

Supplemental Document for

Chip-Integrated Metasurface with Double-Helix Point Spread Functions for Ultra-Large Dynamic Range Wavefront Sensing

Yanhao Chu^{1,2,3,4,†}, Xiaopeng Li^{1,2,3,4,†}, Xian Long^{1,2,3,4,*}, Jiacheng Sun^{1,2,3,4}, Jian Li^{1,2,3,4},
Jitao Ji^{1,2,3,4}, Junyi Wang^{1,2,3,4}, Shining Zhu^{1,2,3,4} & Tao Li^{1,2,3,4,*}

1. National Laboratory of Solid State Microstructures, College of Engineering and Applied Sciences, Nanjing University, Nanjing 210093, China.
2. Key Laboratory of Intelligent Optical Sensing and Manipulations, Nanjing University, Nanjing 210093, China.
3. Jiangsu Key Laboratory of Artificial Functional Materials, Nanjing University, Nanjing 210093, China.
4. Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China.

†These authors contributed equally to this work.

*Correspondence author: loonggodx@nju.edu.cn, taoli@nju.edu.cn.

Supplementary Notes:

1. Design of the Double-Helix Phase Mask
2. Thermomechanical stability of optical adhesive bonding
3. Wavefront Reconstruction Sensitivity Limit
4. Wavefront Reconstruction Algorithm
5. Quantitative Analysis of DH-PSF Decoding Robustness
6. Substrate Propagation Correction
7. Dynamic Wavefront Reconstruction Demonstrations
8. Compatible Dual-Channel Vector Wavefront Sampling
9. Computational Complexity and Real-Time Implementation Analysis

Supplementary Reference.

Supplementary Note 1. Design of the Double-Helix Phase Mask

For a quasi-monochromatic scalar wave, a vortex indicates a variation in the complex amplitude defined in radial coordinates with its center at (r_0, θ_0) as

$$E(r, \theta) = A(r, \theta) (r e^{i\theta} - r_0 e^{i\theta_0})^c, \quad (1)$$

where $A(r, \theta)$ is the amplitude of the beam and c is the charge of the vortex. If the centers of vortex singularities with the same vortex charge are placed roughly along a central line, then the energy of the PSF will be distributed into two main lobes. The corresponding complex amplitude can be written as

$$E_0(r, \theta) = A(r, \theta) \prod_{j=-M}^M (r e^{i\theta} - r_j e^{i\theta_j}), \quad (2)$$

where N ($N=2M+1$) vortices with charge $+1$ are added to the pupil plane, $A(r, \theta)$ is the Gaussian amplitude, r_j and θ_j are respectively defined as follows

$$\begin{cases} r_j = |j| d \\ \theta_j = \frac{\pi}{2} \left(1 - \frac{j}{|j|} \right) + \theta_0, j \neq 0, \end{cases} \quad (3)$$

