1. Proof of Equation (2) of the main paper

Synthetic wavelength interferometry uses illumination comprising two distinct, but narrowly-separated, lasers that are incoherent with each other. We denote the wavelengths of these lasers $\lambda$ and $\lambda/1+\varepsilon$, corresponding to wavenumbers $\kappa \equiv 2\pi/\lambda$ and $(1+\varepsilon)\kappa$, respectively. In full-field SWI, the outputs of these lasers are collimated by a lens into a beam covering the field of view. The laser with wavelength $\lambda$ results in a wavefront parallel to the optical axis. The corresponding field propagating towards the beamsplitter is

$$u^\lambda_s(x, l) = \exp(-i\kappa l_0)$$ (1)

When the scene point at $x$ is placed at a distance $d(x)$ from the beamsplitter, the illumination travels 2$d(x)$ in the scene arm, accounting for the propagation to the scene and back. Then, the field due to the scene at the sensor pixel $x$ is

$$u^\lambda_s(x) = \exp(-i\kappa (2d(x) + l_0))$$ (2)

where $l_0$ is the travel distance from the lens to the beamsplitter and the beamsplitter to the sensor. Similarly, when the reference arm is at a distance $l$ from the beamsplitter, the field due to the reference arm at sensor pixel $x$ is

$$u^\lambda_r(x, l) = \exp(-i\kappa (2l + l_0))$$ (3)

Then, the correlation $C^\lambda (x, l)$ for the wavelength $\lambda$ equals

$$C^\lambda (x, l) = u^\lambda_s(x)u^\lambda_r(x, l)$$ (4)

$$= \exp(-2i\kappa (d(x) + l_0))\exp(i\kappa (l + l_0))$$ (5)

$$= \exp(-2i\kappa (d(x) - l))$$ (6)

As the two lasers are incoherent with each other, there are no cross-correlations between the fields. Thus, $C(x, l)$ for both lasers equals the incoherent sum of their correlations:

$$C(x, l) = C^\lambda (x, l) + C^{(1+\varepsilon)\kappa} (x, l)$$ (7)

$$= \exp(-2i\kappa (d(x) - l)) + \exp(-2i\kappa (1+\varepsilon)(d(x) - l))$$ (8)

$$= \exp(-2i\kappa (d(x) - l)) [1 + \exp(-2i\kappa (d(x) - l))]$$ (9)

thus proving Equation (2) from the main paper.

2. Scanning versus full-field comparison

Scan points for equal-time acquisition. In Figure 9 of the main paper, we show depth reconstructions with full-field swept-angle SWI and upsampling point scanning SWI for the {4, 4}-shift algorithm. To emulate point scanning, we downsampled the SWI depth by a factor of 35 in each dimension, and claimed that this corresponds to an equal-time comparison for a 30 kHz MEMS scanner. Here, we detail this calculation.

Our swept-angle SWI system, for scenes with very low reflectivity, operates at 1 Hz. In the equivalent time of 1 s, the scanner must perform 16 passes over the scene to take the 16 measurements required for the {4, 4}-shift algorithm. This makes the maximum number of points the scanner can measure in two dimensions $30000/16 = 1875$. In one dimension, this translates to $\sqrt{1875} \approx 43$ points. Our images have approximate dimension $1600 \times 1300$. Distributing these points equally along the larger dimension yields a downsampling factor of $1600/43 \approx 35$.

Challenges for achieving micrometer lateral resolutions with scanning systems. The main paper discusses some of the challenges associated with achieving micrometer lateral resolution using a scanning system operating in resonant mode (e.g., using Lissajous scanning). Here, we discuss in more detail additional challenges in achieving micrometer lateral resolutions using a scanning system. Doing so requires: (i) a laser beam that can be collimated or focused at few micrometers; (ii) a MEMS mirror capable of scanning at high-enough angular resolution to translate the laser beam a few microns on the scene surface; and (iii) acquisition time long enough to scan a megapixel-size grid on the scene. Each of these requirements is difficult to achieve:

(i) The diameter of a Gaussian laser beam is inversely proportional to its divergence [5, Chapter 4]. The smaller the beam diameter, the larger the divergence. At 780 nm, a laser beam with a diameter of 1 µm grows in diameter by 10% every 2 m. Therefore, maintaining collimation diameter of 1 µm is challenging except for very small working distances.
As an alternative to using a thin, collimated laser beams, we can use a beam that is focused at each point on the scene. Contrary to micron-scale beam waists, it is possible to focus single-mode lasers to pixel-size spot sizes [5, Chapter 9]. However, focusing the laser beam onto the scanned scene points sharply decreases the depth of field of the imaging system. Whereas, in the case of a collimated beam, the depth of field is limited by the divergence of the collimated beam, in the case of a focused beam, it is limited by the quadratic phase profile of the focused spot. Effectively, to use this focused setup, we need another axial scan to ensure that the scanned post is within the depth of field, which only adds to acquisition time.

(i) Top-of-the-line scanning micromirrors typically have angular scanning resolutions of 10 µrad [4]. The maximum working distance such that this would correspond to micrometer lateral resolution is 10 cm.

(ii) The scanning micromirror needs to be run in “point-to-point scanning mode” [3], where the micromirror stops at every desired position. The best settling times for step mirror deflections are around 100 µs [4]. Using these numbers, for a megapixel image, the micromirror rotations require acquisition time around 100 s.

A swept-angle full-field interferometer does not need to perform lateral scanning of the image plane. Instead, it accomplishes direct-only (i.e., coaxial) imaging by scanning an area in the focal plane of the collimating lens, an operation that can be done in the resonant mode of a MEMS mirror within exposure and at low lateral resolution.

3. Additional experiments

Trade-off between acquisition time and depth quality.

The theoretical minimum number of measurements needed to reconstruct depth using SWI is \( M \cdot N = 3 \). However, increasing \( M \) and \( N \) makes the depth reconstruction robust to measurement and speckle noise, yielding higher quality depth. There is, therefore, a trade-off between number of measurements \( M \cdot N \) and depth quality.

Besides number of measurements, another factor contributing to acquisition time is the MEMS mirror scan we perform to create swept-angle illumination. If we decrease acquisition time, giving the mirror less time to complete one full scan of the focal plane of the collimating lens, the spatial density of scanned points on focal plane decreases. This reduces the effectiveness of rejecting indirect light, and therefore reduces depth quality. This, again, creates a trade-off between acquisition time and depth quality.

In Figure 1, we demonstrate the effect of both these factors on depth quality. On the horizontal dimension, we use different \( \{ M, N \} \)-shift phase retrieval algorithms, and on

![Figure 1. Depth quality and acquisition time.](image)

the vertical dimension, we use different per-image acquisition times, corresponding to different focal plane scanning resolutions. Using higher \( M \) and \( N \) allows us to reduce the per-image acquisition time, by requiring a lower scanning density for equal depth quality. In particular, the 10 ms scan with the \( \{4, 5\} \)-shift algorithm performs as well as the 250 ms scan with the \( \{5, 5\} \)-shift algorithm, allowing us to reduce total acquisition time from 10 s to 250 ms.

Tunable depth range.

The use of two wavelengths in synthetic wavelength interferometry makes it possible to control the unambiguous depth range: By decreasing the separation \( \kappa \) between the two laser wavelengths, we increase the unambiguous depth range, at the cost of decreasing depth resolution. In particular, picometer separations in wavelengths result in synthetic wavelengths of centimeters. We use this to scan the macroscopic scenes in Figure 2, which have a depth range of approximately 1 cm. In all three scenes, the use of swept-angle illumination greatly improves reconstruction quality, by mitigating the effects of the significant subsurface scattering present in all scenes.

We note that achieving picometer-scale wavelength separation requires using current-based tuning of the wavelength of the DBR lasers in our setup. The DBR lasers have a linear response of wavelength to current near their operating point, and tuning current by 50 mA gives picometer-scale wavelength separations.

Robustness to ambient light.

In Figure 3, we demonstrate the robustness of our method to ambient light on the toy cup scene from Figure 2. We shine a spotlight on the scene such that the signal-to-background ratio (SBR) of the laser
**Depth reconstruction.** Depth maps (left) and surface renderings (right) acquired using full-field SWI with: (c) swept-angle scanning and bilateral filtering; (d) swept-angle scanning and Gaussian filtering; (e) no swept-angle scanning.

