} e^{i(\mathbf{k}_{\text{out}}^{\text{sam}} - \mathbf{k}_{\text{in}}^{\text{sam}})\cdot\mathbf{r}} R(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega), \tag{S13a}$$

$$R(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega) \equiv e^{-i(k_{z,\text{out}}^{\text{air}} - k_{z,\text{in}}^{\text{air}}) z_{\text{ref}}^{\text{air}}} R'(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega). \tag{S13b}$$

This recovers Eq. (1) of the main text. Eq. (S13b) shifts the reference plane of the reflection matrix: the original R' has its in-air reference plane at $z = z_{\text{ref}}^{\text{air}} = z_{\text{mirror}}$, and the new R has its in-medium reference plane at the air-sample interface at $z = 0$. We also see explicitly that SMT uses momenta $\mathbf{k}_{\text{in/out}}^{\text{sam}}$ inside the sample to focus inside the sample medium.

2. *Double-interface index mismatch*

For a sample under a coverslip with thickness h^{glass}, there are two interfaces: an air-glass interface at $z = -h^{\text{glass}}$ and a glass-sample interface at $z = 0$ (**Supplementary Fig. 8**b). We choose the zero of the coordinate such that z is the depth into the sample. Here,

$$u_{\text{in}}(\mathbf{r}; \mathbf{k}_{\parallel}^{\text{in}}, \omega) = \begin{cases} e^{i\mathbf{k}_{\text{in}}^{\text{air}}\cdot(\mathbf{r} - \mathbf{r}_{\text{ref}}^{\text{air}})}, & z \le -h^{\text{glass}} \\ e^{i[\mathbf{k}_{\text{in}}^{\text{glass}}\cdot\mathbf{r} - k_{z,\text{in}}^{\text{air}}(h^{\text{glass}} + z_{\text{ref}}^{\text{air}}) + k_{z,\text{in}}^{\text{glass}} h^{\text{glass}}]}, & -h^{\text{glass}} < z \le 0 \\ e^{i[\mathbf{k}_{\text{in}}^{\text{sam}}\cdot\mathbf{r} - k_{z,\text{in}}^{\text{air}}(h^{\text{glass}} + z_{\text{ref}}^{\text{air}}) + k_{z,\text{in}}^{\text{glass}} h^{\text{glass}}]}, & z > 0 \end{cases} \tag{S14}$$

and

$$u_{\text{out}}(\mathbf{r}; \mathbf{k}_{\parallel}^{\text{out}}, \omega) = \begin{cases} e^{i\mathbf{k}_{\text{out}}^{\text{air}}\cdot(\mathbf{r} - \mathbf{r}_{\text{ref}}^{\text{air}})}, & z \le -h^{\text{glass}} \\ e^{i[\mathbf{k}_{\text{out}}^{\text{glass}}\cdot\mathbf{r} - k_{z,\text{out}}^{\text{air}}(h^{\text{glass}} + z_{\text{ref}}^{\text{air}}) + k_{z,\text{out}}^{\text{glass}} h^{\text{glass}}]}, & -h^{\text{glass}} < z \le 0 \\ e^{i[\mathbf{k}_{\text{out}}^{\text{sam}}\cdot\mathbf{r} - k_{z,\text{out}}^{\text{air}}(h^{\text{glass}} + z_{\text{ref}}^{\text{air}}) + k_{z,\text{out}}^{\text{glass}} h^{\text{glass}}]}, & z > 0 \end{cases} \tag{S15}$$

which ensure that the in-air reference point is at $\mathbf{r} = \mathbf{r}_{\text{ref}}^{\text{air}}$ and continuity across the interfaces. The definitions are the same as in the previous section, and we additionally introduce the momentum in the coverslip as $k_{z,\text{in}}^{\text{glass}} = \sqrt{(n_{\text{glass}}\omega/c)^2 - |\mathbf{k}_{\parallel}^{\text{in}}|^2}$, $\mathbf{k}_{\text{in}}^{\text{glass}} = (\mathbf{k}_{\parallel}^{\text{in}}, k_{z,\text{in}}^{\text{glass}})$, $k_{z,\text{out}}^{\text{glass}} = -\sqrt{(n_{\text{glass}}\omega/c)^2 - |\mathbf{k}_{\parallel}^{\text{out}}|^2}$, and $\mathbf{k}_{\text{out}}^{\text{glass}} = (\mathbf{k}_{\parallel}^{\text{out}}, k_{z,\text{out}}^{\text{glass}})$.

When computing k_z^{glass} and k_z^{sam}, a rough estimate of n_{glass} and n_{sam} will suffice since we also perform dispersion compensation and wavefront optimizations later. Here, we take $n_{\text{glass}}(\omega)$ to be that of SCHOTT D263T Thin Glass [82] and we take $n_{\text{sam}}(\omega)$ to be that of water [83] and polydimethylsiloxane (PDMS) [84] respectively for the USAF-target-under-tissue sample and the nanoparticle colloid sample, with data from refractiveindex.info [85]. If the material is unknown, one can carry out an image metric optimization (similar to the dispersion compensation in Supplementary Sect. III E below) to estimate its average refractive index.

To create an ideal spatiotemporal focus at $\mathbf{r}_{\text{in}} = (x_{\text{f}}, y_{\text{f}}, z_{\text{f}})$ inside the sample (with $z_{\text{f}} > 0$), we digitally generate an input wavefront

$$\tilde{E}_{\text{in}}(\mathbf{k}_{\parallel}^{\text{in}}, \omega) = e^{-i[\mathbf{k}_{\text{in}}^{\text{sam}}\cdot\mathbf{r}_{\text{in}} - k_{z,\text{in}}^{\text{air}}(h^{\text{glass}} + z_{\text{ref}}^{\text{air}}) + k_{z,\text{in}}^{\text{glass}} h^{\text{glass}}]}. \tag{S16}$$

Like Eq. (S11), the resulting incident field $E_{\text{in}}(\mathbf{r}, t)$ forms an ideal spatiotemporal focus at $\mathbf{r} = \mathbf{r}_{\text{in}}$, $t = 0$. The corresponding output field inside the sample ($z > 0$) is

$$E_{\text{out}}(\mathbf{r}, t; \mathbf{r}_{\text{in}}) = \sum_{\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega} R'(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega) e^{i[\mathbf{k}_{\text{out}}^{\text{sam}}\mathbf{r} - \mathbf{k}_{\text{in}}^{\text{sam}}\cdot\mathbf{r}_{\text{in}} - (k_{z,\text{out}}^{\text{air}} - k_{z,\text{in}}^{\text{air}})(h^{\text{glass}} + z_{\text{ref}}^{\text{air}}) + (k_{z,\text{out}}^{\text{glass}} - k_{z,\text{in}}^{\text{glass}}) h^{\text{glass}} - i\omega t]}. \tag{S17}$$

As before, we align the three gates to obtain

$$\psi_{\text{SMT}}(\mathbf{r}) = E_{\text{out}}(\mathbf{r}, t = 0; \mathbf{r}_{\text{in}} = \mathbf{r}) = \sum_{\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega} e^{i(\mathbf{k}_{\text{out}}^{\text{sam}} - \mathbf{k}_{\text{in}}^{\text{sam}})\cdot\mathbf{r}} R(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega), \tag{S18a}$$

$$R(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega) \equiv e^{-i(k_{z,\text{out}}^{\text{air}} - k_{z,\text{in}}^{\text{air}})(h^{\text{glass}} + z_{\text{ref}}^{\text{air}})} e^{i(k_{z,\text{out}}^{\text{glass}} - k_{z,\text{in}}^{\text{glass}}) h^{\text{glass}}} R'(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega), \tag{S18b}$$

which again recovers Eq. (1) of the main text. The original R' has an in-air reference plane at $z = z_{\text{ref}}^{\text{air}} = z_{\text{mirror}}$. The new R has its in-medium reference plane at the glass-sample interface at $z = 0$ [by shifting the reference plane to the air-glass interface with the first factor of Eq. (S18b) and then to the glass-sample interface with the second factor of Eq. (S18b)].

D. Layer-thickness measurement and depth calibration

Accurately correcting the refractive index mismatches requires accurate knowledge of the coverslip thickness $h^{\rm glass}$ and of the reference plane location $z^{\rm air}_{\rm ref} = z_{\rm mirror}$. They are also needed for us to determine the precise imaging depth.

To determine $h^{\rm glass}$ and $z^{\rm air}_{\rm ref}$, we first place the air-glass interface at the objective lens focal plane, which makes the iris in the incidence path clearly imaged on the camera (Supplementary Sect. I A). We then use the motorized translation stage to shift the sample backward along the axial z direction by 70 μm (so that $z^{\rm air}_{\rm ref}$ is known) and then perform SMT to image the glass-sample interface, using the single-interface index mismatch correction of Sect. III C 1. The SMT image gives $h^{\rm glass} = 141.5$ μm for the USAF-target-under-tissue sample and $h^{\rm glass} = 154$ μm for the nanoparticle colloid sample. This procedure can be repeated if multiple layers are present.

Then, we move the sample forward to place the air-glass interface at the objective lens focal plane again, shift the sample backward in z by $h^{\rm glass} + z^{\rm air}_{\rm ref}$, and perform SMT imaging there. Given the depth range of interest, we use geometric optics to estimate the $z^{\rm air}_{\rm ref}$ to set: $(h^{\rm glass} + z^{\rm air}_{\rm ref}) \tan\theta^{\rm air} = h^{\rm glass} \tan\theta^{\rm glass} + z^{\rm sam}_{\rm f}(\theta^{\rm air}) \tan\theta^{\rm sam}$. Here, $z^{\rm sam}_{\rm f}(\theta^{\rm air})$ is the depth of the focal plane in the sample (illustrated in **Supplementary Fig. 8**b), which depends on the incident angle. $\theta^{\rm air}$, $\theta^{\rm glass}$, and $\theta^{\rm sam}$ are incident angles in air, glass, and the sample, respectively, related through Snell's law: $\sin\theta^{\rm air} = n_{\rm glass}\sin\theta^{\rm glass} = n_{\rm sam}\sin\theta^{\rm sam}$. For the data in this paper, we set $z^{\rm air}_{\rm ref} = 678.5$ μm for the USAF target under brain tissue, resulting in the single-angle focal-plane depth ranging from $z^{\rm sam}_{\rm f}(0^\circ) = 965$ μm to $z^{\rm sam}_{\rm f}(30^\circ) = 1044$ μm (with $n_{\rm glass} = 1.5$ and $n_{\rm sam} = 1.33$) for the angular range (NA = 0.5) being considered. For the nanoparticle colloid sample, we use $z^{\rm air}_{\rm ref} = 996$ μm, leading to a $z^{\rm sam}_{\rm f}(0^\circ) = 1466$ μm and $z^{\rm sam}_{\rm f}(30^\circ) = 1594$ μm (with $n_{\rm sam} = 1.4$). Note that the depth of field of the SMT image can go beyond the range of $z^{\rm sam}_{\rm f}$, as described in Supplementary Sect. VIII B.

In SMT, the absolute image depth z is quantitatively determined by the other variables thanks to the index-mismatch correction. Given the frequency ω measured by the wavemeter, $\mathbf{k}^{\rm out}_{\parallel}$ from off-axis holography (Supplementary Sect. I C), $\mathbf{k}^{\rm in}_{\parallel}$ from incident angle calibration (Supplementary Sect. III A), $h^{\rm glass}$ measured through imaging the glass-sample interface, $z^{\rm air}_{\rm ref}$ set by the motorized translation stage, and the estimated refractive indices, we can reconstruct the image $\psi_{\rm SMT}(x, y, z)$ with a quantitative depth coordinate z.

E. Compensation for sample dispersion

Next, we correct the dispersion of the sample and the output path of the optical system by introducing a frequency-dependent phase shift $\theta(\omega)$ to the scattering matrix. We assume the dispersion is spatially invariant and only depends on the frequency. Since the angles do not affect dispersion, we pre-compute the angular summations as

$$\tilde{\psi}_{\rm SMT}(\mathbf{r}, \omega) \equiv \sum_{\mathbf{k}^{\rm out}_{\parallel}, \mathbf{k}^{\rm in}_{\parallel}} e^{i(\mathbf{k}_{\rm out} - \mathbf{k}_{\rm in})\cdot\mathbf{r}} R(\mathbf{k}^{\rm out}_{\parallel}, \mathbf{k}^{\rm in}_{\parallel}, \omega) \tag{S19}$$

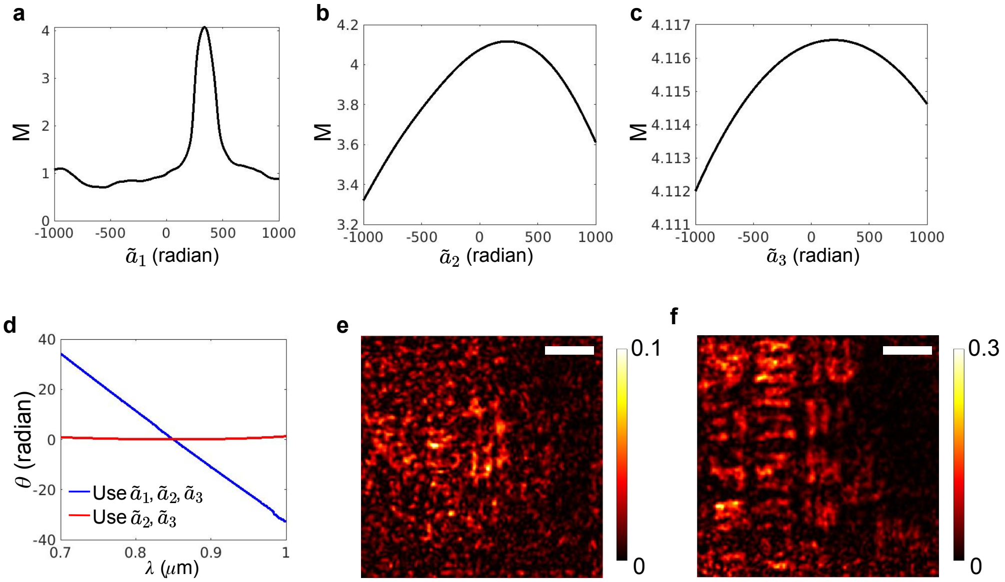
such that $\psi_{\rm SMT}(\mathbf{r}) = \sum_\omega e^{i\theta(\omega)}\tilde{\psi}_{\rm SMT}(\mathbf{r}, \omega)$ and $I_{\rm SMT}(\mathbf{r}) = |\psi_{\rm SMT}(\mathbf{r})|^2$ before we perform wavefront corrections [when $\phi_{\rm out}(\mathbf{k}^{\rm out}_{\parallel}) = \phi_{\rm in}(\mathbf{k}^{\rm in}_{\parallel}) = 0$], following Eq. (2) of the main text. Hereafter, we abbreviate $\mathbf{k}^{\rm sam}_{\rm out}$ and $\mathbf{k}^{\rm sam}_{\rm in}$ as $\mathbf{k}_{\rm out}$ and $\mathbf{k}_{\rm in}$ for conciseness. To determine $\theta(\omega)$, we maximize an image quality metric (also referred to as the "objective function" or "figure of merit" of the optimization)

$$M \equiv \sum_{\mathbf{r}} I_{\rm SMT}(\mathbf{r}) \ln \frac{I_{\rm SMT}(\mathbf{r})}{I_0}. \tag{S20}$$

The precise choice of the image quality metric is not crucial. For example, we have found similar performances with $M' \equiv \sum_{\mathbf{r}} I_{\rm SMT}(\mathbf{r})$ and $M'' \equiv \sum_{\mathbf{r}} I^2_{\rm SMT}(\mathbf{r})$.

For the USAF-target-under-tissue sample, we first scan z and record M for the 2D *en-face* image at each depth. We found $z_1 = 994.5$ μm is the depth where M is maximized before dispersion compensation. We then perform dispersion compensation with M summed over positions $\mathbf{r}$ in a 3D volume containing z_1: $x, y \in [0, 50]$ μm and $z \in [970, 1030]$ μm. The axial range allows the dispersion compensation to sharpen the axial resolution through the temporal gate. For the sample of dense nanoparticle colloid, we sum over positions $\mathbf{r}$ in a volume spanning $x, y \in [0, 50]$ μm and $z \in [1470, 1580]$ μm. A relatively coarse $0.7 \times 0.7 \times 1$ μm^3 voxel size is used to speed up the summation. We choose the normalization constant I_0 to be 10^{-3} times the maximal $I_{\rm SMT}$ before optimization. Our experience is that the value of I_0 does not change the result as long as it is not too large.

While it may be tempting to allow the phase $\theta(\omega)$ at every frequency to be an independent variable in the optimization, doing so will lead to overfitting and make it easier for the optimization to get trapped in a poor local optimum given the very high dimensionality of the parameter space (see Fig. 2b of the main text). To regularize the



Supplementary Fig. 9. **Dispersion compensation for USAF target under tissue.** **a–c** Normalized SMT image quality metric $M = \sum_{\mathbf{r}} I_{\mathrm{SMT}}(\mathbf{r}) \ln I_{\mathrm{SMT}}(\mathbf{r})$ as a function of the dispersion coefficients $\tilde{a}_j$. **d** The optimized $\theta(\omega)$. **e** The USAF target image at $z = z_1 = 994.5$ µm before dispersion compensation. **f** The USAF target image at $z = z_{\mathrm{target}} = 977.5$ µm after dispersion compensation. Scale bars: 10 µm. Note that prior to this sample dispersion compensation, we have already removed the dispersion in the input path of the optical system through a calibration mirror (Supplementary Sect. III B) and removed the dispersion from refractive index mismatch (Supplementary Sect. III C).

optimization while reducing the dimensionality, we express $\theta(\omega)$ as a third-order polynomial [43],

$$\theta(\omega) = \tilde{a}_1 \frac{\omega - \omega_0}{\omega_0} + \tilde{a}_2 \left(\frac{\omega - \omega_0}{\omega_0} \right)^2 + \tilde{a}_3 \left(\frac{\omega - \omega_0}{\omega_0} \right)^3, \tag{S21}$$

where ω_0 is the central frequency, and $\tilde{a}_j$ $(j = 1, 2, 3)$ are dimensionless dispersion coefficients corresponding to the inverse group velocity (correcting for axial shifts of the temporal gate), group velocity dispersion (correcting for symmetric pulse broadening), and third-order dispersion (correcting for asymmetric pulse distortion), respectively.

To maximize the image metric M while avoiding poor local optima, we optimize one variable at a time. We first align the temporal gate with the spatial gates by scanning $\tilde{a}_1$ to obtain $\tilde{a}_1^{\max}$ while keeping $\tilde{a}_2 = \tilde{a}_3 = 0$. Then, we perform symmetric pulse compression by scanning $\tilde{a}_2$ to obtain $\tilde{a}_2^{\max}$ while keeping $\tilde{a}_1 = \tilde{a}_1^{\max}$ and $\tilde{a}_3 = 0$. Finally, we perform asymmetric pulse compression by scanning $\tilde{a}_3$ to obtain $\tilde{a}_3^{\max}$ while keeping $\tilde{a}_1 = \tilde{a}_1^{\max}$, $\tilde{a}_2 = \tilde{a}_2^{\max}$. This process is shown in **Supplementary Fig. 9** for the USAF-target-under-tissue sample. The scanning is the most important for $\tilde{a}_1$ as its objective function landscape is the most nonconvex.

To further push the performance, we additionally perform a local gradient-based optimization to optimize all three variables simultaneously, using $(\tilde{a}_1^{\max}, \tilde{a}_2^{\max}, \tilde{a}_3^{\max})$ as the initial guess. The gradient of M with respect to $\tilde{a}_j$ is

$$\frac{\partial M}{\partial \tilde{a}_j} = -2 \sum_{\mathbf{r},\omega} \mathrm{Im} \left\{ \left[1 + \ln \frac{I_{\mathrm{SMT}}(\mathbf{r})}{I_0} \right] \psi_{\mathrm{SMT}}^*(\mathbf{r}) \tilde{\psi}_{\mathrm{SMT}}(\mathbf{r}, \omega) e^{i\theta(\omega)} \right\} \left(\frac{\omega - \omega_0}{\omega_0} \right)^j. \tag{S22}$$

To compute this gradient efficiently in MATLAB, we use vector-matrix multiplications to perform the spatial and frequency summations. Specifically, $I_{\mathrm{SMT}}(\mathbf{r}_i)$, $\psi_{\mathrm{SMT}}^*(\mathbf{r}_i)$, and $\theta(\omega_k)$ are each stored in a row vector, $\boldsymbol{I}$, $\boldsymbol{\psi}^*$, and $\boldsymbol{\theta}$. $\tilde{\psi}_{\mathrm{SMT}}(\mathbf{r}_i, \omega_k)$ is stored in the ik element of a matrix $\tilde{\boldsymbol{\psi}}$, and $(\frac{\omega_k - \omega_0}{\omega_0})^j$ in the kj element of matrix $\boldsymbol{\Omega}$. We then compute the gradient as a row vector by

$$\frac{\partial M}{\partial \tilde{\boldsymbol{a}}} = -2\mathrm{Im} \left\{ \left[\left(\left(1 + \ln \frac{\boldsymbol{I}}{I_0} \right) \otimes \boldsymbol{\psi}^* \right) \tilde{\boldsymbol{\psi}} \right] \otimes e^{i\boldsymbol{\theta}} \right\} \boldsymbol{\Omega}, \tag{S23}$$

where $\otimes$ denotes the Hadamard (element-wise) product between row vectors, and the other products are vector-matrix products. Note that the matrices $\tilde{\boldsymbol{\psi}}$ and $\boldsymbol{\Omega}$ are both pre-computed and fixed; only the vectors $\boldsymbol{I}$, $\boldsymbol{\psi}^*$, and $\boldsymbol{\theta}$ are updated during the optimization. We use the NLopt library [49] for the optimization, adopting the low-storage

Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm [48]. L-BFGS is a second-order quasi-Newton method that uses the gradient and an approximation of the Hessian matrix of the objective function to determine the search direction and estimates a step size. We set the upper and lower bounds of $\tilde{\boldsymbol{a}}$ to be $(\tilde{a}_1^{\max} \pm 100, \tilde{a}_2^{\max} \pm 100, \tilde{a}_3^{\max} \pm 100)$ in the optimization, with the stopping criterion being when both the objective function M and the variables $\tilde{\boldsymbol{a}}$ change less than 0.01%.