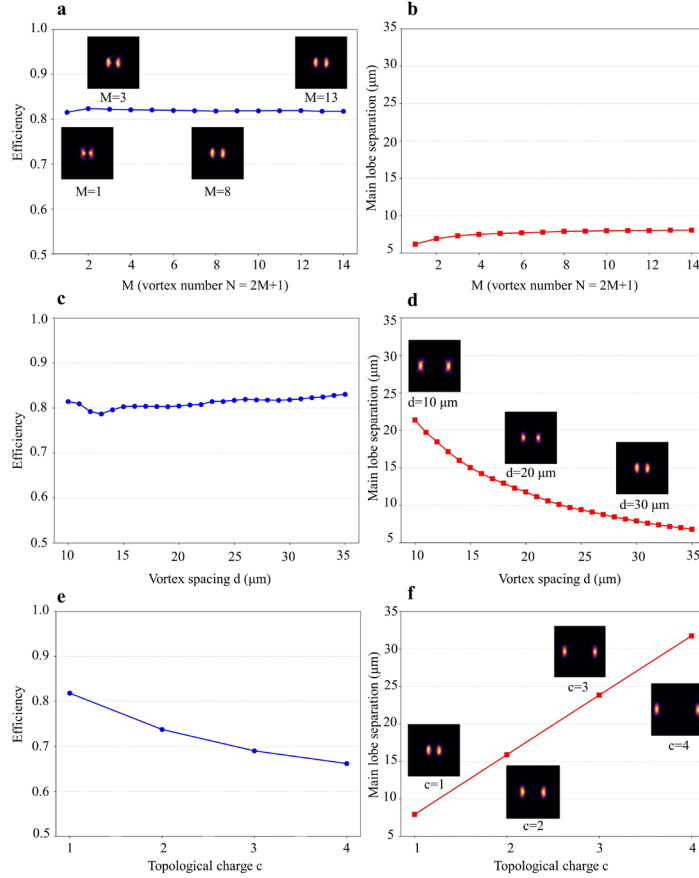
where d is the distance between vortices, θ_0 denotes the orientation angle of the double-helix point spread function (DH-PSF), defined as the angle between the line connecting its two main lobes and the x -axis. If the phase term in Eq. (1) is retained and the amplitude term is replaced with a clear circular aperture, the PSF on the focal plane still maintains the characteristics of a double helix. Meanwhile, it should be noted that even if the centers of some vortices are located outside the radius of the circular aperture, they still contribute to the overall amplitude and phase within the aperture, thereby affecting the PSF. Therefore, the complex transmittance function of the metasurface for a single sub-aperture writes

$$T(r, \theta) = \exp\{i \cdot \arg E_0(r, \theta)\} \exp\left(-i \cdot \frac{kr^2}{2f}\right), \quad (4)$$

where $\arg$ denotes the argument (phase) of a complex number, k is the wave vector and f is the focal length. The quadratic phase in the phase term plays the role of Fourier transform.

The dependence of key characteristics of the DH-PSF (efficiency, main lobe shape, and main lobe separation) on design parameters is subsequently investigated quantitatively. Specifically, Supplementary Figure 1(a) illustrates the efficiency of the DH-PSF as a function of the number of vortices N (where $N = 2M + 1$), under fixed conditions of vortex separation $d = 30 \mu\text{m}$ and topological charge $c = 1$. Efficiency is defined here as the ratio of the energy contained within the two main lobes to the total incident energy. The main lobes are identified as connected regions above a threshold set at 5% of the peak intensity, irrespective of their specific morphology. As a result of this design approach, the generated DH-PSF exhibits virtually no side lobes, with the energy at the focal plane being almost entirely concentrated within the two main lobes, thereby achieving high efficiency (above 80%). Despite the consistently high efficiency observed across all configurations, examination of the focal plane intensity profiles of the DH-PSF in Supplementary Figure 1(a) reveals that the main lobe intensity profile exhibits irregular shape when the number

of vortices is small ($M = 1$ and $M = 3$). As M increases, the DH-PSF gradually assumes a regular and stable shape. Supplementary Figure 1(b) presents the variation of main lobe separation as a function of M under the same parameter conditions, where the separation is defined as the distance between the geometric centers of the two main lobes. A minor variation in lobe separation is observed with increasing M . Supplementary Figure 1(c) and 1(d) illustrate the dependence of efficiency and main lobe separation on the vortex separation d , respectively, with fixed parameters $M = 8$ and $c = 1$. It is observed that as d increases, the efficiency maintains a high value while the main lobe separation decreases monotonically. Supplementary Figure 1(e) and 1(f) depict the variation of efficiency and main lobe separation as functions of the topological charge c , respectively, under fixed conditions of $M = 8$ and $d = 30 \mu\text{m}$. The results show that efficiency decreases with increasing c , whereas the main lobe separation increases approximately linearly with c .