**Robustness to ambient light.** In (c), we shine external light on the sample so that the signal-to-background ratio (SBR) our laser illumination to ambient noise is 0.1. This greatly decreases the contrast of the interference speckle pattern. Despite this, there is little degradation in the quality of our recovered depth.

**Depth accuracy.** We show here the data we captured to assess our depth accuracy in Table 1 of the main paper. Figure 4 plots, on top, the estimated SWI depth relative to the ground truth depth provided by the scene translation stage. Figure 4(a) is captured with the swept-angle mechanism, whereas Figure 4(b) is captured with the mechanism off. Comparing the two figures, we see that the measured depth correlates with the ground truth depth a lot stronger when we use swept-angle illumination versus when we do not.

Figure 4(c) and Figure 4(d) respectively show the same experiment performed at a *macroscopic* synthetic wavelength of 16 mm, the same as in Figure 2. These measurements also depict that swept-angle scanning is critical for micron-scale depth sensing. We show the error numbers from this experiment, similar to Table 1 in the main paper, in Table 1. With a kernel width of 150 µm, we are able to achieve depth accuracies of 50 µm.

**Comparison with full-field OCT.** As mentioned in the main paper, depth sensing with full-field spatially-incoherent OCT achieves unambiguous depth ranges up to centimeters at micrometer axial resolutions. Here, we
Table 1. Depth accuracy with synthetic wavelength 16 mm. MedAE is the median absolute error between ground truth and estimated depth. Kernel width is the lateral size of the speckle blur filter. All quantities are in \( \mu \text{m} \).

<table>
<thead>
<tr>
<th>kernel width</th>
<th>with swept-angle</th>
<th>w/o swept-angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>471.4</td>
<td>1267.2</td>
</tr>
<tr>
<td>15</td>
<td>167.1</td>
<td>577.9</td>
</tr>
<tr>
<td>21</td>
<td>78.7</td>
<td>609.5</td>
</tr>
<tr>
<td>30</td>
<td>81.7</td>
<td>605.7</td>
</tr>
</tbody>
</table>

We experimented with multiple alterations (of the order of \( \frac{1}{2} \) MHz) stable in wavelength and power, and importantly for the light sources to be monochromatic (single longitudinal mode), stable in wavelength and power, and accurately tunable. We found the DBR lasers from Thorlabs optimal in all these aspects.

**Calibrating the synthetic wavelength.** The synthetic wavelength resulting from this illumination is very sensitive to the separation between the two wavelengths, especially at microscopic scales. Therefore, after selecting a pair of lasers or current levels for an approximate synthetic wavelength, it is necessary to estimate the actual synthetic wavelength accurately. To do this, we measure the envelope sinusoid for a planar diffuser scene at a dense collection of reference arm positions. We then fit a sinusoid to these measurements to estimate the synthetic wavelength.

**Mechanism for swept-angle scanning.** We use two fast-rotating mirrors to scan the laser beam in \( \frac{\pi}{4} \times \frac{\pi}{4} \) angular pattern at slightly separated \( \text{kHz} \) frequencies, as shown by Kotwal et al. [2] and Liu et al. [3]. A \( 35 \text{ mm} \) Nikon prime lens then maps beam orientation to spatial coordinates at the focal plane of the illumination lens, creating the effective area light source for swept-angle illumination. The emission area for this effective light source is a dense Lissajous curve that approximates a square. Figure 7 shows an example scanning pattern. In practice, we use much denser scanning patterns, but show the coarse one in the inset only to make the Lissajous pattern visible. We note that our actual scanning patterns are still at much lower resolution than the pixel-level resolution that would be required in a scanning SWI system that raster scans the image plane. For this choice of scanning resolution, we followed Gkioulekas et al. [1] and Kotwal et al. [2], who show that the extent of the scanning pattern is more important than scanning density. Intuitively, as we decrease scanning density, we improve SNR (more light paths contribute to interference component, speckle contrast is stronger), at the cost of rejecting less indirect illumination. Figure 1 shows the scanning patterns we use for experiments (insets), and experimentally quantifies this trade-off.