Supplementary Fig. 9e–f and **Supplementary Fig. 15**a–b show the images before and after dispersion compensation for the USAF-target-under-tissue and the nanoparticle colloid samples respectively. We then determine the depth $z_{\text{target}} = 977.5$ µm of the USAF target as the depth z at which the *en-face* image quality metric M is the highest.

F. Input and output wavefront corrections

After dispersion compensation, we optimize angle-dependent phases $\phi_{\text{in}}(\mathbf{k}_\parallel^{\text{in}})$ and $\phi_{\text{out}}(\mathbf{k}_\parallel^{\text{out}})$ to correct for the output aberrations of the optical system and for the aberrations and multiple scatterings from the heterogeneous sample.

1. *Zernike-regularized image metric optimization*

We maximize the same image quality metric M as in Eq. (S20), with the spatial summation summing over one or several *en-face* x–y planes at depths z_j $(j = 1, 2, \ldots)$ inside the volume of interest. For the USAF-target-under-tissue sample, we use one depth $z_1 = z_{\text{target}}$ since the target is planar. For the nanoparticle colloid sample, we divide the full volume (across 110 µm depth of field in z) into 16 overlapping sub-volumes, and pick five equally spaced slices $z_1, z_2, \ldots, z_5$ within each sub-volume where z_1 and z_5 are the two ends of the sub-volume; the same wavefront correction is then applied to all slices within each sub-volume. We treat the wavefront corrections $\phi_{\text{in}}(\mathbf{k}_\parallel^{\text{in}})$ and $\phi_{\text{out}}(\mathbf{k}_\parallel^{\text{out}})$ to be frequency independent.

Since the spectral phase $\theta(\omega)$ has been determined and since $z = z_j$ is fixed, we pre-compute the frequency summation to obtain the depth-resolved "synthetic time-gated reflection matrix" at each depth z_j,

$$\tilde{R}(\mathbf{k}_\parallel^{\text{out}}, \mathbf{k}_\parallel^{\text{in}}; z_j) \equiv \sum_{\omega} e^{i\theta(\omega)} e^{i(k_z^{\text{out}} - k_z^{\text{in}}) z_j} R(\mathbf{k}_\parallel^{\text{out}}, \mathbf{k}_\parallel^{\text{in}}, \omega). \tag{S24}$$

Recall that $\mathbf{k}_{\text{in}} = (\mathbf{k}_\parallel^{\text{in}}, k_z^{\text{in}})$, $\mathbf{k}_{\text{out}} = (\mathbf{k}_\parallel^{\text{out}}, k_z^{\text{out}})$. Note $k_z^{\text{in/out}} = \pm\sqrt{|\mathbf{k}_\parallel|^2 - (n_{\text{sam}}\omega/c)^2}$ depends on the frequency ω. For the optimization over $\phi_{\text{in}}(\mathbf{k}_\parallel^{\text{in}})$, we further pre-compute the summation over the outgoing angles for each j, as

$$\psi_{\text{SMT}}^{j,\text{in}}(\mathbf{r}_\parallel, \mathbf{k}_\parallel^{\text{in}}; z_j) \equiv \sum_{\mathbf{k}_\parallel^{\text{out}}} e^{i(\mathbf{k}_\parallel^{\text{out}} - \mathbf{k}_\parallel^{\text{in}})\cdot \mathbf{r}_\parallel + i\phi_{\text{out}}(\mathbf{k}_\parallel^{\text{out}})} \tilde{R}(\mathbf{k}_\parallel^{\text{out}}, \mathbf{k}_\parallel^{\text{in}}; z_j), \tag{S25}$$

where $\mathbf{r}_\parallel = (x, y)$, such that $\psi_{\text{SMT}}(\mathbf{r}_\parallel, z_j) = \sum_{\mathbf{k}_\parallel^{\text{in}}} e^{i\phi_{\text{in}}(\mathbf{k}_\parallel^{\text{in}})} \psi_{\text{SMT}}^{j,\text{in}}(\mathbf{r}_\parallel, \mathbf{k}_\parallel^{\text{in}}; z_j)$ on the j-th slice. For the $\phi_{\text{out}}(\mathbf{k}_\parallel^{\text{out}})$ optimization, we pre-compute

$$\psi_{\text{SMT}}^{j,\text{out}}(\mathbf{r}_\parallel, \mathbf{k}_\parallel^{\text{out}}; z_j) \equiv \sum_{\mathbf{k}_\parallel^{\text{in}}} e^{i(\mathbf{k}_\parallel^{\text{out}} - \mathbf{k}_\parallel^{\text{in}})\cdot \mathbf{r}_\parallel + i\phi_{\text{in}}(\mathbf{k}_\parallel^{\text{in}})} \tilde{R}(\mathbf{k}_\parallel^{\text{out}}, \mathbf{k}_\parallel^{\text{in}}; z_j), \tag{S26}$$

such that $\psi_{\text{SMT}}(\mathbf{r}_\parallel, z_j) = \sum_{\mathbf{k}_\parallel^{\text{out}}} e^{i\phi_{\text{out}}(\mathbf{k}_\parallel^{\text{out}})} \psi_{\text{SMT}}^{j,\text{out}}(\mathbf{r}_\parallel, \mathbf{k}_\parallel^{\text{out}}; z_j)$. We evaluate $\mathbf{r}_\parallel$ on a relatively coarse $(0.7\ \mu\text{m})^2$ spatial grid to speed up the computation.

Like the spectral phase optimization, here it is better to avoid overfitting and not to allow the angular phase at every incident angle and every outgoing angle to be an independent variable. To regularize the optimization, we expand the phase-correction maps in Zernike polynomials [86]

$$\phi_{\text{in}}(\mathbf{k}_\parallel^{\text{in}}) = \sum_n c_n^{\text{in}} Z_n\left(\frac{\mathbf{k}_\parallel^{\text{in}}}{k_0^{\max}\text{NA}}\right), \quad \phi_{\text{out}}(\mathbf{k}_\parallel^{\text{out}}) = \sum_n c_n^{\text{out}} Z_n\left(\frac{\mathbf{k}_\parallel^{\text{out}}}{k_0^{\max}\text{NA}}\right), \tag{S27}$$

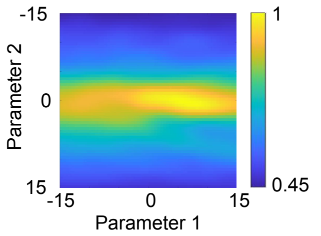
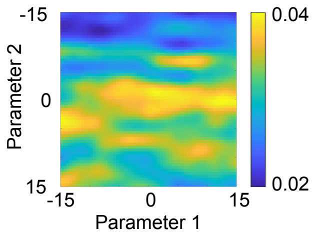
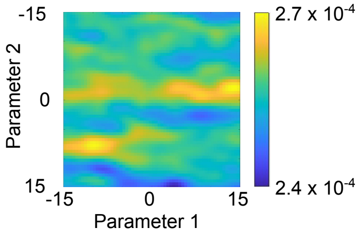
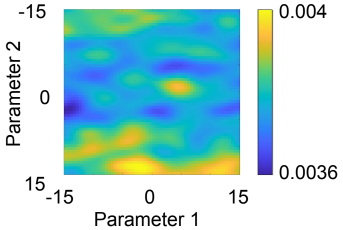
where $\{c_n^{\text{in}}\}$ and $\{c_n^{\text{out}}\}$ are the Zernike coefficients (weights), $k_0^{\max} = \omega_{\max}/c$ is the vacuum wave number for the highest frequency considered, and Z_n is the n-th Zernike polynomial. Note that the Zernike polynomials form a complete basis for functions on a disk, so they can incorporate both the slowly-varying distortions from aberrations and the fast-changing wavefronts caused by multiple scattering.

The gradient of the image metric M in Eq. (S20) with respect to $\{c_n^{\text{in}}\}$ is

$$\frac{\partial M}{\partial c_n^{\text{in}}} = -2 \sum_{j,\mathbf{r}_\parallel,\mathbf{k}_\parallel^{\text{in}}} \text{Im}\left\{\left[1 + \ln\frac{I_{\text{SMT}}(\mathbf{r}_\parallel, z_j)}{I_0}\right] \psi_{\text{SMT}}^{*}(\mathbf{r}_\parallel, z_j)\psi_{\text{SMT}}^{j,\text{in}}(\mathbf{r}_\parallel, \mathbf{k}_\parallel^{\text{in}}; z_j) e^{i\phi_{\text{in}}(\mathbf{k}_\parallel^{\text{in}})}\right\} Z_n\left(\frac{\mathbf{k}_\parallel^{\text{in}}}{k_0^{\text{max}}\text{NA}}\right), \tag{S28}$$

and similarly for the gradient with respect to $\{c_n^{\text{out}}\}$. Like Eqs. (S22)–(S23), these gradients can be evaluated efficiently using vector-matrix multiplications.

We first keep $c_n^{\text{out}} = 0$, pre-compute $\psi_{\text{SMT}}^{j,\text{in}}(\mathbf{r}_\parallel, \mathbf{k}_\parallel^{\text{in}}; z_j)$, and determine $\{c_n^{\text{in}}\}$ by maximizing the image quality metric with the L-BFGS algorithm and with $c_n^{\text{in}} = 0$ as the initial guess. We set the upper and lower bounds as ±10 radians, with the stopping criterion being when both the objective function M and the variables c_n^{in} change less than 0.01%. This optimization is very fast since the most time-consuming step of computing $\psi_{\text{SMT}}^{j,\text{in}}(\mathbf{r}_\parallel, \mathbf{k}_\parallel^{\text{in}}; z_j)$ has already been done prior. Then, we fix $\{c_n^{\text{in}}\}$, pre-compute $\psi_{\text{SMT}}^{j,\text{out}}(\mathbf{r}_\parallel, \mathbf{k}_\parallel^{\text{in}}; z_j)$, and optimize $\{c_n^{\text{out}}\}$. We alternate between $\phi_{\text{in}}(\mathbf{k}_\parallel^{\text{in}})$ optimization and $\phi_{\text{out}}(\mathbf{k}_\parallel^{\text{out}})$ optimization until when M changes less than 5%, each time using the previous phase maps as the initial guess.

$\sum_{\mathbf{r}}$	$\sum_{\mathbf{k}_{\text{out}}}$	$\sum_{\mathbf{k}_{\text{in}}}$	$\sum_{\omega}$	Image quality metric landscape
Full (x, y from 0 to 50 μm)	✓	✓	✓	
Reduced (x, y from 25 to 35 μm)	✓	✓	✓	
Full	✓		✓	
Full	✓	✓		

Supplementary Fig. 10. **Effect of the quadruple summation on the image metric landscape.** The right-most column plots the image quality metric M of the USAF-target-under-tissue sample as a function of two Zernike coefficients of $\phi_{\text{out}}(\mathbf{k}_{\text{out}})$ when we perform the complete quadruple summation (top row), when we reduce the range of the spatial summation $\mathbf{r}$ (second row), when we restrict to a single incident angle $\mathbf{k}_{\text{in}}$ (third row), and when we restrict to a single frequency ω (bottom row) in the evaluation of M.

2. Image metric landscape without the quadruple summation

Evaluation of the image quality metric M in Eq. (S20) [Eq. (3) of the main text] involves a quadruple summation: summation over space $\mathbf{r}$ in Eq. (S20) and the triple gating through summations over $\mathbf{k}_{\text{out}}$, $\mathbf{k}_{\text{in}}$, and ω in the evaluation of I_{SMT} [Eq. (2) of the main text]. As discussed in the main text, all four of these summations are crucial in ensuring that the landscape of the image-quality metric reflects contributions from the target rather than from the multiple-scattering background or other unintended contributions. **Supplementary Fig. 10** shows the metric landscape as a function of two of the Zernike coefficients of $\phi_{\text{out}}(\mathbf{k}_{\text{out}})$ in the presence of the full quadruple summation and when one of the summations are skipped or reduced. We see that with the full quadruple summation, the metric landscape in this case exhibits only one local optima, so the outcome of the optimization is not sensitive to the initial guess. Without the quadruple summation, the landscape exhibits many local optima, and the optimization may end up in any of these local optima depending on the initial guess. The multitude of local optima arises because each speckle in the multiple-scattering background may be maximized by a different wavefront, unlike the target signals where all targets within the same isoplanatic patch share the same optimal wavefront.

3. Progressive reduction of zone size and increase of Zernike order

Since the 50 × 50 µm^2 field of view is larger than the isoplanatic patch size at the depths we consider, the spatial variation of the wavefront corrections must be accounted for. A simple strategy would be directly performing wavefront optimizations within spatial zones smaller than the isoplanatic patch size. However, as described in Sect. II of the main text and illustrated in **Supplementary Fig. 10**, summing over more spatial points inside the isoplanatic patch can boost the relative contribution of the target signals, making it less likely for the optimization to get stuck in a poor local optimum that corresponds to optimizing speckles. Therefore, the ideal zone size is the isoplanatic patch size—the maximal zone size within which the wavefront correction is invariant. As mentioned in the main text, the isoplanatic patch size depends on the order of the Zernike polynomials. This motivates us to adopt a progression strategy: first optimize the image over the full field of view with the low-order Zernike polynomials and with the whole image sharing the same phase map, then bisect the image into 2 × 2 zones and optimize the wavefronts for each zone while adding higher-order Zernike polynomials. The wavefronts obtained from the previous (larger-zone) optimization are used as the initial guess for the smaller zone; such "progression" ensures that the non-convex optimization ends up in a good local optimum. As we continue to bisect into smaller zones, we observe the image quality to converge, at which point we stop the bisection; this is where the zone size has become smaller than the isoplanatic patch size of the high-order corrections.

After optimization, the zones are stitched together to yield the full image. To avoid image discontinuity at the zone boundaries, we let neighboring zones overlap by 2.4 µm and stitch them using a weighted average with a linear ramp.

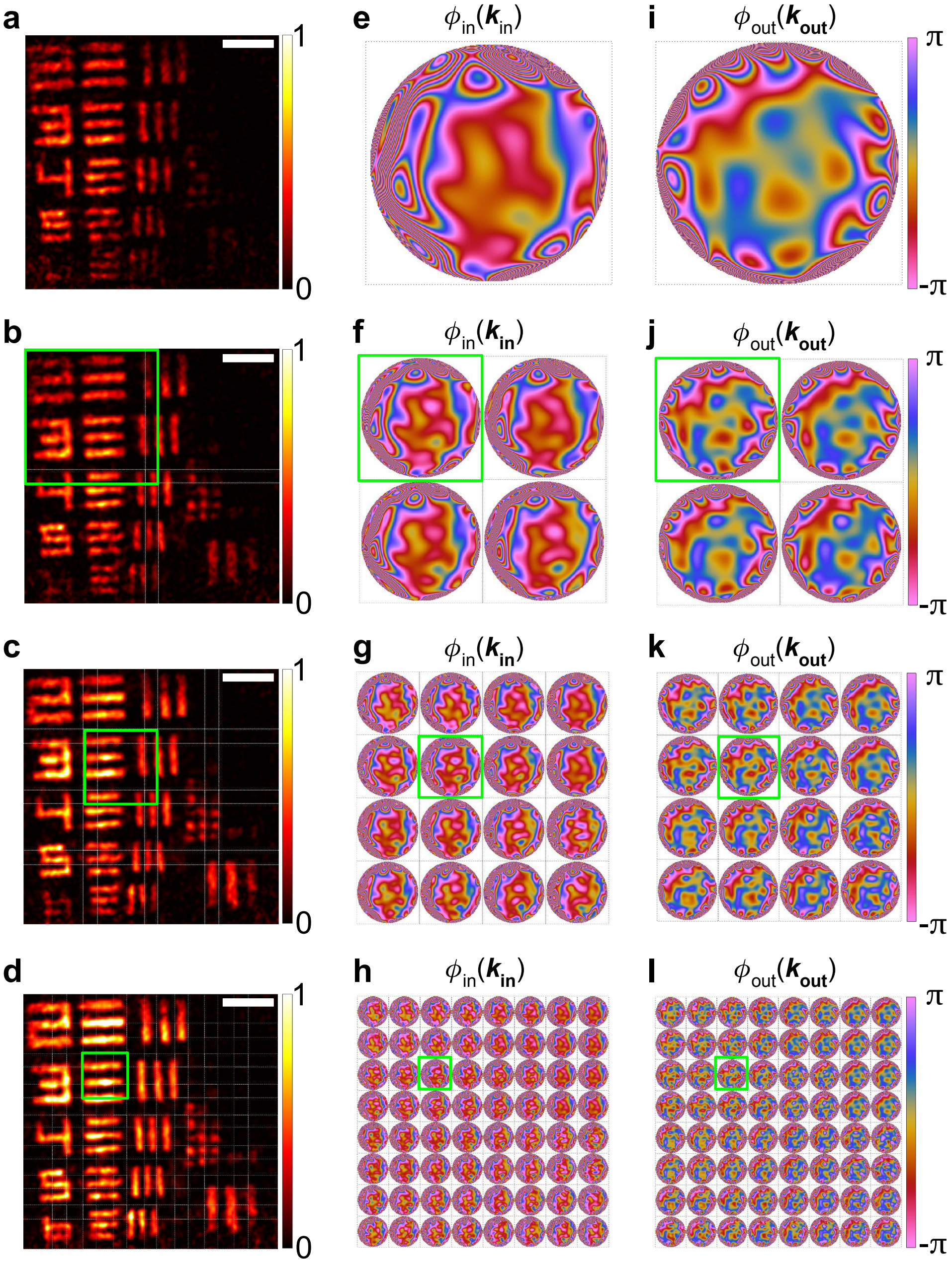
Supplementary Figs. 11–12 show the evolution of the image, the phase maps, and the image quality metric during this progression process for the USAF-target-under-tissue sample. Here, we use 11, 15, 19, and 22 radial orders of Zernike polynomials for the full image, 2 × 2 zones, 4 × 4 zones, and 8 × 8 zones respectively (corresponding to 77, 135, 209, and 275 Zernike polynomials). The number of Zernike polynomials can be relatively flexible, as long as it is not too small (insufficient to capture the higher-order wavefront distortions) or too large (which causes overfitting). Note that the x-axis of **Supplementary Fig. 12** is the number of objective function evaluations, not the L-BFGS iteration number. Within each L-BFGS iteration, there is a line search to refine the step size, during which the objective function may dip.