Supplementary Figure 1. Design parameter dependence of double-helix point spread function characteristics. a-b: Efficiency and main lobe separation of the double helix point spread function as a function of M . c-d: Efficiency and main lobe separation of the DH-PSF as a function of the vortex separation d . e-f: Efficiency and main lobe separation of the DH-PSF as a function of the topological charge c . The insets in a, d, and f showing the corresponding intensity profiles of the DH-PSF.

Based on the above results, it is evident that the parameter d offers a wider tuning range for adjusting the main lobe separation compared to M . Furthermore, relative to the topological charge c , d enables more precise adjustments without compromising efficiency. To enable clear discrimination between the two polarization channels while maintaining regular lobe shapes that facilitate reliable identification, the parameters (d , M) for the left circularly polarized (LCP) and right circularly polarized (RCP) channels were selected as

(30×10^{-6} m, 8) and (15×10^{-6} m, 10), respectively. The resulting efficiencies and main lobe separations were (81%, 7.9 μ m) for the LCP channel and (80%, 15.4 μ m) for the RCP channel. This parameter selection ensures sufficient separation between the DH-PSF of the two polarization channels to prevent cross-talk, while maintaining regular lobe shapes that facilitate reliable identification and localization. Additionally, both channels achieve comparably high efficiencies, thereby enhancing the signal-to-noise ratio and ensuring accurate detection and positioning.

Supplementary Note 2. Thermomechanical stability of optical adhesive bonding

The optical adhesive used in this study is [OCA8146-2 Thorlabs.], and its relative temperature resistance, sourced from the official technical documentation, are as follows:

Short Term High Temperature: 177°C

Long Term High Temperature: 85°C

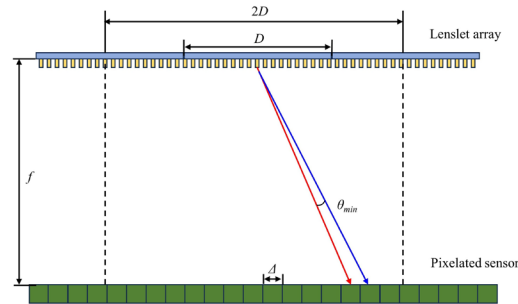
Long Term Low Temperature: -40°C

which confirm our device's reliability under typical laboratory and environmental conditions.

Thermal expansion or contraction of the CMOS cover glass and the adhesive layer, induced by temperature fluctuations, could in principle alter the optical path length between the metasurface and the detector, potentially leading to a slight rotation of the DH-PSF. However, the encoding scheme employed in this work incorporates substantial tolerance to such variations. The four primary angular encoding states of the DH-PSFs, oriented at 0°, 45°, 90°, and 135°, are designed with an allowable decoding margin of $\pm 22.5^\circ$. Within this range, minor rotations of the PSF do not affect orientation identification or centroid localization accuracy. Consequently, the precision of wavefront reconstruction remains uncompromised. This inherent design robustness ensures stable operation under realistic temperature variations.

Supplementary Note 3. Wavefront Reconstruction Sensitivity Limit

The sensitivity limit of our integrated metasurface wavefront sensor, defined as the smallest resolvable angle θ_{min} , is fundamentally determined by the geometric parameters of the system: the focal length f of the lenslet and the pixel pitch Δ of the CMOS image sensor.



Supplementary Figure 2. Side view schematic of the wavefront sensor. f denotes the focal length of the lenslet, and D denotes the diameter of the lenslet. The black dashed lines indicate the allowed displacement region for the double-helix focal spot of each lenslet. θ_{min} represents the minimum resolvable angle, and Δ is the detector pixel pitch.