**Illumination lens.** We place the above light source in the focal plane of a \( 200 \text{ mm} \) Nikon prime lens to generate the swept-angle illumination. Photographic lenses perform superior to AR-coated achromatic doublets in terms of spherical and chromatic aberration, therefore resulting in significantly lesser distortion in the generated wavefront.

**Interreflections.** Interreflections are problematic for us because our illumination is temporally coherent. Interreflections introduce multiple light paths that interfere with each
Figure 5. Comparing swept-angle SWI and full-field OCT in microscopic scenes. The difference maps show the absolute difference between recovered OCT and swept-angle SWI depths, and report the root-mean-square (top) and mean absolute (bottom) differences between the two. The OCT depths are captured at a resolution comparable to swept-angle SWI.

other to create strong spurious fringes. Such fringes suppress the contrast of our speckle signal. The optics we use are coated with anti-reflective films designed for our laser wavelengths to reduce interreflections. We also deliberately misalign our optics with sub-degree rotations from the ideal alignment to avoid strong interreflections.

**Beamsplitter.** We use a 50:50 plate beamsplitter, as pellicle and cube beamsplitters create strong fringes. As above, we misalign the beamsplitter to avoid interreflections.

**Mirrors.** We use high-quality mirrors of guaranteed λ/4 flatness to ensure a uniform phase reference throughout the field of view of the camera.

**Translation stage.** We use a high-precision motorized linear translation stage with a positioning accuracy of up to 10 nm and minimum incremental motion of 1 nm. For high-resolution depth recovery, it is important that the mirror positions images are captured at accurate sub-wavelength scales. In addition, it is important that the translation stage guarantee low-positioning-noise operation.

**Camera lens.** Our scenes are sized at the order of 1 inch.
Therefore, we benefit from a lens that achieves high magnifications (1:1 reproduction ratio). This also provides better contrast due to less speckle averaging (interference signal is blurred with the pixel box when captured with the camera). We use a 180 mm Canon prime macro lens for the camera.

**Camera.** We use a machine vision camera from Allied Vision with a high-sensitivity CCD sensor of 8 MP resolution, and pixel size 3.5 µm. A sensor with a small pixel pitch averages interference speckle over a smaller spatial area, therefore allowing us to resolve finer lateral detail. We use a camera with the sensor protective glass removed. This is critical to avoid spurious fringes from interreflections.

**Neutral density filters.** We use absorptive neutral density filters to optimize interference contrast by making the intensities of both interferometer arms equal.

**Alignment.** For depth estimation accurate to micron-scales, the optical setup requires very careful alignment. To avoid as much human error as possible, we build the illumination side and the beamsplitter holder on a rigid cage system constructed with components from Thorlabs. To ensure a mean direction of light propagation that’s parallel to the optical axis of the interferometer, we tune the steering mirrors electronically by adjusting their driving waveform’s DC offset. We then align the reference mirror and camera using the alignment technique described by Gkioulekas et al. [1].

**Component list.** For reproducibility, Table 2 gives a list of the important components used in our implementation.

## 5. Code and algorithms

We provide in Figure 8 an implementation of the \( \{4, 4\} \)-shift phase retrieval algorithm for recovering depth from measurements made with four subwavelength shifts and the
four-bucket algorithm. The code assumes that the measurements are stored in a variable frames of size \( H \times W \times 4 \times 4 \), where \( H \) and \( W \) are the height and width of the measured images respectively, with the third dimension varying over sub-wavelength shifts and the fourth varying over four-bucket positions. The variable scene stores an ambient light image of the scene to serve as the guide image for the bilateral filter, and the variable lam denotes the synthetic wavelength. The function bilateralFilter executes bilateral filtering of its first argument with its second argument as the guide image with spatialWindow and intensityWindow.

In addition, we provide in Algorithms 1 and 2 pseudocode for acquisition and reconstruction respectively using the general \( \{M, N\} \)-shift phase retrieval algorithm.

Algorithm 1: Acquiring intensity measurements with swept-angle synthetic wavelength interferometry. The steps are captioned with reference to Figure 6 of the main paper.