4. Spatial-basis truncation and angular-basis down sampling

A caveat of such image-metric-based wavefront optimization is that when there are strong scatterers outside the volume considered in the $\mathbf{r}$ summation of Eq. (S20), the optimization may overfit by distorting the wavefront to smear or to move such outside targets such that they contribute to the volume being optimized. Such a nuisance is particularly relevant for the progression strategy, where targets may lie on the zone boundaries. To address this issue, we Fourier transform the synthetic time-gated reflection matrix $\tilde{R}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}; z_j)$ in Eq. (S24) to spatial basis

$$\bar{R}(\mathbf{r}_{\parallel}^{\text{out}}, \mathbf{r}_{\parallel}^{\text{in}}; z_j) \equiv \sum_{\mathbf{k}_{\parallel}^{\text{out}}} e^{i\mathbf{k}_{\parallel}^{\text{out}}\cdot\mathbf{r}_{\parallel}^{\text{out}}} \sum_{\mathbf{k}_{\parallel}^{\text{in}}} e^{-i\mathbf{k}_{\parallel}^{\text{in}}\cdot\mathbf{r}_{\parallel}^{\text{in}}} \tilde{R}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}; z_j), \tag{S29}$$

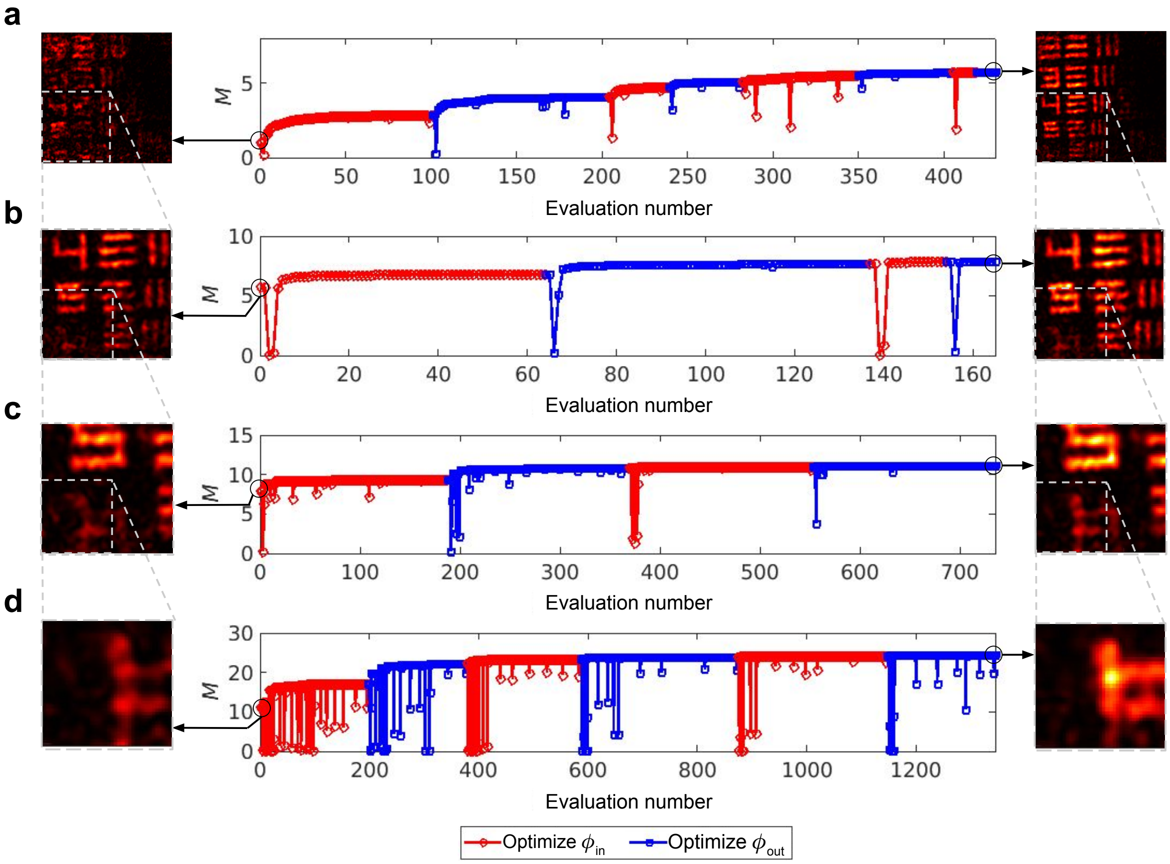
set the matrix elements of $\bar{R}(\mathbf{r}_{\parallel}^{\text{out}}, \mathbf{r}_{\parallel}^{\text{in}}; z_j)$ outside the zone boundary to zero, and then Fourier transform back to the $\mathbf{k}_{\parallel}$ basis before proceeding with the optimization. Doing so effectively prevents overfitting.



Supplementary Fig. 11. **Progressive bisection and wavefront correction maps for the USAF-target-under-tissue sample. a–d** The wavefront-corrected image during a progression that bisects the area at each step. Scale bars: 10 µm. All images share the same normalization. White dotted lines indicate zone boundaries, with 2.4 µm of overlap between neighboring zones. **e–l** The input and output wavefront correction maps of the zones.

Since $\bar{R}(\mathbf{r}_{\parallel}^{\text{out}}, \mathbf{r}_{\parallel}^{\text{in}}; z_j)$ has been truncated to a smaller spatial region, we can use $\tilde{R}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}; z_j)$ with a larger angular spacing when we convert back to the $\mathbf{k}_{\parallel}$ basis; doing so significantly reduces the computation time. Here, we set the input angular spacing and the output angular spacing of a divided zone both to $\delta k_x = \delta k_y = \sqrt{2}\pi/W_{\text{zone}}$ (slightly over-sampled above the Nyquist rate of $2\pi/W_{\text{zone}}$ but down-sampled compared to the original $\delta k = 2\pi/\text{FOV}$ spacing mentioned in Supplementary Sect. I C) where W_{zone} is the zone width. Note that the Zernike polynomials in Eq. (S27) can be evaluated at any set of angular grid points, so we can flexibly choose the angular spacing without affecting the optimization variables $\{c_n^{\text{in}}\}$ and $\{c_n^{\text{out}}\}$.

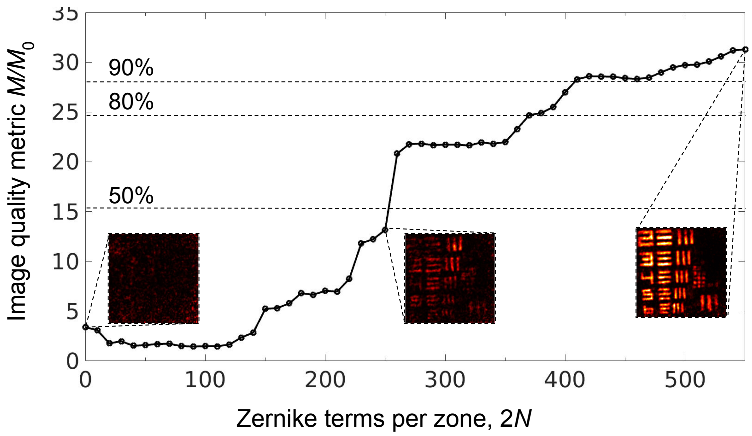
This spatial-basis truncation and angular-basis down-sampling are only used during the optimization. We use the complete scattering matrix data for the image reconstruction (after the phase maps are determined).



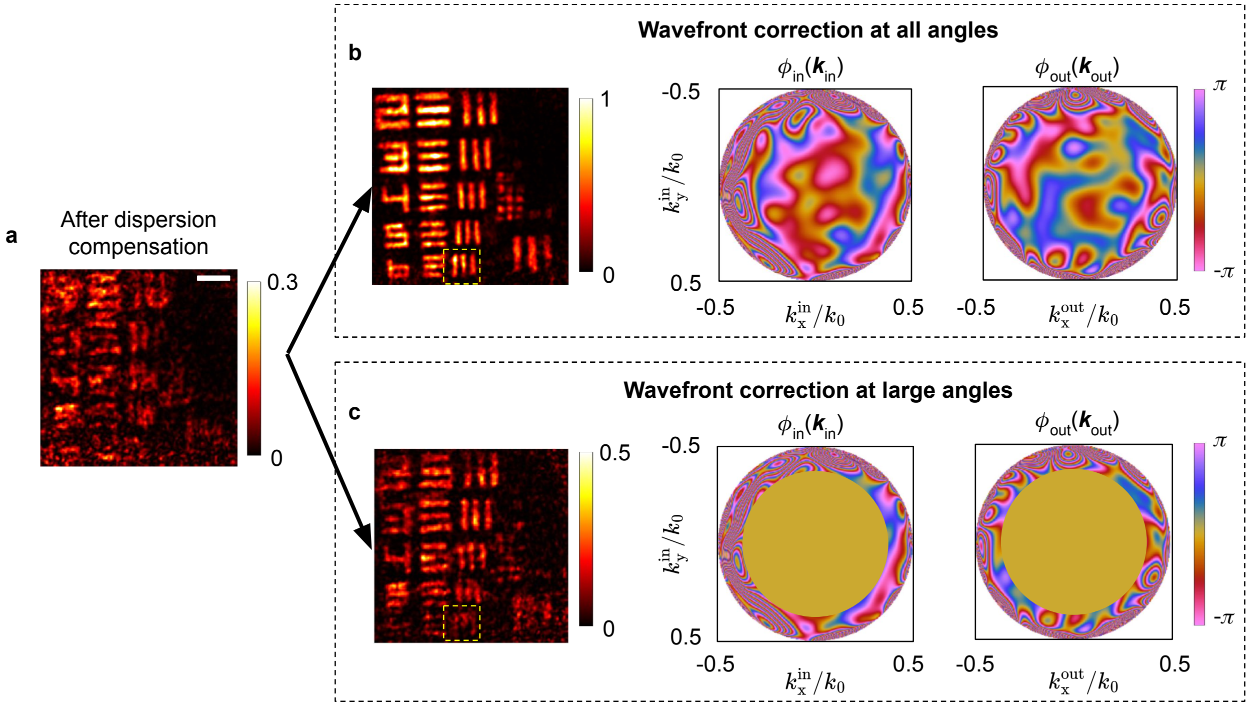
Supplementary Fig. 12. **Evolution of the image quality metric M during the optimization process**. Each panel shows the evolution for one zone as the zone size is reduced. The inset figures on the left and right are the images before and after optimization. All images share the same normalization as **Supplementary Fig. 11**.

5. *Contribution of the high-order terms*

Our wavefront correction employees a large number of variables. In the last iteration of the progression, in each of the 64 spatial zones and for both the incident and the return paths, we include 275 Zernike polynomials for the USAF-target-under-tissue sample, bringing the total number of wavefront optimization variables to $64 \times 2 \times 275 = 35,200$. One may wonder if all of these variables actually contribute. For example, it is possible that the highest-order terms



Supplementary Fig. 13. **Contribution of the high-order wavefront correction terms.** The plot shows the image quality metric M of the USAF-target-under-tissue sample when the number of Zernike terms per zone is truncated to $2N$ (N terms for ϕ_{in}; N terms for ϕ_{out}), normalized by the image quality metric M_0 before dispersion compensation.



Supplementary Fig. 14. **Contribution of wavefront correction at large angles.** SMT image with full-angle correction (top row) and with correction only for large angles (bottom row). In the latter case, the Zernike coefficients are reused from the first case but with $\phi_{\rm in/out}(\mathbf{k}_{\|}^{\rm in/out}) = 0$ for all $|\mathbf{k}_{\|}^{\rm in/out}| < 0.38k_0$. Both wavefront-corrected images are divided into 8 × 8 zones, and we show the phase maps in the zone marked by a yellow dashed box. Scale bar: 10 µm.

have a small amplitude or are large but random numbers that do not contribute to the image quality improvement.

To find out, we take the optimized Zernike coefficients (from **Supplementary Fig. 11**h,l), keep only N lowest-order Zernike terms for each zone for each path (*i.e.*, set $c_n^{\rm in} = c_n^{\rm out} = 0$ for all $n > N$), and build the SMT image with wavefront corrections $\phi_{\rm in}(\mathbf{k}_{\|}^{\rm in})$ and $\phi_{\rm out}(\mathbf{k}_{\|}^{\rm out})$ using the truncated Zernike coefficients. **Supplementary Fig. 13** shows the image quality metric M as a function of Zernike terms per zone ($2N$). We find there is no plateau at large N, indicating that the highest-order terms do contribute meaningfully to improving the image quality. Note that there is a dip at $2N < 40$, which highlights that the low-order terms are insufficient for correcting the wavefront distortion and that they need to work together with the high-order terms to provide a meaningful correction.

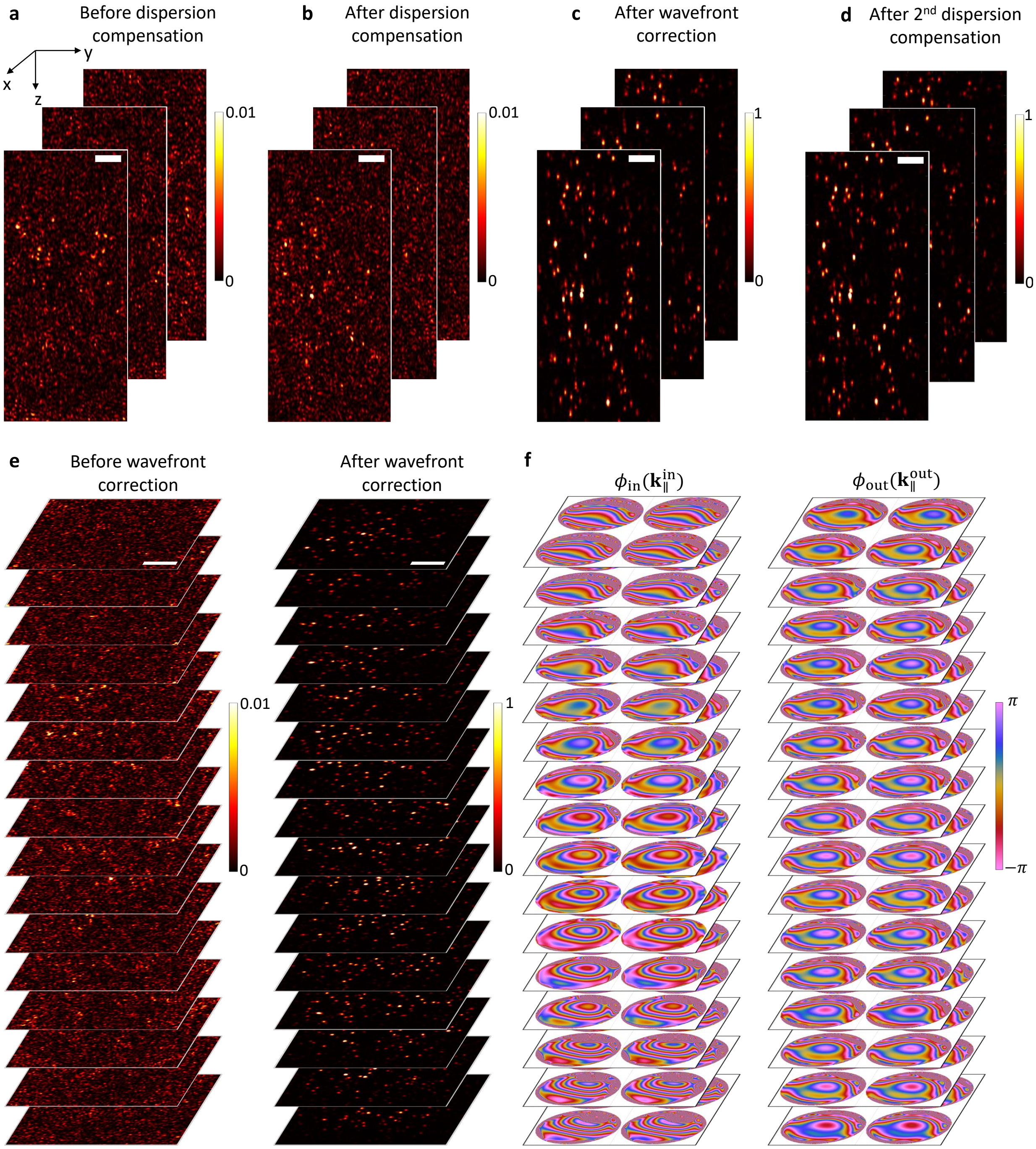
One may use this plot to quantify a number of effective degrees of freedom in the wavefront correction. For example, if one is willing to tolerate a 20% reduction of the image quality metric, we can follow the 80% horizontal dashed line to find that 370 Zernike terms per zone are needed, totaling 23,680 degrees of freedom across the 64 zones.

6. *Contribution of the large-angle corrections*

As shown in **Fig. 4**h of the main text, the optimized wavefront correction patterns $\phi_{\rm in}(\mathbf{k}_{\|}^{\rm in})$ and $\phi_{\rm out}(\mathbf{k}_{\|}^{\rm out})$ exhibit faster oscillations at larger angles, corresponding to higher-order corrections. To assess the importance of these large-angle higher-order corrections, we compare in **Supplementary Fig. 14** the SMT image with full-angle correction (top row) and with correction only for angles with $|\mathbf{k}_{\|}^{\rm in/out}| > 0.38k_0$ (bottom row). We see that the wavefront corrections at large angles provide visible image quality improvement. The wavefront corrections at large angles are more fast-varying than at small angles near the center of the pupils, indicating the existence of high-order Zernike modes.

7. *Volumetric wavefront correction*

Supplementary Fig. 15 shows cross sections of the nanoparticle colloid sample before and after wavefront corrections and the associated phase maps. As mentioned in Supplementary Sect. III F 1, the 110 µm depth of field in z is divided into 16 overlapping sub-volumes, and five equally spaced slices within each sub-volume are used to compute the image quality metric for that sub-volume. The images of the sub-volumes are stitched together similar to how



Supplementary Fig. 15. **Volumetric dispersion compensation and wavefront correction for the nanoparticle colloid sample. a–d**, YZ cross-section images at $x = 18$, 24, and 30 µm before dispersion compensation (a), after dispersion compensation (b), after wavefront correction (c), and after the second dispersion compensation (d). **e**, XY *en face* images at $z = 1.476$, 1.483, 1.490, 1.496, 1.502, 1.509, 1.515, 1.522, 1.528, 1.535, 1.541, 1.548, 1.554, 1.561, 1.567, and 1.575 mm before and after wavefront correction. Scale bars: 10 µm. **f**, The corresponding input and output wavefront correction phase maps.

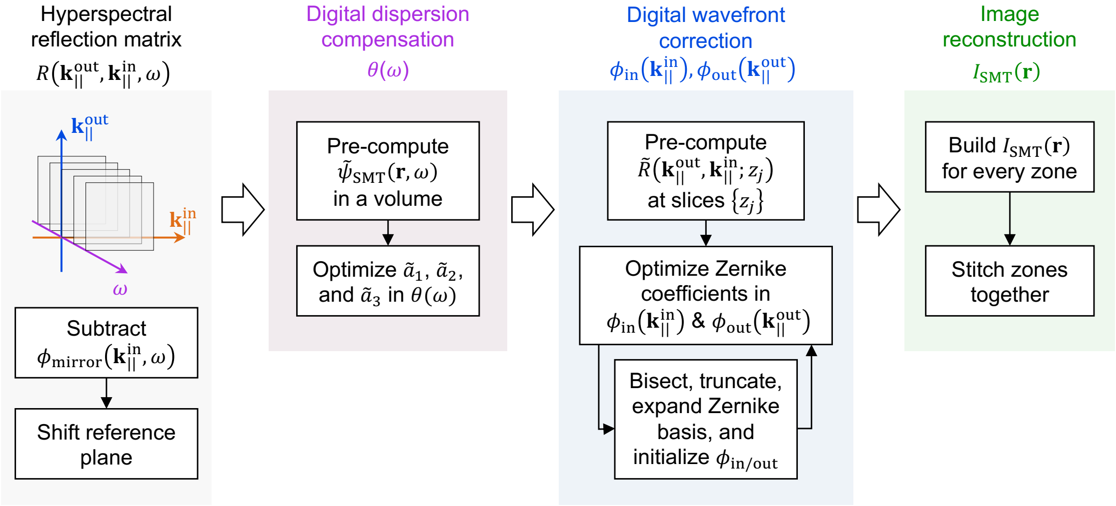
the spatial zones in x-y are stitched. The wavefront distortions are less severe in this sample, so we only divide up to 2×2 zones in the x-y directions, using 11 and 15 radial orders for the full FOV and 2×2 zones respectively (77 and 135 Zernike terms). To maintain the axial resolution, after the wavefront correction, we optimize $\tilde{a}_2$ and $\tilde{a}_3$ in the dispersion compensation again for each sub-volume.

8. Potential solution for frequency-dependent wavefront distortions

The current framework assumes sample-induced aberrations are the same for all frequencies. However, in general, they can be frequency-dependent. In that case, aberration correction has to be carried out for each frequency or each narrow frequency band where the aberrations remain relatively frequency-invariant. In particular, in Eq. S24 , the frequency summation is restricted to each narrow frequency band instead of the entire spectrum. Then, aberration correction is performed on each frequency band. Note that by reducing the frequency to narrow bands, there are fewer frequencies to sum over to suppress multiple scattering, the wavefront correction is now more susceptible to multiple scattering noise. To ensure the optimization's robustness to multiple scattering, the multiscale progression strategy can also be adopted, assuming the change of aberrations from one frequency to another is gradual. In particular, optimization can be performed on all frequencies first. Then, the frequency band is gradually shrunk and optimized again until the image no longer improves. The resulting correction phases of the wider frequency band act as the initial guess for the narrower frequency band. On the USAF target under mouse brain, this approach achieves a similar final image as the all-frequency aberration correction shown on **Supplementary Fig. 11**. Thus, further testings on other samples are needed to assess if frequency-dependent aberration correction leads to better images than frequency-independent correction.

G. Digital correction flowchart

Supplementary Fig. 16 shows a flowchart of the digital correction and image reconstruction process.



Supplementary Fig. 16. Flowchart of SMT digital corrections and image construction.

H. Time usage and memory consumption

Supplementary Table 1 summarizes the time and memory usage for the digital dispersion compensation, wavefront correction, and image reconstruction. All tasks are performed using MATLAB on a personal laptop with an Intel Xeon E-2276M CPU with no GPU, 2 × 16 GB Dell DDR4 SODIMM 2666 MHz memory, and a KXG60ZNV512G Toshiba SSD. All data are stored and processed in the single-precision floating-point format. Note that the table does not contain the time and memory consumption of the data transferring phase between the camera and the computer, which takes around 20 minutes to load 100 GB of raw interferograms of both the samples and the reference mirror.

IV. RCM, OCT, OCM/CASS, CAO-OCM, VRM, AND EIGENCHANNELS APPROACH

With the measured hyperspectral reflection matrix, we also perform virtual experiments to mimic other reflection-based imaging methods. Here, we describe how we synthesize reflectance confocal microscopy (RCM), optical coherence tomography (OCT), optical coherence microscopy (OCM) (which is equivalent to "collective accumulation of single scattering" (CASS) microscopy), computational adaptive optics OCM (CAO-OCM), and "volumetric reflection-matrix microscopy" (VRM).