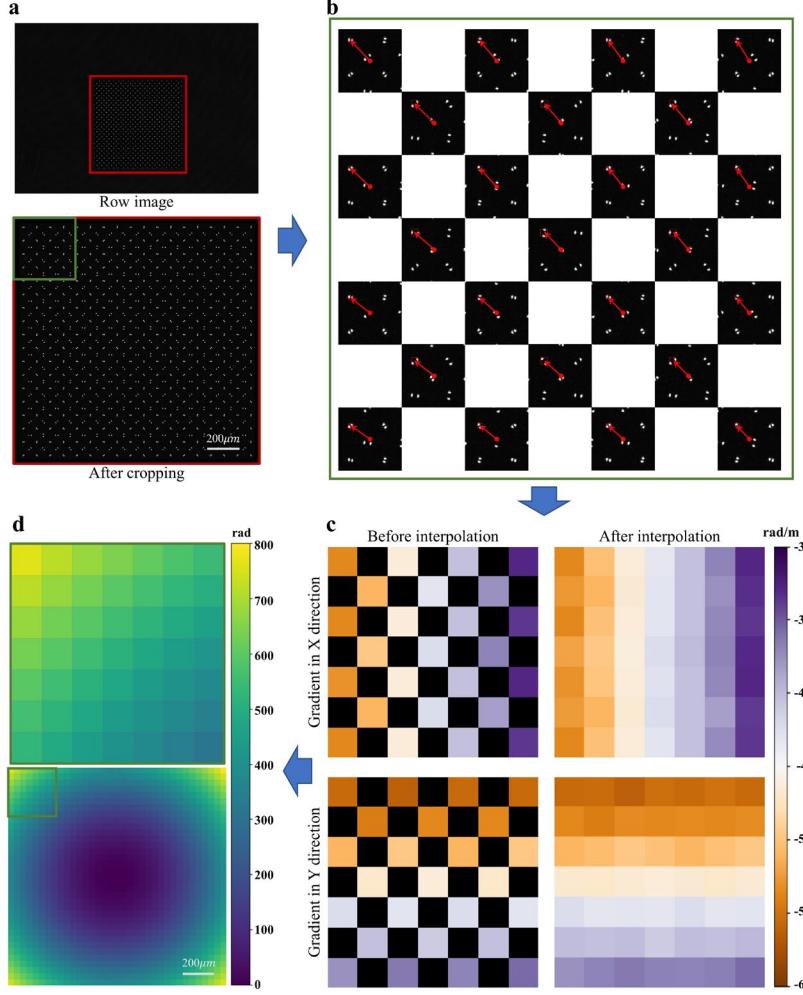
For a Shack–Hartmann type wavefront sensor as shown in Supplementary Figure 2, the local wavefront tilt θ is encoded as the displacement of the focal spot at the detector plane. Under the small-angle approximation (valid for $\theta < 10^\circ$, which holds for all measurements in this work), the smallest resolvable angle is given by $\theta_{min} = \arctan(\Delta/f)$. The minimum detectable displacement is limited by the pixel pitch $\Delta = 1.67 \mu\text{m}$ of the CMOS sensor (Imaging source: DMM 27UJ003-ML). The focal length f in our design is $470 \mu\text{m}$, yielding $\theta_{min} \approx 0.2^\circ$. The corresponding minimum resolvable phase gradient is $k \sin(\theta_{min})$, with the wave number $k = 2\pi/\lambda$. At the operating wavelength $\lambda = 532 \text{ nm}$, we obtain a minimum detectable gradient of approximately $4.2 \times 10^4 \text{ rad/m}$.

This sensitivity limit represents the fundamental lower bound imposed by detector pixelation and optical geometry under ideal conditions. Together with the upper dynamic range bound ($\pm 9.4^\circ$) discussed in the main text, these two metrics characterize the measurement capabilities of the sensor. Supplementary Figure 2 also illustrates the allowable displacement range of the focal spot after orientation encoding of the DH-PSF, as indicated by the black dashed lines. Compared to the conventional Shack–Hartmann limit of D , this range is extended to $2D$. The relationship also provides a design guideline: improving sensitivity requires either a longer focal length or a smaller effective pixel pitch, though both choices may trade off against other performance parameters such as device footprint or dynamic range.