**Data:** synthetic wavelength \( \lambda_s \); optical wavelength \( \lambda_c \); start position \( l \)

**Result:** intensity images \( \mathcal{I}(x, l^m) \) at reference position \( l^m_n \) (defined below)

\[ l^m_n = l + n\lambda_c/N + m\lambda_s/M \]

for \( n \in \{0, \ldots, N-1\} \) and \( m \in \{0, \ldots, M-1\} \);

/* Capture the intensity images in Figure 6(a) */

for bucket positions \( n \in \{0, \ldots, N-1\} \) do

for sub-wavelength shifts \( m \in \{0, \ldots, M-1\} \) do

move reference mirror to position \( l^m_n \);

capture image \( I(x, l^m_n) \);

end

end

return \( I(x, l^m_n) \)

Algorithm 2: Processing intensity measurements to estimate depth in swept-angle synthetic wavelength interferometry. The steps are captioned with reference to Figure 6 of the main paper.

**Data:** synthetic wavelength \( \lambda_s \); optical wavelength \( \lambda_c \); start position \( l \); bilateral filter hyperparameters: spatial kernel size \( \sigma_s \) and intensity kernel size \( \sigma_i \); intensity measurements \( I(x, l^m_n) \) at reference position \( l^m_n \) (defined below); scene ambient-light image \( S(x) \)

**Result:** depth map \( d(x) \)

/* Initialization */

\[ l_{mn} = l + n\lambda_c/N + m\lambda_s/M \]

for \( m \in \{0, \ldots, M\} \)

for bucket positions \( n \in \{0, \ldots, M\} \) do

/* Figure 6(b) */

estimate interference-free image

\[ \mathcal{I}(x, l_n) = \left( \sum_{m=0}^{M-1} I(x, l^m_n) \right)/M; \]

/* Figure 6(c) */

estimate real parts of correlations

\[ \mathcal{C}(x, l^m_n) = I(x, l^m_n) - \mathcal{I}(x, l_n); \]

denoise envelope using the bilateral filter

\[ \mathcal{E}^{bf}(x, l_n) = \text{BilateralFilter} (\mathcal{E}^{2}(x, l_n), S(x), \sigma_s, \sigma_i); \]

end

/* Figure 6(e) */

estimate \( d(x) = \frac{1}{2\pi\sigma_c} \arctan \left[ \frac{\mathcal{E}^{bf}(x, l_n) - \mathcal{E}^{bf}(x, l_{n-1})}{\mathcal{E}^{bf}(x, l_n) - \mathcal{E}^{bf}(x, l_{n+1})} \right] + \hat{l}; \)

return \( d(x) \)

References


function depth = reconstruct(frames, lam, spatialWindow, ...
intensityWindow, scene)
    % Reconstruct depth from \{4, 4\}-shift swept-angle synthetic wavelength interferometry
    % frames: HxWx4x4 array of measurements, where the third dimension
    % varies over subwavelength shifts and fourth over
    % four-bucket positions
    % lam: synthetic wavelength
    % spatialWindow: spatial window for the bilateral filter
    % intensityWindow: intensity window for the bilateral filter
    % scene: ambient light image of the scene

    frames = im2double(frames)*4;
    scene = im2double(scene)*4;

    % Get interference-free images at each four-bucket position by averaging
    % images captured with sub-wavelength shifts
    interferenceFreeFrames = mean(frames, 3);

    % Get interference images at each four-bucket position by subtracting
    % interference-free images from the full images.
    interference = frames - interferenceFreeFrames;

    % Estimate the absolute values of the envelope by squaring and adding interference images
    envelope = squeeze(sum(interference.^2, 3));

    % Filter the estimated envelope with bilateral filtering using an ambient
    % light image of the scene as the guide image
    for position = 1:4
        envelope(:, :, position) = bilateralFilter(envelope(:, :, position), ...
            scene, spatialWindow, ...
            intensityWindow);
    end

    % Apply the four-bucket phase retrieval algorithm to estimate phase
    phase = atan2(envelope(:, :, 4) - envelope(:, :, 2), ...
        envelope(:, :, 1) - envelope(:, :, 3));

    % Convert phase to depth
    depth = phase*lam/(2*pi);
end

Figure 8. Matlab code for recovering depth from \{4, 4\}-shift measurements