Supplementary Table 1. Time and memory usage of SMT dispersion compensation, wavefront correction, and image reconstruction. Each x-y slice covers 50×50 µm^2, with a $0.7 \times 0.7 \times 1.1$ µm^3 grid spacing. The precomputations refer to Eq. (S19), Eq. (S24), and the Zernike matrices in Eq. (S27). For the nanoparticle colloid, we perform dispersion compensation for the full volume together, wavefront correction for one sub-volume at a time (5 slices of $\{z_j\}$ per sub-volume).

Task		2D USAF target under tissue		3D nanoparticle colloid	
		Time (min)	Memory (GB)	Time (min)	Memory (GB)
Dispersion compensation	Data loading & precomputation	6.8	1.2	8.5	2.7
	Scan $\tilde{a}_1$, $\tilde{a}_3$, $\tilde{a}_3$	0.66		1.0	
	Gradient-based optimization	0.03		0.03	
Wavefront correction	Data loading & precomputation	5.0	0.2	8.7 (per sub-volume)	1.2
	1 x 1 zone	1	4.0	5.0 (per sub-volume)	6.2
	2 x 2 zones (all zones)	1.5		7.5 (per sub-volume)	
	4 x 4 zones (all zones)	1.0		N/A	
	8 x 8 zones (all zones)	1.6			
Image reconstruction	2D slice given precomputation	0.2	0.3	0.03	0.3
	2D slice w/o precomputation	5.0	0.5	7.5	0.5
	3D volume w/o precomputation	N/A	N/A	152 (100 slices)	3.5

A. RCM

RCM uses a single frequency $\omega = \omega_{\mathrm{RCM}}$ with no time gating while spatially filtering both the incident and the reflected light with a high-NA objective lens [4]. Therefore, to synthesize an RCM image with index-mismatch correction, we use

$$I_{\mathrm{RCM}}(\mathbf{r}) = \left| \sum_{\mathbf{k}_\parallel^{\mathrm{out}}, \mathbf{k}_\parallel^{\mathrm{in}}}^{\mathrm{NA}=0.5} e^{i(\mathbf{k}_{\mathrm{out}}^{\mathrm{sam}} - \mathbf{k}_{\mathrm{in}}^{\mathrm{sam}})\cdot\mathbf{r}} R(\mathbf{k}_\parallel^{\mathrm{out}}, \mathbf{k}_\parallel^{\mathrm{in}}, \omega = \omega_{\mathrm{RCM}}) \right|^2, \tag{S30}$$

with $R(\mathbf{k}_\parallel^{\mathrm{out}}, \mathbf{k}_\parallel^{\mathrm{in}}, \omega)$ defined in Eq. (S18b) and with $\mathbf{k}_{\mathrm{out}}^{\mathrm{sam}}$ and $\mathbf{k}_{\mathrm{in}}^{\mathrm{sam}}$ defined below Eq. (S6) and Eq. (S7). Note that, following Supplementary Sect. III B, the $R(\mathbf{k}_\parallel^{\mathrm{out}}, \mathbf{k}_\parallel^{\mathrm{in}}, \omega)$ here has already corrected for the dispersion and aberrations in the input path of the optical system. We use this version for the 3D nanoparticle colloid sample in Fig. 5 of the main text in order to compare the SMT and RCM images at the same depths.

Typically, the index mismatch is not corrected in RCM. To synthesize an RCM image without index-mismatch correction, we use

$$I_{\mathrm{RCM}}(\mathbf{r}') = \left| \sum_{\mathbf{k}_\parallel^{\mathrm{out}}, \mathbf{k}_\parallel^{\mathrm{in}}}^{\mathrm{NA}=0.5} e^{i(\mathbf{k}_{\mathrm{out}}^{\mathrm{air}} - \mathbf{k}_{\mathrm{in}}^{\mathrm{air}})\cdot\mathbf{r}'} R'(\mathbf{k}_\parallel^{\mathrm{out}}, \mathbf{k}_\parallel^{\mathrm{in}}, \omega = \omega_{\mathrm{RCM}}) \right|^2, \tag{S31}$$

with $R'(\mathbf{k}_\parallel^{\mathrm{out}}, \mathbf{k}_\parallel^{\mathrm{in}}, \omega)$ being the sample reflection matrix after removing aberrations in the input path by subtracting $\Delta\phi_{\mathrm{mirror}}(\mathbf{k}_\parallel^{\mathrm{in}}, \omega)$ (see Supplementary Sect. III B) but before index-mismatch correction, which has its in-air reference plane at $z_{\text{ref}}^{\text{air}} = z_{\text{mirror}}$. Here, $\mathbf{r}' = (x, y, z')$, and $z' = z^{\text{air}} - z_{\text{ref}}^{\text{air}}$ is the apparent depth of the spatial focus looking from air relative to the reference plane in air. For the 2D USAF target under tissue sample in Fig. 4 of the main text, we use this version and choose $z' = z'_{\text{target}}$ to be the depth that maximizes the image intensity integrated over x-y.

B. OCT/OCM/CASS

For OCT and OCM, the spatial gate has its depth z_{f} fixed on the focal plane and is scanned in x and y. The z dimension of the image is mapped out through time/coherence gating. They have a limited depth of field because the strength of the spatial gate and the lateral resolution degrades when $|z - z_{\text{f}}|$ is larger than the Rayleigh range z_{R}. We synthesize an OCM image with index-mismatch correction through

$$I_{\text{OCM}}(\mathbf{r}) = \left| \sum_{\omega} e^{i\theta_{\text{OCM}}(\omega) - 2i\omega\Delta t_{\text{f}}} \sum_{\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.5} e^{i(\mathbf{k}_{\text{out}}^{\text{sam}} - \mathbf{k}_{\text{in}}^{\text{sam}})\cdot \mathbf{r}_{\text{f}}} R(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega) \right|^2, \tag{S32}$$

with $\mathbf{r} = (x, y, z)$, $\Delta t_{\text{f}} = n_{\text{sam}}(z - z_{\text{f}})/c$ being the additional (or reduced) travel time to propagate from z_{f} to z in the sample, and $\mathbf{r}_{\text{f}} = (x, y, z_{\text{f}})$ the location of the spatial gate. We perform the same dispersion compensation as in SMT (Supplementary Sect. III E) to determine the spectral phase correction $\theta_{\text{OCM}}(\omega)$. We use this version for the 3D nanoparticle colloid sample in Fig. 5 of the main text in order to compare the SMT and RCM images at the same depths.

Typically, the index mismatch is not corrected in OCM. To synthesize an OCM image without index-mismatch correction, we use

$$I_{\text{OCM}}(\mathbf{r}') = \left| \sum_{\omega} e^{i\theta_{\text{OCM}}(\omega) - 2i\omega\Delta t'_{\text{f}}} \sum_{\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.5} e^{i(\mathbf{k}_{\text{out}}^{\text{air}} - \mathbf{k}_{\text{in}}^{\text{air}})\cdot \mathbf{r}'_{\text{f}}} R'(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega) \right|^2, \tag{S33}$$

with $\mathbf{r}' = (x, y, z')$, $z'_{\text{f}} = z_{\text{f}}^{\text{air}} - z_{\text{ref}}^{\text{air}}$ the apparent depth of the spatial gate relative to the in-air reference plane of the reflection matrix $R'(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega)$, $\mathbf{r}'_{\text{f}} = (x, y, z'_{\text{f}})$, and $\Delta t'_{\text{f}} = [n_{\text{glass}} h^{\text{glass}} + n_{\text{sam}} z - (h^{\text{glass}} + z_{\text{f}}^{\text{air}})]/c$ the additional (or reduced) travel time to propagate to z in the sample relative to propagating to the apparent depth $z_{\text{f}}^{\text{air}}$ of the spatial gate in air (see **Supplementary Fig. 8**). For the 2D USAF target under tissue sample in Fig. 4 of the main text, we use this version and choose $z' = z'_{\text{f}} = z'_{\text{target}}$ to be the depth that maximizes the image intensity integrated over x-y.

To synthesize OCT images, we use a smaller NA = 0.1. For $\mathbf{k}_{\parallel}^{\text{in}}$ near the normal incidence, data points with $\mathbf{k}_{\parallel}^{\text{out}}$ near the normal incidence are excluded from our reflection matrix due to reflection from the objective lens (Supplementary Sect. I C). Therefore, for the synthesized OCT here, we use $\mathbf{k}_{\parallel}^{\text{in}}$ and $\mathbf{k}_{\parallel}^{\text{out}}$ within a disk with NA = 0.1 centered around angle $\alpha_{\text{in}} = \alpha_{\text{out}} = 10°$. With index-mismatch correction, we have

$$I_{\text{OCT}}(\mathbf{r}) = \left| \sum_{\omega} e^{i\theta_{\text{OCT}}(\omega) - i\omega\Delta t_{\text{f}}\left(\frac{1}{\cos\alpha_{\text{in}}} + \frac{1}{\cos\alpha_{\text{out}}}\right)} \sum_{\mathbf{k}_{\parallel}^{\text{out}}}^{\text{NA}=0.1,\alpha_{\text{out}}} \sum_{\mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.1,\alpha_{\text{in}}} e^{i(\mathbf{k}_{\text{out}}^{\text{sam}} - \mathbf{k}_{\text{in}}^{\text{sam}})\cdot \mathbf{r}_{\text{f}}} R(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega) \right|^2. \tag{S34}$$

Similar to OCM, we build OCT images without index-mismatch correction with

$$I_{\text{OCT}}(\mathbf{r}') = \left| \sum_{\omega} e^{i\theta_{\text{OCT}}(\omega) - i\omega\Delta t'_{\text{f}}\left(\frac{1}{\cos\alpha_{\text{in}}} + \frac{1}{\cos\alpha_{\text{out}}}\right)} \sum_{\mathbf{k}_{\parallel}^{\text{out}}}^{\text{NA}=0.1,\alpha_{\text{out}}} \sum_{\mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.1,\alpha_{\text{in}}} e^{i(\mathbf{k}_{\text{out}}^{\text{air}} - \mathbf{k}_{\text{in}}^{\text{air}})\cdot \mathbf{r}'_{\text{f}}} R'(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega) \right|^2. \tag{S35}$$

In the "collective accumulation of single-scattering" (CASS) microscopy [22], the time-gated reflection matrix at one focal plane $z = z_{\text{f}}$ is measured using a pulsed laser. The time-gated reflection matrix $\tilde{\tilde{R}}(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{in}}; t_{\text{f}})$ at arrival time t_{f} is related to the frequency-dependent reflection matrix through [87]

$$\tilde{\tilde{R}}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}; t_{\text{f}}) = \sum_{\omega} f(\omega) e^{-i\omega t} R(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega), \tag{S36}$$

where $f(\omega)$ is the field spectrum of the incident pulse. Comparing this to the synthetic time-gated reflection matrix $\tilde{R}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}; z_j)$ we defined earlier in Eq. (S24), we see that $\tilde{\tilde{R}}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}; t_{\text{f}}) \approx \tilde{R}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}; z_j = z_{\text{f}})$ if we ignore the dispersion compensation $\theta(\omega)$ in $\tilde{R}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}; z_j)$ and set the reference plane at $z_j = z_{\text{f}} = 0$ and $t = 0$.

In CASS, matrix elements with the same in-plane momentum difference $\Delta\mathbf{k}_\parallel = \mathbf{k}^{\rm out}_\parallel - \mathbf{k}^{\rm in}_\parallel$, denoted as $\mathbf{k}^{\rm s}$ in Ref. [22], are summed over to accumulate the single-scattering contribution. Then, an inverse Fourier transform is performed to go from the momentum space $\Delta\mathbf{k}_\parallel$ to the real space $\mathbf{r}_\parallel$ to produce an image on the focal plane:

$$I_{\rm CASS}(\mathbf{r}_\parallel) = \left| \sum_{\Delta\mathbf{k}_\parallel} e^{i\Delta\mathbf{k}_\parallel\cdot\mathbf{r}_\parallel} \left[\sum_{\mathbf{k}^{\rm in}_\parallel} \tilde{\tilde{R}}(\mathbf{k}^{\rm in}_\parallel + \Delta\mathbf{k}_\parallel, \mathbf{k}^{\rm in}_\parallel; t_{\rm f}) \right] \right|^2. \tag{S37}$$

By treating $\tilde{\tilde{R}}(\mathbf{k}^{\rm out}_\parallel, \mathbf{k}^{\rm in}_\parallel; t_{\rm f}) \approx \tilde{R}(\mathbf{k}^{\rm out}_\parallel, \mathbf{k}^{\rm in}_\parallel; z_j = z_{\rm f})$, inserting the definition of $\tilde{R}(\mathbf{k}^{\rm out}_\parallel, \mathbf{k}^{\rm in}_\parallel; z_j)$ from Eq. (S24), and comparing to Eq. (S32), we can see that the CASS image equals the OCM image on the focal plane,

$$I_{\rm CASS}(\mathbf{r}_\parallel) \approx I_{\rm OCM}(\mathbf{r}_\parallel, z_{\rm f}). \tag{S38}$$

In the frequency-domain OCM images we synthesize, we have removed the dispersion and aberrations in the input path of the optical system by subtracting $\Delta\phi_{\rm mirror}(\mathbf{k}^{\rm in}_\parallel, \omega)$ (see Supplementary Sect. III B) and additionally compensated for the sample dispersion (Supplementary Sect. III E). A CASS image does not come with these corrections, so it would have a lower quality than the OCM image we synthesize.

C. CAO-OCM

The interferometric measurement of OCT and OCM provides phase information that can be used for digital aberration correction [18–21, 88, 89]. However, the correction only applies to the phase image itself; it cannot apply separately to the input spatial gate and the output spatial gate since OCT/OCM does not measure the reflection matrix. In this section, we consider computational adaptive optics (CAO) corrections onto the confocal OCM image, referred to as CAO-OCM [19–21, 88]. CAO for full-field OCT without input spatial gate [18, 89] has a lower performance due to the lack of confocal gating, which we consider later as a limiting case of SMT in Supplementary Sect. V B.

We start by synthesizing the phase image of OCM with index-mismatch correction and dispersion compensation, from Eq. (S32), at the focal plane $z = z_{\rm f}$,

$$\psi_{\rm OCM}(\mathbf{r}_\parallel; z_{\rm f}) = \sum_{\omega} e^{i\theta_{\rm OCM}(\omega)} \sum_{\mathbf{k}^{\rm out}_\parallel, \mathbf{k}^{\rm in}_\parallel} e^{i(\mathbf{k}^{\rm sam}_{\rm out} - \mathbf{k}^{\rm sam}_{\rm in})\cdot\mathbf{r}_{\rm f}} R(\mathbf{k}^{\rm out}_\parallel, \mathbf{k}^{\rm in}_\parallel, \omega). \tag{S39}$$

In CAO-OCM, the aberrations are corrected in the reciprocal space of the *en face* image [19–21, 88], $\mathbf{q}_\parallel = \mathbf{k}^{\rm out}_\parallel - \mathbf{k}^{\rm in}_\parallel$. So, we consider the 2D Fourier transform of the phase image,

$$\tilde{\psi}_{\rm OCM}(\mathbf{q}_\parallel; z_{\rm f}) = \sum_{\mathbf{r}_\parallel} e^{-i\mathbf{q}_\parallel\cdot\mathbf{r}_\parallel} \psi_{\rm OCM}(\mathbf{r}_\parallel; z_{\rm f}). \tag{S40}$$

After applying phase correction $e^{i\phi(\mathbf{q}_\parallel)}$, the image is inverse Fourier transformed back to the real space

$$\psi_{\rm CAO}(\mathbf{r}_\parallel; z_{\rm f}) = \sum_{\mathbf{q}_\parallel} e^{i\mathbf{q}_\parallel\cdot\mathbf{r}_\parallel + i\phi(\mathbf{q}_\parallel)} \tilde{\psi}_{\rm OCM}(\mathbf{q}_\parallel; z_{\rm f}) \tag{S41}$$

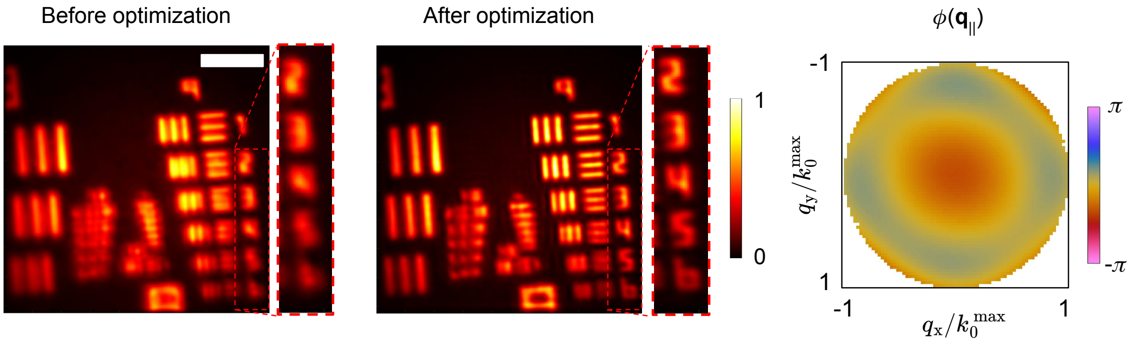
to form the corrected image $I_{\rm CAO}(\mathbf{r}_\parallel; z_{\rm f}) = |\psi_{\rm CAO}(\mathbf{r}_\parallel; z_{\rm f})|^2$.

To determine the phase correction $\phi(\mathbf{q}_\parallel)$, we follow the same strategy as in Supplementary Sect. III F, expanding it in Zernike polynomials like in Eq. (S27),

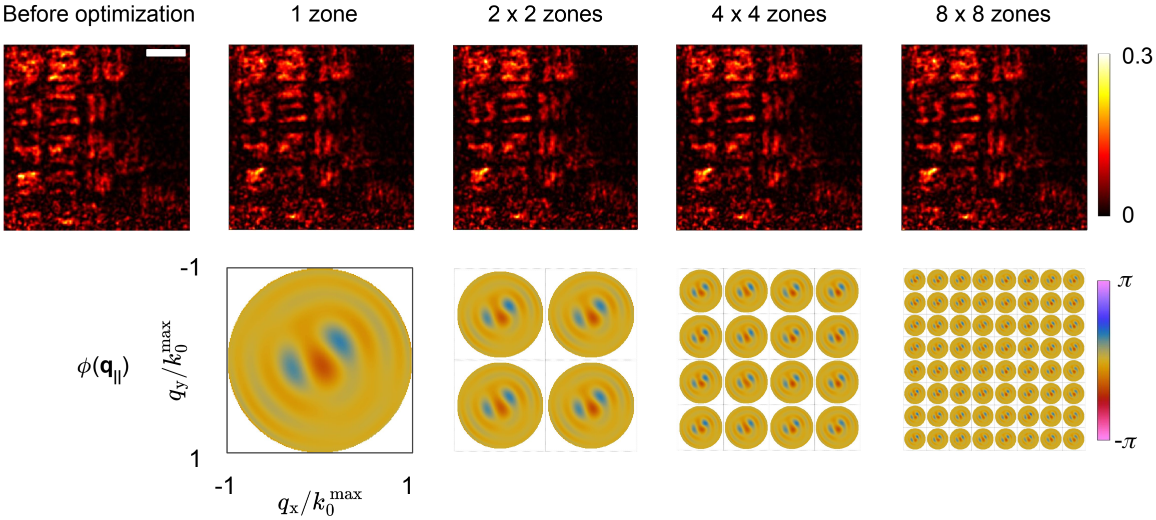
$$\phi(\mathbf{q}_\parallel) = \sum_{n} c_n Z_n\left(\frac{\mathbf{q}_\parallel}{2k_0^{\max}{\rm NA}}\right), \tag{S42}$$

with $k_0^{\max} = \omega_{\max}/c$. The Zernike coefficients $\{c_n\}$ are then determined by maximizing the same image quality metric $M = \sum_{\mathbf{r}} I_{\rm CAO}(\mathbf{r}) \ln\left[I_{\rm CAO}(\mathbf{r})/I_0\right]$ using the same L-BFGS algorithm with initial guess $c_n = 0$. Here, the gradient of M with respect to $\{c_n\}$ is

$$\frac{\partial M}{\partial c_n} = -2 \sum_{\mathbf{r}_\parallel, \mathbf{q}_\parallel} {\rm Im}\left\{ \left[1 + \ln\frac{I_{\rm CAO}(\mathbf{r}_\parallel; z_{\rm f})}{I_0}\right] \psi^*_{\rm CAO}(\mathbf{r}_\parallel; z_{\rm f}) \tilde{\psi}_{\rm OCM}(\mathbf{q}_\parallel; z_{\rm f}) e^{i\mathbf{q}_\parallel\cdot\mathbf{r}_\parallel + i\phi(\mathbf{q}_\parallel)} \right\} Z_n\left(\frac{\mathbf{q}_\parallel}{2k_0^{\max}{\rm NA}}\right). \tag{S43}$$



Supplementary Fig. 17. **Validation of OCM corrected with computational adaptive optics (CAO-OCM).** The sample is a USAF target in air, not covered by any scattering medium. Scale bar: 10 µm.



Supplementary Fig. 18. **CAO-OCM images of USAF target under 1-mm of mouse brain tissue.** The same progressive optimization as in SMT is applied. The bottom row shows the corresponding phase correction maps in the reciprocal space $\mathbf{q}_{\|}$. Only some low-order variations get corrected despite allowing up to 275 Zernike polynomials per zone because the high-order wavefront distortions can only be corrected separately through $\phi_{\rm in}(\mathbf{k}_{\|}^{\rm in})$ and $\phi_{\rm out}(\mathbf{k}_{\|}^{\rm out})$. Scale bar: 10 µm.