Supplementary Note 4. Wavefront Reconstruction Algorithm

4.1 Algorithm description.

As illustrated in Supplementary Figure 3(a–d), the complete wavefront reconstruction process consists of four sequential steps: image registration and cropping, channel-wise DH-PSF localization, gradient formation with neighborhood interpolation, and TV-regularized Poisson integration. This framework follows the general principle of gradient-based wavefront reconstruction commonly used in Shack–Hartmann sensors [1-2], with modifications introduced here for dual-channel DH-PSF localization and sampling-density enhancement.



Supplementary Figure 3. Pipeline of the wavefront reconstruction algorithm. **a** Dual-channel image registration and cropping based on the plane-wave DH-PSF array. **b** Sub-pixel DH-PSF localization via rotation-bank correlation within $\pm 22.5^\circ$, yielding local gradients for each aperture. **c** Neighborhood interpolation applied to the interleaved gradient field to obtain a complete and continuous gradient map. **d** Total-variation-regularized Poisson integration reconstructing the final phase map.

(a) Image registration and cropping.

Both polarization channels are simultaneously captured under identical illumination. Before patch-level analysis, all raw frames are rigidly aligned to a reference image obtained under plane-wave incidence, ensuring consistent spatial registration across channels. The registered image is then cropped according to the DH-PSF array size determined from the plane-wave reference, guaranteeing that each cropped region precisely corresponds to a single sub-aperture. This alignment ensures accurate localization and subsequent gradient mapping for all DH-PSFs.

(b) DH-PSF localization.

For each sub-aperture, DH-PSF localization is performed independently in the two polarization channels using normalized cross-correlation with a compact rotation bank spanning $\pm 22.5^\circ$. The best-matching orientation and sub-pixel centroid are jointly determined, from which the local PSF displacement is extracted. Through calibration with

known system parameters (pixel pitch, magnification, wavelength, and DH-PSF sensitivity), the displacement vector is directly converted to local gradients (g_x, g_y) . Although both channels measure full two-dimensional gradients, their interleaved aperture layout effectively doubles the spatial sampling density without redundancy.

(c) Gradient interpolation.

Because the dual-channel sampling follows an interleaved aperture pattern, the recovered gradient field is spatially incomplete, with unmeasured nodes between alternating sub-apertures. To obtain a continuous field suitable for integration, the missing values are filled using local neighborhood averaging of valid four-neighbors for each gradient component. This lightweight interpolation preserves local smoothness and structural consistency while avoiding global bias.

(d) TV-regularized Poisson reconstruction.

The phase map $\phi(x, y)$ is reconstructed by minimizing the discrepancy between the estimated and observed gradients under total-variation regularization, following the standard TV-regularized Poisson framework [3-4]. The optimization problem is formulated as

$$\min_{\phi} \|\nabla_x \phi - g_x\|_2^2 + \|\nabla_y \phi - g_y\|_2^2 + \lambda TV(\phi) \quad (5)$$