We adopt the same progressive bisection strategy as described in Supplementary Sect. III F, gradually dividing the FOV into smaller zones while increasing the number of Zernike polynomials per zone and truncating $\psi_{\rm CAO}(\mathbf{r}_{\|}; z_{\rm f})$ to within the target zone.

To validate our implementation of CAO-OCM, we first test it on a bare USAF target not covered by any scattering medium. Before the CAO-OCM optimization, the aberrations in the input path of the optical system are already removed by subtracting $\Delta\phi_{\rm mirror}(\mathbf{k}_{\|}^{\rm in}, \omega)$ (Supplementary Sect. III B), but there are still aberrations in the output path, which is visible when we image group 9 of the USAF target with a reduced NA of 0.25. As shown in **Supplementary Figs. 17**, CAO-OCM successfully removes these low-order aberrations and improves the image quality. Progressive bisection is not necessary since the isoplanatic patch size of these low-order aberrations is very large.

Next, we test CAO-OCM for the USAF-target-under-tissue sample considered in the main text. **Supplementary Fig. 18** shows the CAO-OCM images and the corresponding phase correction maps during the progressive optimization. The final image here is also shown in Fig. 5f of the main text. Here we use the same focal depth $z_{\rm f} = 977.5$ µm, the same number of Zernike polynomials, and the same momentum spacing δq as in **Supplementary Figs. 11**. Despite applying all of these strategies and allowing up to 275 Zernike polynomials per zone, the optimization only identified some low-order aberrations, with no significant improvement to the image. This is because (1) prior to this optimization, the low-order aberrations in the input path of the optical system have already been removed through a calibration mirror (Supplementary Sect. III B), and (2) double-path wavefront correction is crucial for the high-order terms [21] but is not possible with CAO-OCM.

D. Volumetric reflection-matrix microscopy (VRM) and CLASS

Here we detail our implementation of the VRM method [35], present analyses that reveal the quantities that VRM implicitly maximizes, and compare the VRM method to the SMT method. The wavefront correction part of VRM is identical to the "closed-loop accumulation of single-scattering" (CLASS) [26, 28], which is also mathematically equivalent to the "distortion matrix" approach with iterative phase reversal [32–34].

1. *VRM spectral dispersion compensation*

The very first step of VRM is to "minimize defocus error" through "a numerical propagation to the depth with the maximum reflectance ($z_{\max}$)" [35]:

$$R_{\max}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) \equiv R(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) e^{i(k_z^{\rm out} - k_z^{\rm in}) z_{\max}}. \tag{S44}$$

Doing so places the reference plane of the reflection matrix at $z_{\max}$. This step is only mentioned on page 5 of the Supplementary Materials of Ref. [35], but it is important because our analysis below shows that the VRM dispersion compensation implicitly maximizes the image brightness at this particular depth $z_{\max}$. To find the "maximum reflectance," we build the volumetric SMT image without corrections using Eq. (1) of the main text, and choose $z_{\max}$ to be the depth where the x-y-integrated image intensity is the largest. We start with inclusion of the index-mismatch correction, so the $R(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega)$ in Eq. (S44) is the corrected one defined in Eq. (S18b), the $k_z^{\rm in}$ and $k_z^{\rm out}$ in Eq. (S44) are abbreviations of $k_{z,\rm in}^{\rm sam}$ and $k_{z,\rm out}^{\rm sam}$, and the $\mathbf{k}_{\rm out}$ and $\mathbf{k}_{\rm in}$ below are abbreviations of $\mathbf{k}_{\rm out}^{\rm sam}$ and $\mathbf{k}_{\rm in}^{\rm sam}$.

VRM then uses an iterative phase conjugation scheme to estimate the spectro-angular dispersion $\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in}, \omega)$ and $\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out}, \omega)$ relative to the center frequency $\omega_{\rm c}$. At the n-th iteration, a compensation of $-\Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in}, \omega)$ and $-\Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \omega)$ is applied onto the columns and rows of the reflection matrix,

$$R^{(n+1)}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) = e^{-i\Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \omega)} R^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) e^{-i\Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in}, \omega)}, \tag{S45}$$

with $n = 0, 1, 2, \ldots$ and with initial condition

$$\begin{aligned} R^{(0)}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) &= R_{\max}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega), \\ \Delta\phi_{\rm in}^{(0)}(\mathbf{k}_{\parallel}^{\rm in}, \omega) &= \Delta\phi_{\rm out}^{(0)}(\mathbf{k}_{\parallel}^{\rm out}, \omega) = 0. \end{aligned} \tag{S46}$$

The compensation at the n-th iteration is determined by phase conjugation,

$$\Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in}, \omega) = \arg\left[\sum_{\mathbf{k}_{\parallel}^{\rm out}} R^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) R^{(n)*}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega_{\rm c})\right], \tag{S47}$$

$$\Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \omega) = \arg\left[\sum_{\mathbf{k}_{\parallel}^{\rm in}} R^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) R^{(n)*}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega_{\rm c})\right]. \tag{S48}$$

Here, we iterate to the $N_{\rm dc}$-th iteration, until $|\Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in}, \omega)|$ and $|\Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \omega)|$ are both smaller than 10° for every pair of $(\mathbf{k}_{\parallel}^{\rm in}, \omega)$ and every pair of $(\mathbf{k}_{\parallel}^{\rm out}, \omega)$. The reflection matrix after the VRM dispersion compensation (dc) is then

$$R_{\rm dc}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) \equiv R^{(N_{\rm dc}+1)}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) = e^{-i\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out}, \omega)} R_{\max}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) e^{-i\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in}, \omega)}, \tag{S49}$$

where $\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in}, \omega) = \sum_{n=1}^{N_{\rm dc}} \Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in}, \omega)$ and $\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out}, \omega) = \sum_{n=1}^{N_{\rm dc}} \Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \omega)$. Note we can see from Eq. (S47)–(S48) that $\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in}, \omega_{\rm c}) = \phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out}, \omega_{\rm c}) = 0$ at the center frequency throughout the iterations.

These VRM dispersion compensation steps are very different from the SMT dispersion compensation steps described in Supplementary Sect. III E. To facilitate comparison and to gain insights into VRM, here we identify the objective functions that the VRM dispersion compensation implicitly maximizes. Define

$$M_{\rm in}^{\rm VRM}(\mathbf{k}_{\parallel}^{\rm in}, \omega) \equiv \sum_{\mathbf{r} \in z_{\max}} I_{\rm in}(\mathbf{r}, \mathbf{k}_{\parallel}^{\rm in}, \omega), \quad M_{\rm out}^{\rm VRM}(\mathbf{k}_{\parallel}^{\rm out}, \omega) \equiv \sum_{\mathbf{r} \in z_{\max}} I_{\rm out}(\mathbf{r}, \mathbf{k}_{\parallel}^{\rm out}, \omega), \tag{S50}$$

where $\sum_{\mathbf{r}\in z_{\max}}$ denotes $\mathbf{r} = (x, y, z_{\max})$ with summations over x and y in the FOV while fixing the depth at $z_{\max}$. Here,

$$I_{\text{in}}(\mathbf{r}, \mathbf{k}_{\parallel}^{\text{in}}, \omega') \equiv \left| \sum_{\omega=\omega',\omega_{\text{c}}} \sum_{\mathbf{k}_{\parallel}^{\text{out}}} e^{i(\mathbf{k}_{\text{out}}-\mathbf{k}_{\text{in}})\cdot\mathbf{r} - i\phi_{\text{out}}(\mathbf{k}_{\parallel}^{\text{out}},\omega) - i\phi_{\text{in}}(\mathbf{k}_{\parallel}^{\text{in}},\omega)} R(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{in}}, \omega) \right|^2, \tag{S51}$$

$$I_{\text{out}}(\mathbf{r}, \mathbf{k}_{\parallel}^{\text{out}}, \omega') \equiv \left| \sum_{\omega=\omega',\omega_{\text{c}}} \sum_{\mathbf{k}_{\parallel}^{\text{in}}} e^{i(\mathbf{k}_{\text{out}}-\mathbf{k}_{\text{in}})\cdot\mathbf{r} - i\phi_{\text{out}}(\mathbf{k}_{\parallel}^{\text{out}},\omega) - i\phi_{\text{in}}(\mathbf{k}_{\parallel}^{\text{in}},\omega)} R(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{in}}, \omega) \right|^2. \tag{S52}$$

By substituting Eq. (S44) into Eqs. (S50)–(S51), using Parseval's theorem, and taking the derivative, we see that

$$\frac{\partial M_{\text{in}}^{\text{VRM}}}{\partial \phi_{\text{in}}(\mathbf{k}_{\parallel}^{\text{in}}, \omega)} \propto \text{Im}\left[\sum_{\mathbf{k}_{\parallel}^{\text{out}}} R_{\text{dc}}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega) R_{\text{dc}}^{*}(\mathbf{k}_{\parallel}^{\text{out}}, \mathbf{k}_{\parallel}^{\text{in}}, \omega_{\text{c}}) \right]. \tag{S53}$$

When the VRM dispersion compensation converges, $\Delta\phi_{\text{in}}^{(N)}(\mathbf{k}_{\parallel}^{\text{in}}, \omega) \approx 0$, so the term inside the square brackets of Eq. (S47) is purely real, and $\frac{\partial M_{\text{in}}^{\text{VRM}}}{\partial \phi_{\text{in}}(\mathbf{k}_{\parallel}^{\text{in}}, \omega)} \approx 0$. We can similarly show that $\frac{\partial M_{\text{out}}^{\text{VRM}}}{\partial \phi_{\text{out}}(\mathbf{k}_{\parallel}^{\text{out}}, \omega)} \approx 0$ when VRM converges. Therefore, we conclude that the iterative phase conjugation of VRM dispersion compensation reaches a local maximum of $M_{\text{in}}^{\text{VRM}}(\mathbf{k}_{\parallel}^{\text{in}}, \omega)$ and $M_{\text{out}}^{\text{VRM}}(\mathbf{k}_{\parallel}^{\text{out}}, \omega)$ in Eq. (S50) for all $\mathbf{k}_{\parallel}^{\text{in}}, \mathbf{k}_{\parallel}^{\text{out}}, \omega$. Note this analysis is new and was not carried out in Ref. [35].

This analysis allows direct comparison between the VRM dispersion compensation above and the SMT dispersion compensation of Supplementary Sect. III E. We see four major differences:

1. The SMT image (I_{SMT}, defined in Eq. (2) of the main text) used in the SMT objective function of Eq. (S20) employs both input spatial gating (through a summation over $\mathbf{k}_{\text{in}}$) and output spatial gating (through a summation over $\mathbf{k}_{\text{out}}$). The confocal spatial gate is crucial for the optimization, as shown in **Supplementary Fig. 10**. Meanwhile, the I_{in} and I_{out} of Eqs. (S51)–(S52) used in the VRM dispersion compensation drops one of the two spatial gates, which reduces its ability to suppress multiple scattering and other undesired signals.

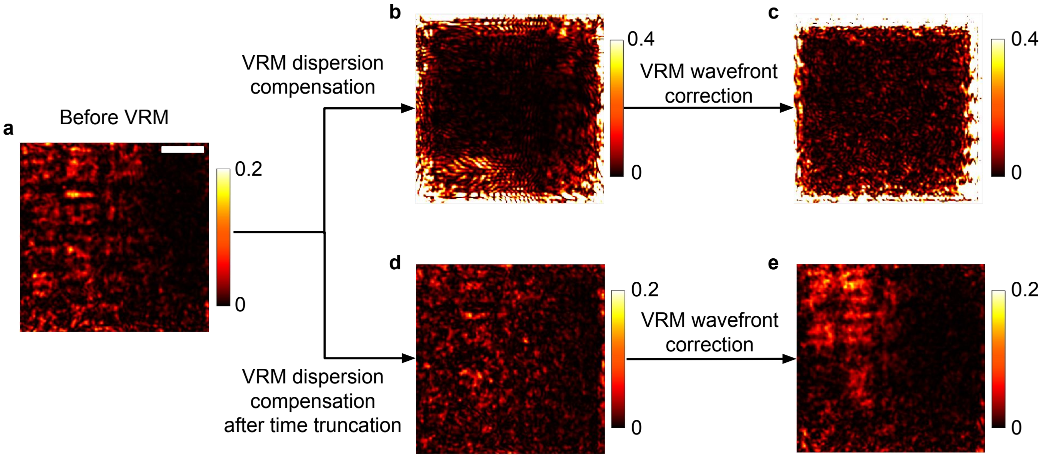
2. The SMT image I_{SMT} used in the SMT dispersion compensation also employees temporal gating through a summation over all frequencies ω. The temporal gate is crucial for the optimization, as shown in **Supplementary Fig. 10**. The VRM frequency summation in Eqs. (S51)–(S52) only sums over two frequencies, ω' and ω_{c}, so the temporal gate is gone, which further reduces its ability to suppress multiple scattering and other undesired signals.

3. The spatial summation of SMT dispersion compensation in Eq. (S20) sums over a 3D volume, which avoids overfitting. The spatial summation of VRM dispersion compensation in Eq. (S50) sums over a 2D slice at a single depth $z_{\max}$, which can create serious artifacts. This is the reason that the volumetric VRM image in Fig. 5 of the main text shows bright slices, with each slice being the $z_{\max}$ of that sub-volume.

4. To avoid overfitting and to alleviate the curse of dimensionality in an optimization problem, SMT employs only three variables for dispersion compensation: $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{a}_3$, each having a physical meaning (group velocity, group velocity dispersion, and asymmetric pulse distortion). Meanwhile, VRM uses $\phi_{\text{in}}(\mathbf{k}_{\parallel}^{\text{in}}, \omega)$ and $\phi_{\text{out}}(\mathbf{k}_{\parallel}^{\text{out}}, \omega)$ as free variables for every pair of $(\mathbf{k}_{\parallel}^{\text{in}}, \omega)$ and every pair of $(\mathbf{k}_{\parallel}^{\text{out}}, \omega)$.

Also, with the SMT dispersion compensation, we pre-compute the angular summations over $\mathbf{k}_{\text{in}}$ and $\mathbf{k}_{\text{out}}$ in Eq. (S19), which makes the SMT dispersion compensation much faster than the VRM dispersion compensation. For the nanoparticle colloid sample here, the SMT dispersion compensation took less than 10 minutes for the full volume (across the 110 µm depth of field), while the VRM dispersion compensation took around 2 hours per sub-volume (32 hours in total for the full volume).

In our tests, we find the VRM dispersion compensation to work reasonably well with high-contrast planar targets in a weakly scattering system, such as a USAF target in air or under a thin layer of tissue. However, for the strongly scattering systems we consider in the main text, the VRM dispersion compensation introduces very bright artifacts on the edges of the image (**Supplementary Fig. 19**b). We observe the same behavior with VRM when using a reflection matrix measured without any sample. This is because instead of maximizing the signal from the intended targets, VRM maximizes the signal from the objective lens reflection. Recall that after triple gating and after we exclude the peak of the objective lens signal from the reflection matrix data (Supplementary Sect. I C), the sensitivity of $I_{\text{SMT}}(\mathbf{r})$ is still limited by the residual reflection from the objective lens (Supplementary Sect. II C). Since the objective lens lies at depths very far away from the sample (but still within the coherence length of the laser), the



Supplementary Fig. 19. **Evolution of the VRM image of the USAF target under tissue without and with temporal truncation.** Scale bar: 10 µm.

most effective way to filter out the objective lens reflection is time gating. But the $I_{\rm in}(\mathbf{r}, \mathbf{k}_\parallel^{\rm in}, \omega')$ and $I_{\rm out}(\mathbf{r}, \mathbf{k}_\parallel^{\rm out}, \omega')$ that the VRM dispersion compensation maximizes only use two frequencies (ω' and $\omega_{\rm c}$) with no time gating, so the objective lens signal dominates over the weak target signal in $I_{\rm in}(\mathbf{r}, \mathbf{k}_\parallel^{\rm in}, \omega')$ and $I_{\rm out}(\mathbf{r}, \mathbf{k}_\parallel^{\rm out}, \omega')$ and is maximized by VRM. This may not be as much of an issue in Ref. [35], which used lasers with much shorter coherence lengths and considered samples dominated by the aberrations (namely, single scattering) of a thin plastic layer rather than the multiple scattering from a volumetric medium. To improve the VRM performance, here we suppress the objective lens signal by incorporating a coarse temporal gate prior to the VRM dispersion compensation. We do so by (1) shifting the reference plane of the reflection matrix to the center of a truncation volume [same as Eq. (S44) but to the center of the truncation volume instead], (2) perform a non-uniform discrete Fourier transform to convert the reflection matrix data from frequency domain ω to time domain t, (3) perform another non-uniform discrete Fourier transform to convert from time domain t back to frequency domain ω, and (4) shift the reference plane of the reflection matrix back. In time domain, we only use the values of t within the truncation volume, which sets the coarse temporal gate. As shown in **Supplementary Fig. 19**d, this "temporal truncation" suppresses the objective-lens contributions in the reflection matrix data and prevents the VRM dispersion compensation from maximizing the objective lens signal. However, the performance of VRM is still substantially worse than SMT due to the four differences described above.

For the 3D nanoparticle colloid sample, we perform the first SMT dispersion compensation simultaneously over the full 110 µm depth of field, $z \in [1470, 1580]$ µm, to avoid overfitting. However, simultaneous dispersion compensation over such a large volume is not possible with VRM: when the temporal truncation spans such a large volume, there is not enough suppression of the objective lens signal (since the VRM dispersion compensation itself has no temporal gate), and the VRM dispersion compensation always maximizes the objective lens signal instead. So, here we divide the 110 µm depth of field into 16 overlapping sub-volumes (same as in the SMT wavefront correction) and perform VRM separately for each sub-volume (both dispersion compensation and wavefront correction), each with its temporal truncation. The final images of the sub-volumes are then stitched together, same as in SMT.

2. VRM angular wavefront correction and reconstruction

After estimating the spectro-angular dispersion relative to the center frequency $\omega_{\rm c}$, VRM performs another iterative phase conjugation to estimate the angular wavefront distortion $\phi_{\rm in}(\mathbf{k}_\parallel^{\rm in}, \omega_{\rm c})$ and $\phi_{\rm out}(\mathbf{k}_\parallel^{\rm out}, \omega_{\rm c})$ at $\omega_{\rm c}$, and then adds the same wavefront correction of $-\phi_{\rm in}(\mathbf{k}_\parallel^{\rm in}, \omega_{\rm c})$ and $-\phi_{\rm out}(\mathbf{k}_\parallel^{\rm out}, \omega_{\rm c})$ to all frequencies.

This wavefront correction procedure is identical to CLASS [26, 28] except that VRM builds the coherence-gated matrices from frequency domain while generalizing CLASS to include multiple depth slices (referred to as "volumetric dispersion correction" in Ref. [35]). It starts with selecting J depth slices $\{z_1, \ldots, z_J\}$ and building the coherence-gated (cg) reflection matrices at these depths,

$$\tilde{R}_{\rm cg}(\mathbf{k}_\parallel^{\rm out}, \mathbf{k}_\parallel^{\rm in}; z_j) = \sum_{\omega} R_{\rm dc}(\mathbf{k}_\parallel^{\rm out}, \mathbf{k}_\parallel^{\rm in}, \omega) e^{i(k_z^{\rm out} - k_z^{\rm in})(z_j - z_{\rm max})}, \tag{S54}$$

with the dispersion-compensated reflection matrix $R_{\rm dc}(\mathbf{k}_{||}^{\rm out}, \mathbf{k}_{||}^{\rm in}, \omega)$ from Eq. (S49). Another iterative phase conjugation follows, as

$$\tilde{R}_{{\rm cg},j}^{(n+1)}(\mathbf{k}_{||}^{\rm out}, \mathbf{k}_{||}^{\rm in}) = e^{-i\Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out})} \tilde{R}_{{\rm cg},j}^{(n)}(\mathbf{k}_{||}^{\rm out}, \mathbf{k}_{||}^{\rm in}) e^{-i\Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in})}, \tag{S55}$$

with $n = 0, 1, 2, \ldots$ and with initial condition $\tilde{R}_{{\rm cg},j}^{(0)}(\mathbf{k}_{||}^{\rm out}, \mathbf{k}_{||}^{\rm in}) = \tilde{R}_{\rm cg}(\mathbf{k}_{||}^{\rm out}, \mathbf{k}_{||}^{\rm in}; z_j)$, $\Delta\phi_{\rm in}^{(0)}(\mathbf{k}_{\parallel}^{\rm in}) = \Delta\phi_{\rm out}^{(0)}(\mathbf{k}_{\parallel}^{\rm out}) = 0$. The correction at the n-th iteration is determined by phase conjugation,

$$\Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in}) = \arg\left[\sum_{j,\Delta\mathbf{k}_{\parallel}} \tilde{R}_{{\rm cg},j}^{(n)}(\mathbf{k}_{\parallel}^{\rm in} + \Delta\mathbf{k}_{\parallel}, \mathbf{k}_{\parallel}^{\rm in}) \tilde{R}_{{\rm ccg},j}^{(n)*}(\Delta\mathbf{k}_{\parallel})\right], \tag{S56}$$

$$\Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out}) = \arg\left[\sum_{j,\Delta\mathbf{k}_{\parallel}} \tilde{R}_{{\rm cg},j}^{(n)}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm out} - \Delta\mathbf{k}_{\parallel}) \tilde{R}_{{\rm ccg},j}^{(n)*}(\Delta\mathbf{k}_{\parallel})\right], \tag{S57}$$

where

$$\tilde{R}_{{\rm ccg},j}^{(n)}(\Delta\mathbf{k}_{\parallel}) = \sum_{\mathbf{k}_{\parallel}^{\rm in}} \tilde{R}_{{\rm cg},j}^{(n)}(\mathbf{k}_{\parallel}^{\rm in} + \Delta\mathbf{k}_{\parallel}, \mathbf{k}_{\parallel}^{\rm in}) \tag{S58}$$

is the confocal and coherence-gated (ccg) reflection matrix at depth z_j. Here, we iterate to the $N_{\rm wc}$-th iteration, until $|\Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in})|$ and $|\Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out})|$ are both smaller than 10° for every $\mathbf{k}_{\parallel}^{\rm in}$ and every $\mathbf{k}_{\parallel}^{\rm out}$.

Ref. [35] only reported images at these individual slices $\{z_j\}$, obtained by taking the inverse Fourier transform of the final confocal and coherence-gated reflection matrices similar to Eq. (S37),

$$I_{\rm VRM}(\mathbf{r}_{\parallel}, z_j) = \left|\sum_{\Delta\mathbf{k}_{\parallel}} e^{i\Delta\mathbf{k}_{\parallel}\cdot\mathbf{r}_{\parallel}} \tilde{R}_{{\rm ccg},j}^{(N_{\rm wc}+1)}(\Delta\mathbf{k}_{\parallel})\right|^2. \tag{S59}$$

In order to examine how VRM works for volumetric reconstruction, here we generalize Eq. (S59) to

$$I_{\rm VRM}(\mathbf{r}) = \left|\sum_{\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega} e^{i(\mathbf{k}_{\rm out} - \mathbf{k}_{\rm in})\cdot\mathbf{r}} R_{\rm VRM}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega)\right|^2, \tag{S60}$$

with the wavefront-corrected (wc) reflection matrix

$$R_{\rm VRM}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) \equiv e^{-i\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out}) - i\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in}) - i(k_z^{\rm out} - k_z^{\rm in})z_{\rm max}} R_{\rm dc}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega), \tag{S61}$$

where $\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in}) = \sum_{n=1}^{N_{\rm wc}} \Delta\phi_{\rm in}^{(n)}(\mathbf{k}_{\parallel}^{\rm in})$ and $\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out}) = \sum_{n=1}^{N_{\rm wc}} \Delta\phi_{\rm out}^{(n)}(\mathbf{k}_{\parallel}^{\rm out})$, with $R_{\rm dc}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega)$ from Eq. (S49). Note that Eq. (S60) was not derived and not used in Ref. [35], but it mathematically reduces to Eq. (S59) when $z = z_j$, so we use it to extend the formalism of Ref. [35] from individual slices to volumetric reconstruction. This $I_{\rm VRM}(\mathbf{r})$ in Eq. (S60) is the VRM image we build.

To facilitate comparison with SMT, here we carry out an analysis similar to that in the dispersion compensation section above. The objective function that the VRM wavefront correction implicitly maximizes is

$$M_{\rm wc}^{\rm VRM} \equiv \sum_{\mathbf{r}\in\{z_j\}} I_{\rm VRM}(\mathbf{r}), \tag{S62}$$

where $\sum_{\mathbf{r}\in\{z_j\}}$ denotes $\mathbf{r} = (x, y, z)$ with summations over x and y in the FOV and summations over z at the chosen J depth slices $\{z_1, \ldots, z_J\}$. We can show that

$$\frac{\partial M_{\rm wc}^{\rm VRM}}{\partial\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in})} \propto {\rm Im}\left[\sum_{j,\Delta\mathbf{k}_{\parallel}} \tilde{R}_{{\rm cg},j}(\mathbf{k}_{\parallel}^{\rm in} + \Delta\mathbf{k}_{\parallel}, \mathbf{k}_{\parallel}^{\rm in}) \tilde{R}_{{\rm ccg},j}^{*}(\Delta\mathbf{k}_{\parallel})\right]. \tag{S63}$$

When the VRM wavefront correction converges, $\Delta\phi_{\rm in}^{(N)}(\mathbf{k}_{\parallel}^{\rm in}) \approx 0$, so the term inside the square brackets of Eq. (S56) is purely real, and $\frac{\partial M_{\rm wc}^{\rm VRM}}{\partial\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in})} \approx 0$. We can similarly show that $\frac{\partial M_{\rm wc}^{\rm VRM}}{\partial\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out})} \approx 0$ when the VRM wavefront correction converges. Therefore, we conclude that the iterative phase conjugation of VRM wavefront correction (and similarly of CLASS) reaches a local maximum of $M_{\rm wc}^{\rm VRM}$ in Eq. (S62).

This analysis allows direct comparison between the VRM wavefront correction above and the SMT wavefront correction of Supplementary Sect. III F. We see four major differences:

1. To avoid overfitting, SMT regularize the problem by expanding $\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in})$ and $\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out})$ in Zernike polynomials and only using the Zernike coefficients as the free variables. In CLASS and VRM, every $\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in})$ and every $\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out})$ is a free variable.

2. To avoid getting trapped in a poor local optimum, SMT first optimizes $\phi_{\rm in/out}$ over the full volume and then progressively optimizes over smaller zones while using the previous $\phi_{\rm in/out}$ as the initial guess. There is no such progression in CLASS and VRM. As shown in Fig. 2c–f of the main text, optimizing the wavefront inside smaller zones without regularization and progression can result in an image that is brighter but with worse quality.

3. SMT also increases the number of Zernike terms when progressing to smaller zones to account for the relation between isoplanatic patch size and the smoothness of wavefront variation. This is not possible in CLASS and VRM.

4. SMT starts from an initial guess of no wavefront correction, $\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in}) = \phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out}) = 0$. VRM starts with the $\phi_{\rm in}(\mathbf{k}_{\parallel}^{\rm in}, \omega)$ and $\phi_{\rm out}(\mathbf{k}_{\parallel}^{\rm out}, \omega)$ from its dispersion compensation step.

While the VRM wavefront correction uses multiple slices, it cannot reverse the over-optimization of the intensity at the single slice $z_{\rm max}$ during the VRM dispersion compensation because the wavefront correction does not modify the temporal gate.

The formalism above starts from the reflection matrix $R(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega)$ of Eq. (S18b), which has corrected for the refractive index mismatch at interfaces. This is what we use for the 3D nanoparticle colloid sample in Fig. 5 of the main text in order to compare the SMT and VRM images at the same depths.

The VRM formalism from Ref. [35] does not include index-mismatch correction, so we do not include such correction for the 2D USAF target under tissue sample in Fig. 4 of the main text, for which quantitative values of z is not necessary. In this case, we build the VRM image with

$$I_{\rm VRM}(\mathbf{r}') = \left| \sum_{\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega} e^{-2i\omega\Delta t' + i(\mathbf{k}_{\rm out}^{\rm air} - \mathbf{k}_{\rm in}^{\rm air})\cdot \mathbf{r}'} R'_{\rm VRM}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega) \right|^2, \tag{S64}$$

with $\mathbf{r}' = (x, y, z')$, $z' = z^{\rm air} - z_{\rm ref}^{\rm air}$ the apparent depth of the spatial gate relative to the in-air reference plane $z_{\rm ref}^{\rm air}$, and with $R'_{\rm VRM}(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega)$ built from the reflection matrix $R'(\mathbf{k}_{\parallel}^{\rm out}, \mathbf{k}_{\parallel}^{\rm in}, \omega)$ without index-mismatch correction. Here, $\Delta t' = [n_{\rm glass} h^{\rm glass} + n_{\rm sam} z - (h^{\rm glass} + z^{\rm air})]/c$ is the additional travel time to propagate to z in the sample relative to propagating to the apparent depth $z^{\rm air}$ in air, which aligns the temporal gate with the spatial gate in the presence of index mismatch (also considered in Supplementary Sect. 6 of Ref. [61]). For this 2D target, we build the VRM image at $z' = z'_{\rm target}$, the depth that maximizes the image intensity integrated over x-y.

In terms of computational cost, although VRM or CLASS implements phase conjugation rather than gradient-based optimization, its wavefront correction for the whole FOV still takes ≈1 minute to complete due to the large data size. Note that VRM and SMT share similar computational overhead in the reconstruction of the time-gated reflection matrices and the precomputations. Thus, when performing wavefront correction for only the whole FOV instead of each zone, SMT's computational time and memory consumption in wavefront correction is comparable to VRM's. As SMT goes through multiple progressions to correct each zone, enabling superior correction of strong distortions, it ultimately takes more time and memory.

3. *VRM/CLASS and the optical reciprocity principle*

Having examined the implementation details of VRM/CLASS, it is instructive to understand these methods from the perspective of optical reciprocity. The optical reciprocity principle, also known as the Lorentz reciprocity or the Helmholtz reciprocity, states that if a source at point A creates a field at point B, then the same source at B will create an identical field at A. In imaging applications, this principle is often implemented through phase conjugation, where a time-reversed wave can exactly retrace its path through the optical system, effectively canceling wavefront distortions. This principle underlies various wavefront correction methods, including guide-star-based adaptive optics and guide-star-based wavefront shaping where conjugating the wavefront coming out from the guide-star, such as a fluorescence bead, forms a tight focus at the guide-star location, enhancing the image quality in that region.

We can understand how VRM/CLASS implements this principle by examining its matrix operations. In VRM/CLASS, the scattered wavefront carries phase distortions $-\phi_{\text{out}}(\mathbf{k}_{\text{out}})$. Multiplying $e^{i\phi_{\text{out}}(\mathbf{k}_{\text{out}})}$ to the corresponding matrix rows to correct aberrations is computationally equivalent to forming a phase-conjugated wavefront at the output pupil plane. Transposing the matrix interchanges inputs and outputs, meaning that the phase-conjugated wavefront becomes the incident wavefront and is sent back to the sample to form focused/' spots. After the transpose, the inputs become outputs and vice versa, with $-\phi_{\text{in}}(\mathbf{k}_{\text{in}})$ becoming the scattered wavefront. Multiplying $e^{i\phi_{\text{in}}(\mathbf{k}_{\text{in}})}$ to the corresponding rows of the transposed matrix (columns of the non-transposed matrix) to correct aberrations computationally forms a conjugate wavefront at the output pupil plane (previously the input pupil plane before the transpose). Another second transpose sends this conjugate wavefront back to the sample to focus inside it and improve the image. This process carries on until the image can no longer be improved. Through these iterative operations, VRM/CLASS employs the optical reciprocity principle to iteratively improve image quality by computationally implementing wavefront conjugation and propagation.

E. Eigenchannel-based approach

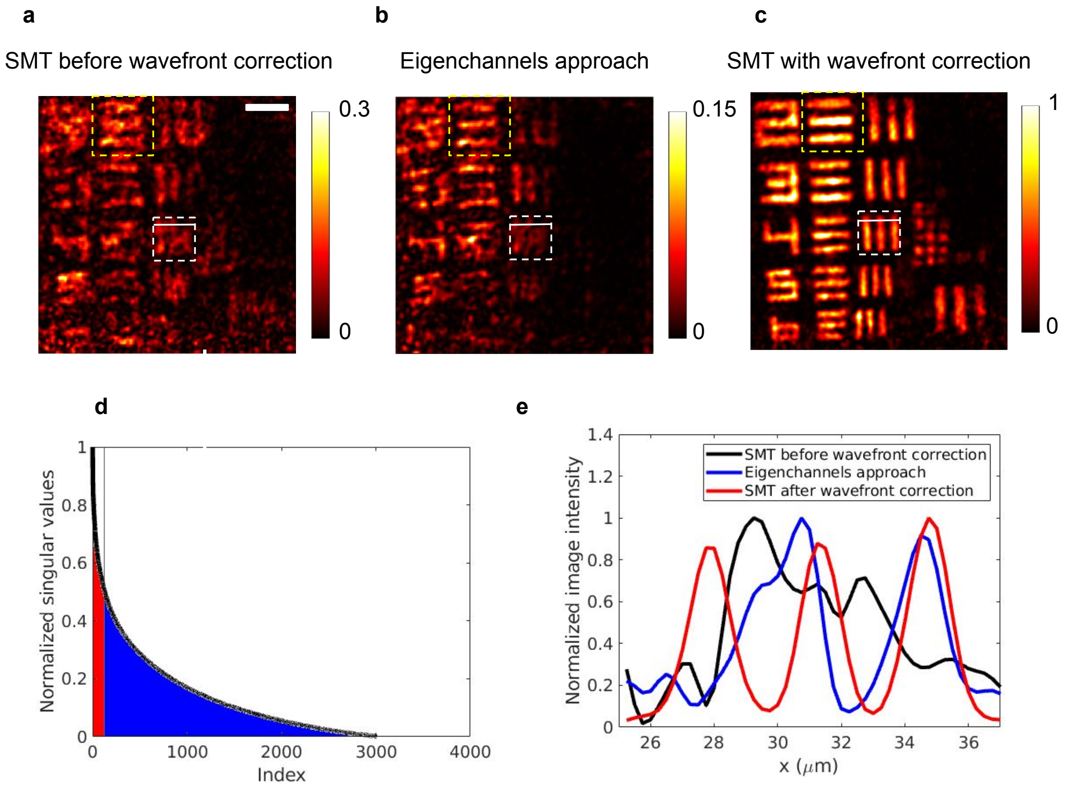
The eigenchannel-based approach has emerged as powerful tools to suppress multiple scattering, leading to weaker speckles on the image [23, 55, 56]. This method employs singular value decomposition (SVD) to decompose the scattering matrix into the product of three matrices $\mathbf{U\Sigma V}^\dagger$, where $\mathbf{U}$ and $\mathbf{V}$ are two unitary matrices whose each row contains the output and input eigenchannels, $\mathbf{\Sigma}$ is a diagonal matrix where each diagonal element (singular value) represents the weight of the corresponding input and output eigenchannels. When the scattered signal is not overwhelmingly distorted by strong aberrations and scattering, the large singular values correspond to single scattering eigenchannels while smaller ones correspond to multiple scattering eigenchannels. To suppress the multiple scattering, singular values smaller than a certain threshold are set to zeros. The image will then be found from the diagonal of the new matrix with suppressed multiple scattering. Here, we apply this approach to the time-gated reflection matrix in the spatial basis $\bar{R}(\mathbf{r}_\parallel^{\text{out}}, \mathbf{r}_\parallel^{\text{in}}, z_{\text{target}})$. We set singular values smaller than 50% of the maximum to zero, with this threshold selected to yield the best image visually.

However, there is not a strict definition to this threshold. Rather, it varies from case to case. Therefore, there is a chance that after thresholding, not only the multiple scattering is suppressed but also some of the weaker single scattering eigenchannels. In addition, as the eigenchannels approach only suppresses multiple scattering, it does not correct for aberrations. Therefore, after being processed with the eigenchannels approach, the image may have weaker multiple scattering speckles but it will still be distorted due to the uncorrected aberrations.

Supplementary Fig. 20 shows a comparison between SMT and the eigenchannels approach. In **Supplementary Fig. 20**(b), inside the yellow dashed region, we notice the image becomes better as the three horizontal bars become more well-separated after suppressing the multiple scattering eigenchannels, although the still worse than the SMT-with-wavefront-correction (**Supplementary Fig. 20**(c)). However, in many other parts of the image, notably the region marked by the white dashed box, eigenchannels approach fails to produce any visible improvement due to the lack of an effective wavefront correction. **Supplementary Fig. 20**(e) shows a quantitative comparison between SMT and the eigenchannels approach. The target along the white solid line in **Supplementary Fig. 20**(a-c) contains three vertical lines with a 3-µm separation, which are only separated in the SMT image after wavefront correction. In the 1D profile along this line, only the SMT with correction shows three discernible high peaks. On the other hand, the eigenchannels approach reveals only one broad peak and another narrower one. This result confirms a better performance of SMT over the eigenchannels approach. It is safe to say SMT can image deeper due to the powerful aberration corrections, as SMT reveals all targets while the eigenchannels approach can only reveal some of them with worse quality. SMT also achieves better resolution as the eigenchannels cannot separate the two vertical bars along the white solid line, meaning that its resolution is worse than 3 µm. Meanwhile, SMT clearly separates all bars and as the smallest elements of the target have a 1.1 µm separation. The resolution of SMT is thus equivalent to or better than 1.1 µm. Eigenchannels approach, nevertheless, still beats SMT in terms of computational complexity, as only one SVD is involved, which took less than 1 second to complete in our case, compared to several minutes in SMT.

V. SMT WITH REDUCED GATING

As shown in **Supplementary Fig. 10**, triple gating is important for the SMT optimization. But reducing the gating strength does have the benefit of requiring less matrix data to be measured, which can speed up the measurement time. In this section, we explore to what extent the gating can be reduced.



Supplementary Fig. 20. (a) SMT without wavefront correction. (b) Image with eigenchannels approach. (c) SMT with wavefront correction. In (a-c), the yellow dashed box marks the region where eigenchannels approach improves the image while the white dashed box marks one of many regions where it fails to produce visible improvement. (d) The normalized singular values of the time-gated reflection matrix. The vertical line separates the single and multiple scattering. The red-colored singular values on the left of the line correspond to single scattering eigenchannels and are kept. The blue-colored singular values on the right of the line are all set to zeros. (e) A comparison between the normalized image intensity of the three images in (a-c) along the white solid line. Each 1D profile is normalized by its own maximum value. Scale bar: 10 µm.

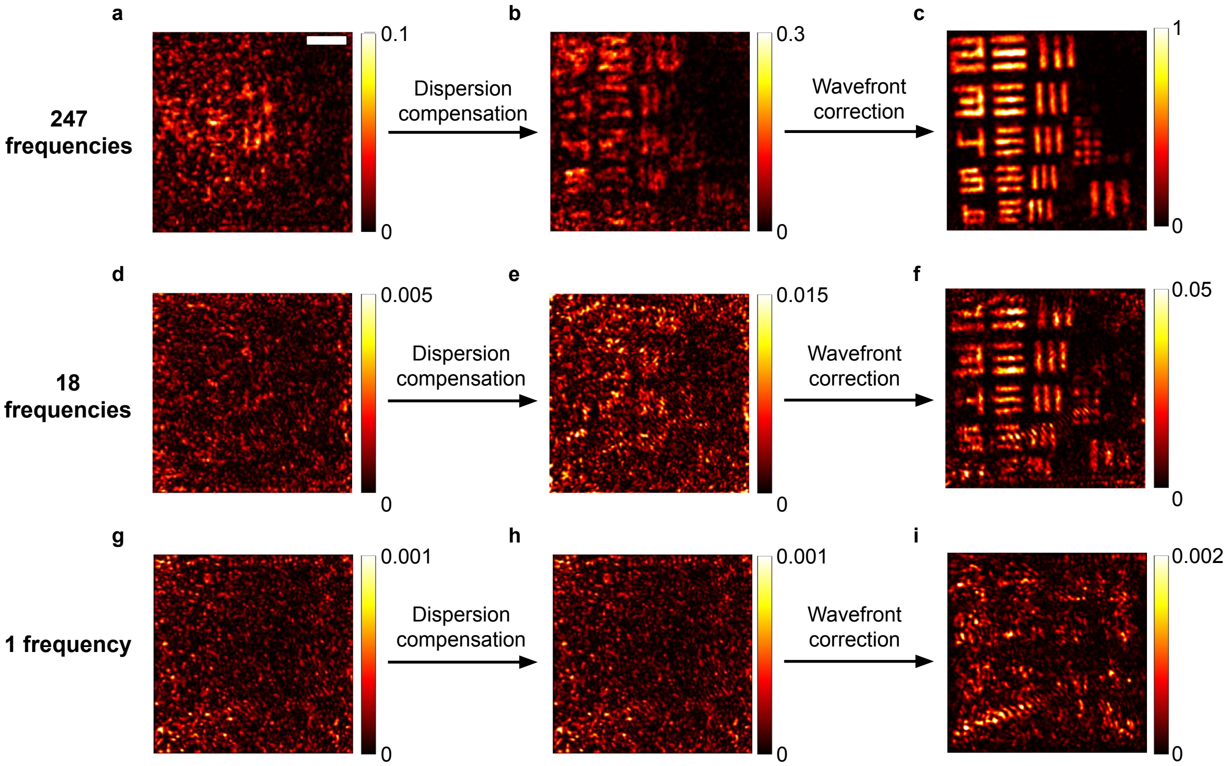
A. SMT with fewer frequencies

We start with time gating. **Supplementary Fig. 21** shows the SMT images of the USAF-target-under-tissue sample when we keep all 247 frequencies (top row), when we reduce the temporal gate by using only 18 frequencies while keeping the 741–899 nm wavelength range (middle row), and when we completely remove the temporal gate by only using the central frequency (bottom row). SMT still works quite well even when the number of frequencies is reduced by over an order of magnitude. But it fails when the temporal gate is completely removed.

B. SMT with fewer incident angles

Next, we consider spatial gating. **Supplementary Fig. 22** shows the SMT images of the USAF-target-under-tissue sample when we keep all 3761 incident angles (top row), when we reduce the input spatial gate by reducing the input NA from $\mathrm{NA}_{\mathrm{in}} = 0.5$ to $\mathrm{NA}_{\mathrm{in}} = 0.15$, corresponding to 341 incident angles at the highest frequency (middle row), and when we completely remove the input spatial gate by only using one incident angle at normal incidence (bottom row), corresponding to $\mathrm{NA}_{\mathrm{in}} = 0$. The last case, **Supplementary Fig. 22**g–i, is also shown as Fig. 5a–c in the main text. When the input spatial gate is fully removed, the image reconstruction fails even after the optimizations.