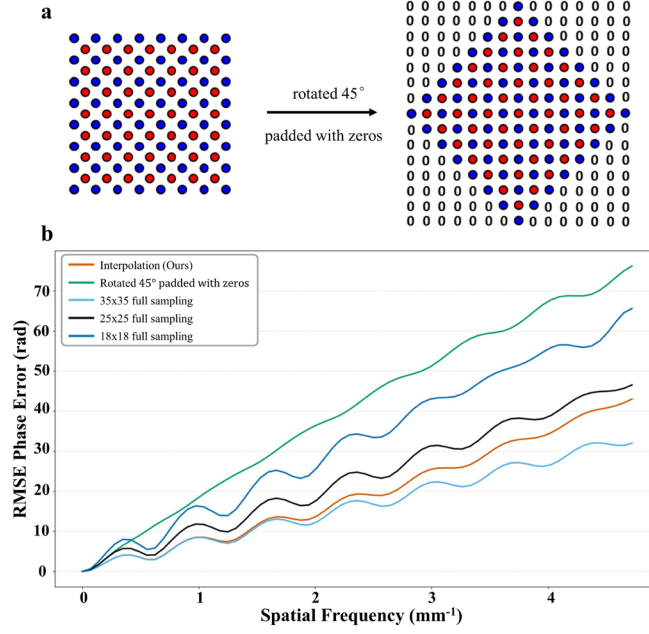
subject to Neumann boundary conditions, and solved using a convex optimization scheme implemented in an Alternating Direction Method of Multipliers (ADMM) solver [5]. The TV term stabilizes inversion against noise and missing samples, yielding an edge-preserving, physically consistent phase map that captures both global curvature and localized aberrations.

4.2 Validation of interpolation strategy.

To rigorously justify the choice of local 4-neighbor averaging and assess whether alternative approaches could eliminate the need for interpolation, a comprehensive quantitative comparison was conducted. One such alternative approach involves applying a 45° rotation coordinate transformation, which in principle maps the sampled points onto a uniform rectangular lattice.

The original interleaved sampling pattern retains only those positions where both coordinates (i, j) are either even or odd, forming a checkerboard-like grid. A coordinate transformation is applied to map these points onto a new uniform grid: $u=(i+j)/2$, $v=(i-j)/2+(N-1)/2$ with $N=35$, where $i, j=0, 1, \dots, N-1$ are the indices in the original Cartesian coordinate system. This transformation rotates the original grid by 45° and scales the coordinates appropriately. The gradient components are transformed accordingly: $g_u=(g_y+g_x)/\sqrt{2}$ and $g_v=(g_y-g_x)/\sqrt{2}$.

In the rotated coordinate system (u, v) , the valid data points occupy a diamond-shaped region corresponding to the original field of view rotated by 45°. Since numerical solvers require data on a rectangular grid, the remaining positions within the $N \times N$ matrix that fall outside this diamond-shaped region must be padded with zeros as shown in Supplementary Figure 4(a). This zero-padding is necessary to maintain a rectangular computational domain but introduces artificial boundaries where the gradient field discontinuously drops to zero.



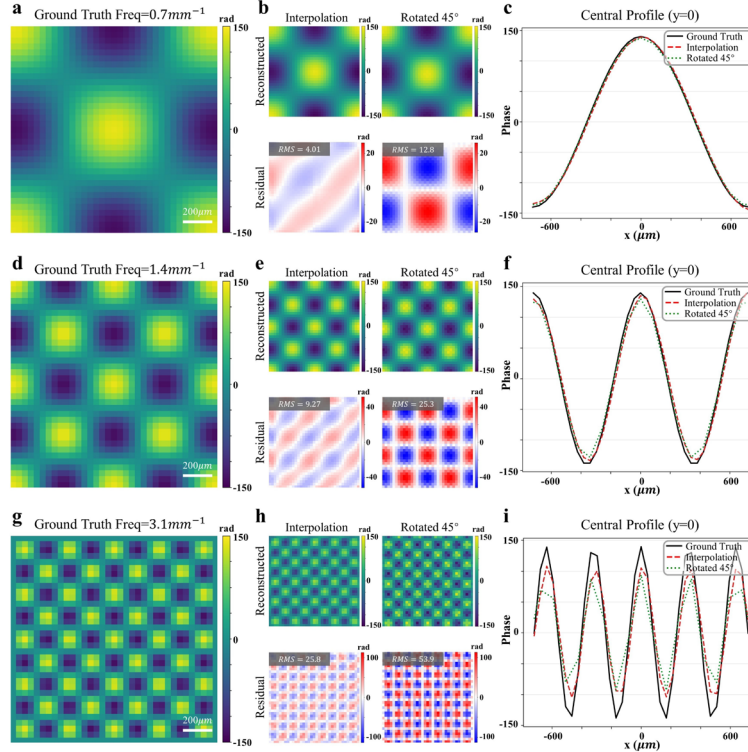
Supplementary Figure 4. Comparison of different gradient handling strategies for wavefront reconstruction. **a** Schematic illustration of the 45° rotation method with zero padding. **b** The root mean square errors (RMSE) of the reconstructed phase as a function of spatial frequency. The light blue curve corresponds to the ideal 35×35 full sampling benchmark representing the native sampling density. The orange curve shows the interpolation method used in this work, where missing gradients at interleaved sampling positions are filled by local 4-neighbor averaging. The green curve represents the 45° rotation method with zero padding. The black curve shows the ideal 25×25 rotated sampling with physical spacing of $40\sqrt{2}$ μm and without zero padding, which serves to isolate the effect of zero padding. The dark blue curve corresponds to the 18×18 single-channel sampling in vector mode, representing the reduced resolution case.