Full-field OCT corrected by computational adaptive optics (CAO-FF-OCT) [18, 89] also has no input spatial gate; their output spatial gate is also weaker due to the lower NA used. Previously, CAO-FF-OCT was applied to correct for the aberrations from a plastic layer and thin film of milk [89] and from the retina [18]. The results in **Supplementary Fig. 22**g–i indicate that CAO-FF-OCT would not work in a strongly scattering medium even after increasing the NA and after adopting more advanced optimization strategies because it does not have a confocal spatial gate.



Supplementary Fig. 21. SMT images of the USAF-target-under-tissue sample using all 247 frequencies measured (top row), using 18 frequencies across the full bandwidth (middle row), and using only 1 frequency (bottom row). Scale bar: 10 µm.

VI. POINT SPREAD FUNCTION (PSF)

To evaluate the point spread function (PSF) of the USAF-target-under-tissue sample, we fix the input at $\mathbf{r}_{\text{in}}$ while varying the position of the output $\mathbf{r}$. Specifically, for SMT, we generalize Eq. (2) of the main text to get

$$\text{PSF}_{\text{SMT}}(\mathbf{r}) = \left| \sum_{\omega} e^{i\theta(\omega)} \sum_{\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.5} e^{i[\mathbf{k}_{\text{out}}^{\text{sam}}\cdot\mathbf{r}-\mathbf{k}_{\text{in}}^{\text{sam}}\cdot\mathbf{r}_{\text{in}}+\phi_{\text{out}}(\mathbf{k}_{\parallel}^{\text{out}})+\phi_{\text{in}}(\mathbf{k}_{\parallel}^{\text{in}})]} R(\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}},\omega) \right|^2, \tag{S65}$$

with $\mathbf{r} = (x, y, z_{\text{target}})$ and $\mathbf{r}_{\text{in}} = (x_{\text{in}}, y_{\text{in}}, z_{\text{target}})$. For the other methods, we generalize Eqs. (S31), (S33), (S35), (S64) to get

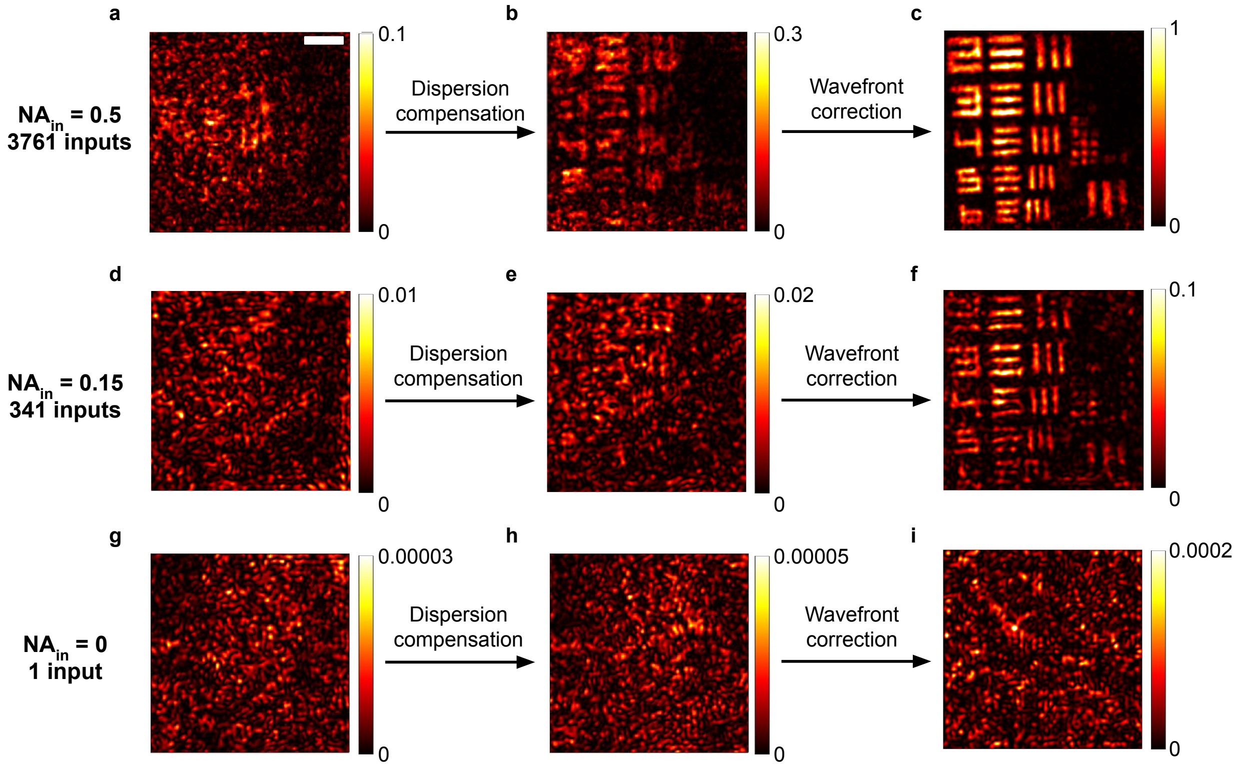
$$\text{PSF}_{\text{RCM}}(\mathbf{r}') = \left| \sum_{\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.5} e^{i(\mathbf{k}_{\text{out}}^{\text{air}}\cdot\mathbf{r}'-\mathbf{k}_{\text{in}}^{\text{air}}\cdot\mathbf{r}'_{\text{in}})} R'(\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}},\omega=\omega_{\text{RCM}}) \right|^2, \tag{S66}$$

$$\text{PSF}_{\text{OCM}}(\mathbf{r}') = \left| \sum_{\omega} e^{i\theta_{\text{OCM}}(\omega)-2i\omega\Delta t'_{\text{f}}} \sum_{\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.5} e^{i(\mathbf{k}_{\text{out}}^{\text{air}}\cdot\mathbf{r}'-\mathbf{k}_{\text{in}}^{\text{air}}\cdot\mathbf{r}'_{\text{in}})} R'(\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}},\omega) \right|^2, \tag{S67}$$

$$\text{PSF}_{\text{OCT}}(\mathbf{r}') = \left| \sum_{\omega} e^{i\theta_{\text{OCT}}(\omega)-i\omega\Delta t'_{\text{f}}(\frac{1}{\cos\alpha_{\text{in}}}+\frac{1}{\cos\alpha_{\text{out}}})} \sum_{\mathbf{k}_{\parallel}^{\text{out}}}^{\text{NA}=0.1,\alpha_{\text{out}}} \sum_{\mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.1,\alpha_{\text{in}}} e^{i(\mathbf{k}_{\text{out}}^{\text{air}}\cdot\mathbf{r}'-\mathbf{k}_{\text{in}}^{\text{air}}\cdot\mathbf{r}'_{\text{in}})} R'(\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}},\omega) \right|^2, \tag{S68}$$

$$\text{PSF}_{\text{VRM}}(\mathbf{r}') = \left| \sum_{\omega} e^{-2i\omega\Delta t'} \sum_{\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}}}^{\text{NA}=0.5} e^{i(\mathbf{k}_{\text{out}}^{\text{air}}\cdot\mathbf{r}'-\mathbf{k}_{\text{in}}^{\text{air}}\cdot\mathbf{r}'_{\text{in}})} R'_{\text{VRM}}(\mathbf{k}_{\parallel}^{\text{out}},\mathbf{k}_{\parallel}^{\text{in}},\omega) \right|^2, \tag{S69}$$

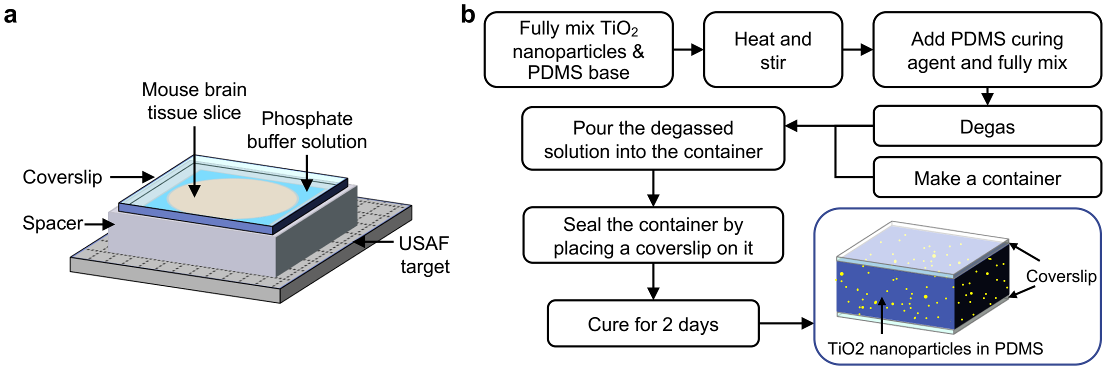
with $\mathbf{r}' = (x, y, z'_{\text{target}})$ and $\mathbf{r}'_{\text{in}} = (x_{\text{in}}, y_{\text{in}}, z'_{\text{target}})$.



Supplementary Fig. 22. SMT images using $\mathrm{NA_{in}} = 0.5$ corresponding to 3761 incident angles (top row), using $\mathrm{NA_{in}} = 0.15$ corresponding to 750 incident angles (middle row), and using $\mathrm{NA_{in}} = 0$ corresponding to only 1 incident angle (bottom row). Scale bar: 10 µm.

VII. SAMPLE PREPARATION

For imaging through tissue, we excised a slice from a fresh CD-1 adult mouse brain that is rinsed of blood (BioChemed Mouse tissue Brain) and placed the tissue slice on top of a USAF target (PhotomaskPORTAL Resolution target RTUSAF519P), immersed in a phosphate buffer solution. The specimen was sealed by a surrounding spacer and a coverslip on top (**Supplementary Fig. 23**a).



Supplementary Fig. 23. **Sample preparation.** **a**, A schematic of the sample with a USAF target beneath mouse brain tissue. The tissue slice is immersed in phosphate buffer solution and sealed by a spacer and a coverslip. **b**, Preparation of the $\mathrm{TiO_2}$ nanoparticle colloid sample.

The nanoparticle colloid sample consists of a mixture of polydimethylsiloxane (PDMS) silicone (MicroLubrol SYL-CAP 284-S) and titanium dioxide ($\mathrm{TiO_2}$) nanoparticles with a mean size of 500-nm diameter. We first dilute the $\mathrm{TiO_2}$ solution (Sigma-Aldrich Titania nanoparticles 914320) with ethanol and mix it with the PDMS base material.

The mixture is stirred and heated at 150 ℃ on a stirring hot plate until it is fully mixed and water and ethanol have evaporated. Then, we add the curing agent of PDMS into the mixture and stir at room temperature. After a thorough mix, the solution is placed into a vacuum chamber for degassing. We make a container by placing tapes (thickness of 2-3 mm) on a coverslip and then pour the degassed solution into the container. The sample is sealed by another coverslip on top and cured for 2 days at room temperature (**Supplementary Fig. 23**b).

The $\mathrm{TiO_2}$ mass concentration of the sample is 1.8 ppm which gives a volume particle number density of $n_v = 6.35 \times 10^{-3}$ µm^{-3} with the material density of $\rho_{\mathrm{TiO_2}} = 4.23$ g/cm^3 [90] and $\rho_{\mathrm{PDMS}} = 0.965$ g/cm^3 [91]. Mie theory [92] yields a scattering cross section of $\sigma = 0.839$ µm^2 and anisotropy factor of $g = 0.599$ for 500-nm particle diameter and wavelength $\lambda = 840$ nm, with the particle and the background refractive indices being 2.51 [93] and 1.41 [84]. The volume fraction of $\mathrm{TiO_2}$ is 0.04%, sufficiently low that spatial correlation effects [94] are expected to be negligible, so we use the independent-particle approximation [95] to obtain the scattering mean free path as $\ell_{\mathrm{sca}} = 1/(n_v\sigma) = 0.19$ mm and the transport mean free path as $\ell_{\mathrm{tr}} = 1/[n_v\sigma(1-g)] = 0.47$ mm.

VIII. RESOLUTION AND DEPTH OF FIELD OF SMT

In this section, we consider the resolution and depth of field (DOF) for the nanoparticle colloid sample, with theoretical estimates and a statistical analysis on the experimental data.

The simplest form of SMT is $\psi_{\mathrm{SMT}}(\mathbf{r}) = \sum_{\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega} e^{i(\mathbf{k}_{\mathrm{out}}-\mathbf{k}_{\mathrm{in}})\cdot\mathbf{r}} S(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$. Here, we can see that the spatial dependence comes from $e^{i(\mathbf{k}_{\mathrm{out}}-\mathbf{k}_{\mathrm{in}})\cdot\mathbf{r}}$, so the spread of $\mathbf{k}_{\mathrm{out}}-\mathbf{k}_{\mathrm{in}}$ in the summation determines the spatial resolution. When the FOV of illumination and detection is infinite, the spread of $\mathbf{k}_{\mathrm{out}}-\mathbf{k}_{\mathrm{in}}$ will be the same everywhere, giving SMT a location-independent resolution and an infinite DOF. In practice, the FOV is finite, so the large-angle beams will cease to illuminate the FOV of interest and/or cease to be detected at depths z sufficiently far from the reference plane under which the reflection matrix was measured. This reduces the lateral resolution and the signal strength away from the reflection-matrix reference plane, giving SMT a finite (but still large) DOF. Roughly speaking, the SMT DOF covers the volume of overlap among the input/output beams in the scattering matrix data.

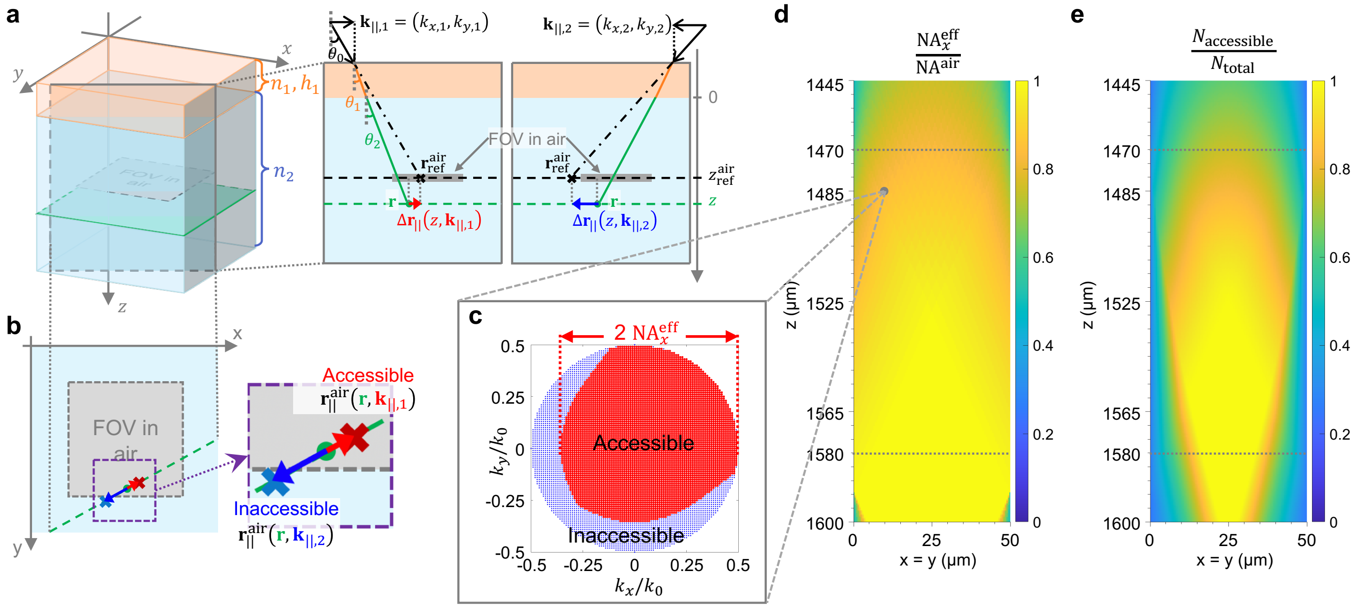
A. Theoretical axial resolution

The frequency summation and the angular summations both contribute to the axial spread of $\mathbf{k}_{\mathrm{out}}-\mathbf{k}_{\mathrm{in}}$. In other words, the temporal gate and the spatial gate both set the axial resolution, with the tighter gate being the main factor. The maximal axial spread of $\mathbf{k}_{\mathrm{out}}-\mathbf{k}_{\mathrm{in}}$ comes from small incident angles, so it is roughly depth independent. For the nanoparticle colloid sample, the expected diffraction-limited axial resolution from the spatial gate is approximately $2z_{\mathrm{R}} = 2n_{\mathrm{sam}}\lambda_0/(\pi\mathrm{NA}^2) \approx 2.99$ µm, with center wavelength $\lambda_0 = 840$ nm and $n_{\mathrm{sam}} = 1.4$ for PDMS. Meanwhile, the expected bandwidth-limited FWHM axial resolution from the temporal gate is $\delta z = \gamma\frac{2\ln 2}{n_{\mathrm{sam}}\pi}\frac{\lambda_0^2}{\Delta\lambda} = 1.42$ µm, where $\Delta\lambda = 187$ nm is the bandwidth and $\gamma = 1.19$ for a flat-top spectrum [96]. The latter is tighter and determines the axial resolution here. Dispersion, aberration, and scattering degrade the resolution, but the dispersion compensation and wavefront correction of SMT can restore the resolution to near the ideal value.

B. Theoretical lateral resolution and depth of field

For the lateral resolution and the DOF, we analyze the range of accessible angles at point $\mathbf{r}$ in the sample, using geometric optics. The range of accessible angles sets the lateral resolution through the lateral spread of $\mathbf{k}_{\mathrm{out}}-\mathbf{k}_{\mathrm{in}}$, and it also provides the signal strength at $\mathbf{r}$. In our setup (Supplementary Sect. I A), the camera sensor plane is conjugate to the focal plane of the objective lens, so the reference planes for the illumination and detection are both at $z_{\mathrm{ref}}^{\mathrm{air}}$ in air. The area that the camera sensor images at $z_{\mathrm{ref}}^{\mathrm{air}}$ is labeled by the gray square as “FOV in air” in **Supplementary Fig. 24**a–b. Given a point $\mathbf{r} = (\mathbf{r}_\parallel, z)$ in the sample and a transverse momentum $\mathbf{k}_\parallel$, we can backpropagate the associated ray to air (while accounting for refraction at the interfaces) and then forward propagate the ray (without refraction) to find its would-be intersection point with the reference plane in air if the sample is not there, $\mathbf{r}_{\mathrm{ref}}^{\mathrm{air}} = (x^{\mathrm{air}}, y^{\mathrm{air}}, z_{\mathrm{ref}}^{\mathrm{air}})$ (**Supplementary Fig. 24**a). If $\mathbf{r}_\parallel^{\mathrm{air}} = (x^{\mathrm{air}}, y^{\mathrm{air}})$ is within the FOV in air (the gray square), such transverse momentum $\mathbf{k}_\parallel$ is accessible at point $\mathbf{r}$ in the sample. If $\mathbf{r}_\parallel^{\mathrm{air}}$ is outside the gray square, then this $\mathbf{k}_\parallel$ is not accessible, meaning such ray cannot be detected by the camera. **Supplementary Fig. 24**a–b schematically illustrate examples of an accessible $\mathbf{k}_\parallel$ (in red) and an inaccessible $\mathbf{k}_\parallel$ (in blue).

Specifically, here we have the sample refractive index $n_2 = n_{\mathrm{sam}} = 1.4$ (PDMS), a coverslip with thickness $h_1 = h^{\mathrm{glass}} = 154$ µm and refractive index $n_1 = n_{\mathrm{glass}} = 1.5$, and a reference plane of $z_{\mathrm{ref}}^{\mathrm{air}} = 996$ µm. Using Snell’s law and



Supplementary Fig. 24. Theoretical lateral resolution, signal strength, and depth of field for the nanoparticle colloid sample. a, For each point $\mathbf{r}$ in the sample and each transverse momentum $\mathbf{k}_{\|}$, we can determine its would-be position $\mathbf{r}_{\text{ref}}^{\text{air}}$ at the reference plane $z_{\text{ref}}^{\text{air}}$ in air, which is the plane imaged by the camera. **b**, When $\mathbf{r}_{\text{ref}}^{\text{air}}$ falls within the FOV in air, the $\mathbf{k}_{\|}$ is accessible (red). When $\mathbf{r}_{\text{ref}}^{\text{air}}$ falls outside the FOV in air, the $\mathbf{k}_{\|}$ is not accessible (blue). **c**, Example of accessible and inaccessible $\mathbf{k}_{\|}$ points, at $\mathbf{r} = (10, 10, 1485)$ μm. **d**, Effective NA defined by the range of accessible angles, which determines the ideal (diffraction-limited) lateral resolution. **e**, Ratio of accessible $\mathbf{k}_{\|}$ points, which determines the overall signal strength in SMT.

geometry, we obtain

$$\Delta\mathbf{r}_{\|} \equiv \mathbf{r}_{\|}^{\text{air}} - \mathbf{r}_{\|} = \left[(z_{\text{ref}}^{\text{air}} + h_1)\tan\theta_0 - h_1\tan\theta_1 - z\tan\theta_2\right]\frac{\mathbf{k}_{\|}}{\left|\mathbf{k}_{\|}\right|}, \quad \sin\theta_0 = n_1\sin\theta_1 = n_2\sin\theta_2 = \left|\mathbf{k}_{\|}\right|/k_0, \tag{S70}$$

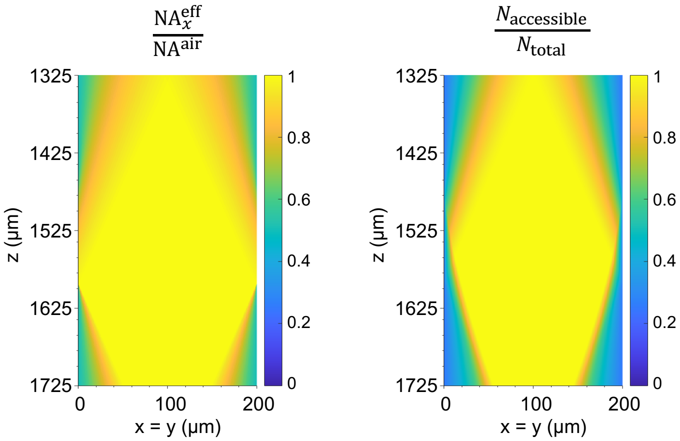
where θ_0, θ_1 and θ_2 are the propagation angles relative to the z axis in different media, and $k_0 = 2\pi/\lambda$ is the vacuum wave number. Note that the in-plane momentum $\mathbf{k}_{\|} = (k_x, k_y)$ is the same in all media. For each point $\mathbf{r}$ in the 3D space of the sample, we loop through the list of in-plane momentum $\mathbf{k}_{\|}$ within the NA of the objective lens and use Eq. (S70) to determine if each $\mathbf{k}_{\|}$ is accessible or not. We denote the accessible ones as $\mathbf{k}_{\|}^{\text{accessible}} = (k_x^{\text{accessible}}, k_y^{\text{accessible}})$. **Supplementary Fig. 24**c provides an example.

The ideal (diffraction-limited) lateral resolution is determined primarily by the maximal spread of $\mathbf{k}_{\|}^{\text{out}} - \mathbf{k}_{\|}^{\text{in}}$. Since we use the same NA for the input and the output, this is determined by the maximal spread of $\mathbf{k}_{\|}$. So, we define an effective NA as

$$\text{NA}_x^{\text{eff}} \equiv \frac{\max(k_x^{\text{accessible}}) - \min(k_x^{\text{accessible}})}{2k_0} \tag{S71}$$

and similarly for NA_y^{eff}. **Supplementary Fig. 24**d plots NA_x^{eff} along a diagonal ($x = y$) slice of the 3D sample. Within the total volume we consider in Fig. 4 of the main text (from $z = 1470$ μm to $z = 1580$ μm, indicated by dashed lines), NA_x^{eff} remains close to the ideal value of $\text{NA}^{\text{air}} = 0.5$. If one defines the DOF as where the lateral resolution increases by a factor of $\sqrt{2}$ (which is the typical definition in RCM and OCM/OCT), the SMT DOF will be much larger than the 110 μm considered.

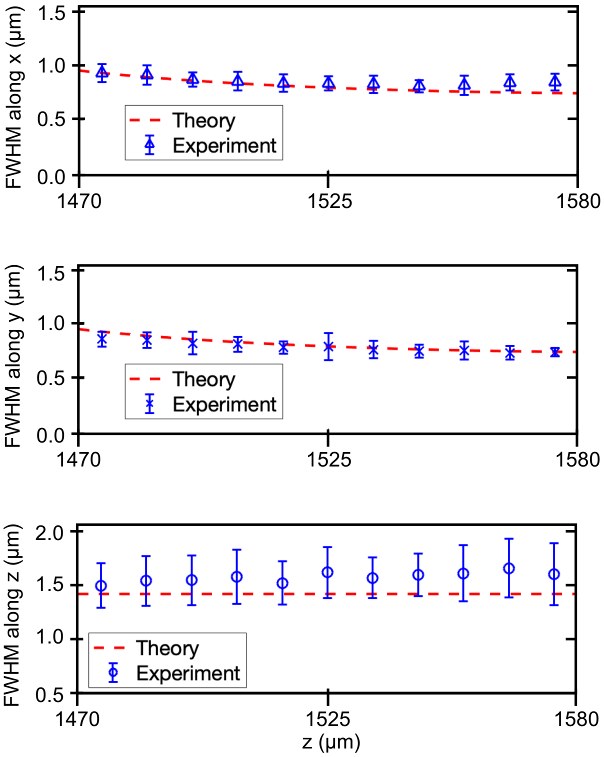
Note that the NA_x^{eff} in **Supplementary Fig. 24**d is asymmetric in z. This is because of the refraction at the air-glass and glass-sample interfaces. As mentioned in Supplementary Sect. III D and illustrated in **Supplementary Fig. 8**b, the depth $z_{\text{f}}^{\text{sam}}$ of the focal plane in the sample depends on the incident angle. Here, $z_{\text{f}}^{\text{sam}}(0°) = 1466$ μm and $z_{\text{f}}^{\text{sam}}(30°) = 1594$ μm. Since NA_x^{eff} is determined by the maximal angle, it is larger near $z_{\text{f}}^{\text{sam}}(30°) = 1594$ μm, so the lateral resolution is the sharpest there. Without refractive index mismatches, $z_{\text{f}}^{\text{sam}}$ will be independent of the incident angle, and NA_x^{eff} will decay symmetrically away from the reference plane.



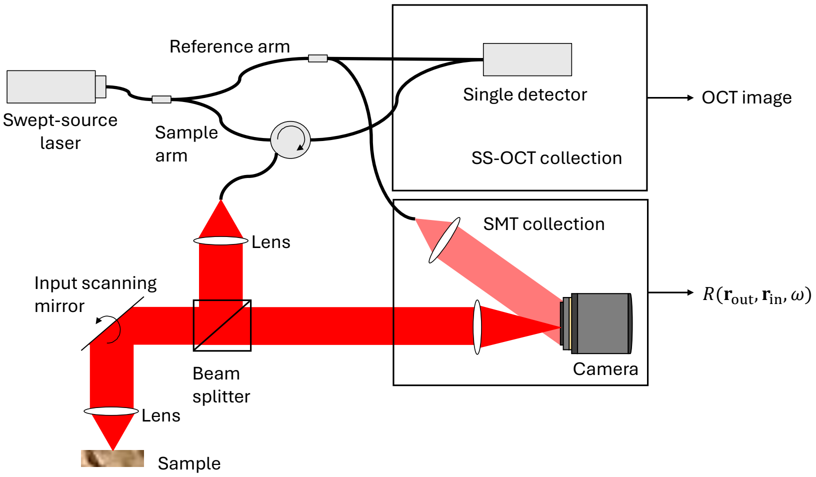
Supplementary Fig. 25. **Depth of field with expanded field of view.** The effective NA (left) and the ratio of accessible $\mathbf{k}_{\parallel}$ (right) when the field of view is 200 µm.

The strength of the signal depends on the number of accessible $\mathbf{k}_{\parallel}$. Therefore, in **Supplementary Fig. 24**e, we plot the ratio $N_{\mathrm{accessible}}/N_{\mathrm{total}}$ between the number of accessible $\mathbf{k}_{\parallel}$ points and the total number of $\mathbf{k}_{\parallel}$ within the NA of the objective lens. This ratio remains above 0.5 across most points within the volume considered in the main text.

With the geometry shown above, we expect the SMT DOF to grow with the FOV. As an example, **Supplementary Fig. 25** shows the corresponding predictions given a FOV of 200 µm. In this case, SMT has a DOF of roughly 400 µm.



Supplementary Fig. 26. **Experimental FWHM resolution of nanoparticle colloid.** Each experimental data point is the mean value over a 10-µm depth interval, with error bars indicating plus/minus one standard deviation.



Supplementary Fig. 27. **Strategy to integrate SMT and swept-source OCT.**

C. Experimental FWHM resolution

To systematically study the resolution of the experimental SMT image, we use the Trackmate [97, 100] plugin in Fiji [98] with a Laplacian of Gaussian (LoG) detector to analyze the 3D SMT image of the nanoparticle colloid, setting the estimated diameter as 4 pixels (0.2 µm per pixel), quality threshold as 130. The analysis locates ∼3,610 particles. For each located particle, we extract the FWHM resolution of its SMT image along the x, y, and z directions using the `findpeaks` function in MATLAB. We bin the data set with a depth interval of 10 µm and calculate the mean and standard deviation within each depth interval, with results shown in **Supplementary Fig. 26**. The FWHM resolution in y is slightly better than in x because our incident angular range in y is larger after excluding incident angles with weak signals (Supplementary Sect. I C).

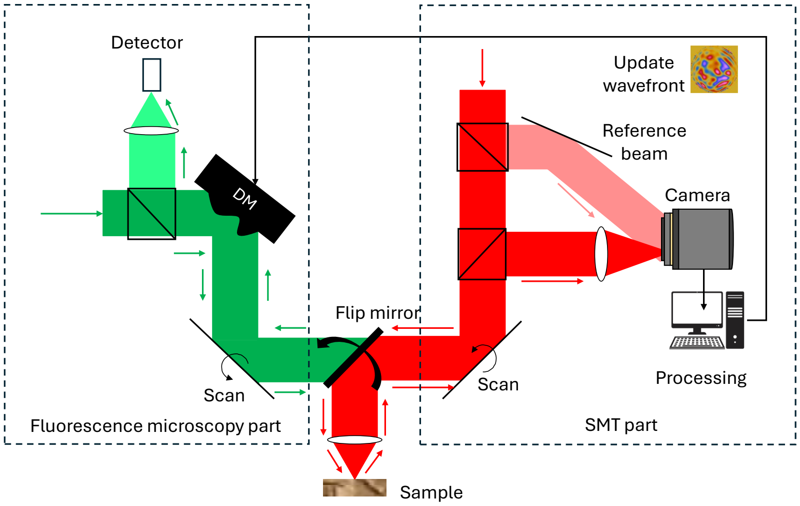
We also plot the theoretically predicted results as red dashed lines. For the lateral resolution, we use the model of Supplementary Sect. VIII B above to calculate $\text{NA}^{\text{eff}}(x, y, z)$ at each position (x, y, z) in the 3D volume and average over (x, y) for each z to obtain $1/\text{NA}^{\text{eff}}(z) \equiv \langle 1/\text{NA}^{\text{eff}}(x, y, z)\rangle_{x,y}$ to model the z dependence of the lateral resolution. For the axial resolution, we show the theoretical estimate of $\delta z = 1.42$ µm from Supplementary Sect. VIII A.

IX. INTEGRATING SMT AND OTHER IMAGING MODALITIES

In this section, we propose some strategies to integrate SMT with other imaging modalities. The current data acquisition speed of SMT is severely limited by the Ti:Sa tunable laser. By integrating SMT with swept-source OCT, the scanning speed of the laser no longer becomes a limiting factor. When measuring the matrix in the spatial basis, SMT captures the scattered light from various locations while OCT only captures the scattered light from the illuminated locations, making it a narrow case of SMT. Therefore, SMT and OCT can share the illumination part of the optical system while differing in the collection part. **Supplementary Fig. 27** illustrates a strategy to incorporate SMT with swept-source OCT, which is a standard imaging modality for medical, especially surgical, applications. Additionally, since SMT provides powerful aberration corrections, its correction maps can become useful for fluorescence or multi-photon microscopy. **Supplementary Fig. 28** illustrates a strategy to incorporate SMT with fluorescence microscopy, adopting the proposed CLASS-single molecule localization microscopy integrated system of Park *et al*. [99], where a flip mirror separates the SMT and the fluorescence microscopy parts. SMT runs and identifies the optimal correction phases before loading it to the DM on the pupil plane of the fluorescence part of the optical system. These are only two examples of future optical systems where SMT can be used alongside other imaging modalities.

X. SMT FOR IMAGING DECORRELATION OF DYNAMIC SAMPLES

Similar to OCT angiography, SMT can also be used to image the decorrelation of dynamic samples such as blood vessels. We take multiple SMT measurements, indexed from 1 to N, and build multiple SMT images $I_j(\mathbf{r})$ (j = 1,..,N).



Supplementary Fig. 28. **Strategy to integrate SMT and fluorescence microscopy.** Here, the red color represents the SMT sample beam, the light red represents SMT reference beam, the green represents incident light of the fluorescence microscopy, and the light green represents the output light of the fluorescence microscopy which has a different wavelength from the incidence.

Then, we compute the decorrelation[74]

$$D(\mathbf{r}) = 1 - 2\sum_{j=1}^{N-1} \frac{I_j(\mathbf{r})I_{j+1}(\mathbf{r})}{I_j^2(\mathbf{r}) + I_{j+1}^2(\mathbf{r})} \tag{S72}$$

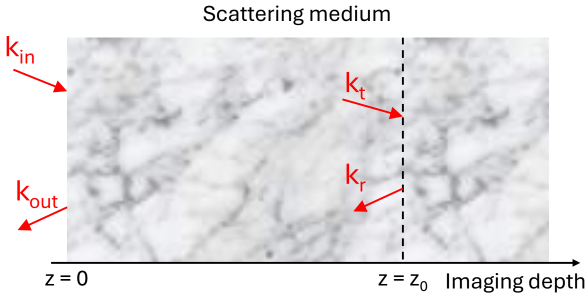
In OCT angiography, as the dynamic samples such as blood vessels decorrelate a lot faster than their surrounding tissues, the decorrelation image $D(\mathbf{r})$ can successfully show the blood vessels that may not show up on the original OCT images. SMT can adopt the same principle. The difference is that each SMT image may have higher quality than OCT images due to better aberration correction.

XI. SMT FOR VECTORIAL WAVES

SMT is capable of measuring the vectorial scattering matrices by controlling the two Fresnel rhombs after the polarized beam splitter, as shown in **Supplementary Fig. 1**. The Fresnel rhomb on the sample arm determines the polarization of the incident wave. The Fresnel rhomb on the reference arm determines the output polarization. For example, when the sample arm's Fresnel rhomb is right-handed circular polarization while the reference arm's Fresnel rhomb is left-handed circular polarization, we get the interference of the reference beam and the left-handed circularly polarized scattered wave on the camera.

Controlling the two Frenel rhombs, we can measure the full vectorial scattering matrices, labeled as $S_{\mathrm{RR}}(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$, $S_{\mathrm{RL}}(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$, $S_{\mathrm{LR}}(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$, and $S_{\mathrm{LL}}(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$. In the subscripts RR, RL, LR, and LL, the first letter represents the polarization of the incident wave (R: right-handed circular polarization, L: left-handed circular polarization), and the second letter represents the polarization of the output waves. For example, $S_{\mathrm{RL}}(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$ contains the scattering coefficients at frequency ω, the incidence is right-handed circularly polarized with wavenumber $\mathbf{k}_{\mathrm{in}}$, the output is left-handed circularly polarized with wavenumber $\mathbf{k}_{\mathrm{out}}$. Applying **Eq.** 2 on each of these matrices, we get four images $I_{\mathrm{RR}}(\mathbf{r})$, $I_{\mathrm{RL}}(\mathbf{r})$, $I_{\mathrm{LR}}(\mathbf{r})$, and $I_{\mathrm{LL}}(\mathbf{r})$. Each of these images shows how the imaging sample responds to polarized incidence, similar to polarization-sensitive OCT. For example, if the incident light is scattered at position r and the polarization remains relatively unchanged, the pixels r are bright on the co-polarized images $I_{\mathrm{RR}}(\mathbf{r})$ and $I_{\mathrm{LL}}(\mathbf{r})$ and dark on the cross-polarized images $I_{\mathrm{RL}}(\mathbf{r})$ and $I_{\mathrm{LR}}(\mathbf{r})$ and vice versa.

Each vectorial scattering matrix can also have its phases modified to correct for sample-induced aberrations by optimizing the corresponding image, the same as how we correct for aberrations in the non-vectorial scattering matrix $S(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$. For example, the aberrations on the right-handed circularly polarized incidence can be corrected by applying a phase shift $\phi_{\mathrm{inR}}(\mathbf{k}_{\mathrm{in}})$ to each input kin of $S_{\mathrm{RR}}(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$ and $S_{\mathrm{RL}}(\mathbf{k}_{\mathrm{out}},\mathbf{k}_{\mathrm{in}},\omega)$, $\phi_{\mathrm{inR}}(\mathbf{k}_{\mathrm{in}})$ can be found by optimizing the quality metrics M of $I_{\mathrm{RR}}(\mathbf{r})$ and/or $I_{\mathrm{RL}}(\mathbf{r})$.



Supplementary Fig. 29. **The propagation of light from the surface of a scattering medium to an imaging depth.**

XII. POTENTIAL USE OF NUMERICAL SIMULATIONS TO IMAGE DEEPER WITH MULTIPLE SCATTERING SIGNALS

Instead of treating multiple scattering as noise and suppressing it, an alternative approach to image deep in scattering media is to utilize multiple scattering as useful signals. Here we discuss the potential use of numerical simulations to obtain multiple scattering signals and use them to image deeper.

First, we have a look at the signal formation at depth $z = z_0$ inside a scattering medium. In **Supplementary Fig. 29**, incident light at frequency ω and wavenumber $\mathbf{k}_{\text{in}}$ travels inside the scattering medium from its surface at $z = 0$ to depth $z = z_0$. The transmitted light with wavenumber $\mathbf{k}_{\text{t}}$ is scattered by the scatterers at depth $z = z_0$ and becomes reflected light with wavenumber $\mathbf{k}_{\text{r}}$. That reflected light passes through the scattering medium back to $z = 0$ and emerges out of the scattering medium with wavenumber $\mathbf{k}_{\text{out}}$. The reflection coefficient $R(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{in}}, \omega)$ can be expressed as:

$$R(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{in}}, \omega) = T^{-1}(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{r}}, \omega, z < z_0) R'(\mathbf{k}_{\text{r}}, \mathbf{k}_{\text{t}}, \omega, z \geq z_0) T(\mathbf{k}_{\text{t}}, \mathbf{k}_{\text{in}}, \omega, z < z_0) \tag{S73}$$

where $R(\mathbf{k}_{\text{r}}, \mathbf{k}_{\text{t}}, \omega, z \geq z_0)$ is the reflection matrix of the scattering medium at depth $z \geq z_0$, linking input $\mathbf{k}_{\text{t}}$ to output $\mathbf{k}_{\text{r}}$ at frequency ω, and $T(\mathbf{k}_{\text{t}}, \mathbf{k}_{\text{in}}, \omega, z < z_0)$ is the transmission matrix of the scattering medium from $z = 0$ to $z = z_0$, linking $\mathbf{k}_{\text{in}}$ to output $\mathbf{k}_{\text{t}}$ at frequency ω. $R'(\mathbf{k}_{\text{r}}, \mathbf{k}_{\text{t}}, \omega, z_0)$ can be found by:

$$R'(\mathbf{k}_{\text{r}}, \mathbf{k}_{\text{t}}, \omega, z \geq z_0) = T(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{r}}, \omega, z < z_0) R(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{in}}, \omega) T^{-1}(\mathbf{k}_{\text{t}}, \mathbf{k}_{\text{in}}, \omega, z < z_0) \tag{S74}$$

The eventual reflection matrix $R(\mathbf{k}_{\text{out}}, \mathbf{k}_{\text{in}}, \omega)$ is what we can measure. $R'(\mathbf{k}_{\text{r}}, \mathbf{k}_{\text{t}}, \omega, z \geq z_0)$ is what we need to find to image at depth z_0 and below. For that purpose, the transmission matrix $T(\mathbf{k}_{\text{t}}, \mathbf{k}_{\text{in}}, \omega, z < z_0)$, which consists of both single and multiple scattering, has to be determined.

To find this matrix with simulations, we image the sample at other shallower depths $z < z_0$. These images may provide information about the locations and sizes of scatterers at $z < z_0$, which can be used to construct a simulation model. Simulations are then conducted to find the transmission matrix of the scattering medium from $z = 0$ to $z = z_0$.

In practice, it remains an open question whether simulations can help push the imaging depth in a real, particular biological sample. A simulation model that exactly resembles the real sample, i.e., the sizes, positions, and refractive indices of all scatterers inside the sample, is needed. This is practically impossible for label-free non-invasive imaging because quantitative phase imaging methods that give the samples' refractive indices require cutting the samples into slices and measuring the transmitted light. In the above example, even if we image and get the sizes and positions of scatterers above z_0, their exact refractive indices are unknown. For some well-known tissues or cells, one can use approximated refractive indices from available databases [85]. However, as we image very deep inside scattering media, approximately 1000 times the wavelength, and the phase that light accumulates during propagation scales linearly with the product of depth and refractive index, a small error in the refractive index can potentially lead to significant error on the phase of the simulated transmission matrix. In addition, before reaching the imaging depth z_0, light has been scattered multiple times, each time light scatters off a scatterer, its refractive index determines how many photons go into which direction. Thus, a small error in the refractive index can lead to huge errors on the amplitude of the simulated transmission matrix. All of the uncertainty in refractive indices renders the simulated multiple scattering signals unusable. Therefore, there is still a long way to go from understanding how light is scattered inside a simulated scattering phantom to accurately evaluating the multiple scattering signals in a real, particular sample to image deeper.