For comparison, five reconstruction scenarios were evaluated: (1) the interpolation method used in this work where missing gradients are filled by averaging the four nearest valid neighbors; (2) the 45° rotation method with zero-padding as described above; (3) ideal 35×35 full sampling using the complete Cartesian grid without missing data; (4) vector-mode single-channel sampling on an 18×18 uniform grid; and (5) ideal rotated 25×25 sampling on a uniform grid with $40\sqrt{2}$ μm spacing covering the entire rotated field of view without zero-padding, serving to isolate the effect of zero-padding from the coordinate transformation itself.

A series of two-dimensional cosine phase objects was generated as $\phi(x,y)=A\cos(2\pi fx)\cos(2\pi fy)$ with amplitude $A=140$ rad and spatial frequency f ranging from 0.1 to 4.7 mm⁻¹, matching the sensor field of view of 1440 μm × 1440 μm. The true gradient fields were computed analytically and sampled according to the actual sub-aperture locations of our sensor.

The root-mean-square errors (RMSE) of the reconstructed phase as a function of spatial frequency are plotted in Supplementary Figure 4(b). In the low frequency region below 0.5 mm⁻¹, all methods exhibit very small errors closely following the 35×35 full sampling curve, indicating that for smooth wavefronts the choice of interpolation strategy has negligible impact. As frequency increases into the intermediate range from approximately 0.5 to 2 mm⁻¹, the 18×18 single-channel reconstruction and the ideal rotated 25×25 sampling show progressively increasing errors due to under-sampling. The 45° rotation method with zero-padding exhibits a rapid error increase that even exceeds the 25×25 sampling curve, despite theoretically containing identical physical information. In

contrast, the interpolation method maintains errors remarkably close to the 35×35 baseline and consistently below the 25×25 sampling curve, demonstrating that local averaging effectively recovers missing gradient information and enables reconstruction quality approaching that of 35×35 full sampling. At high frequencies above 2 mm^{-1} , the interpolation method eventually deviates from the 35×35 baseline due to the inherent low-pass filtering effect of local averaging, but its error remains substantially lower than both the 25×25 sampling curve and, more notably, far below the 18×18 single-channel reconstruction. This directly validates the advantage of dual-channel polarization multiplexing: physically sampling more information through two interleaved channels achieves reconstruction fidelity that greatly exceeds what is possible with a single channel of comparable pitch.



Supplementary Figure 5. Comparison of wavefront reconstruction accuracy between interpolation method and 45° rotation method at different spatial frequencies. **a** Ground truth wavefront at spatial frequency 0.7 mm^{-1} . **b** Top row: reconstructed phases using the interpolation method (left) and the 45° rotation method (right); bottom row: corresponding residuals relative to ground truth. **c** Phase profiles along $y = 0$ comparing the interpolation method reconstruction, the 45° rotation method reconstruction, and the ground truth at 0.7 mm^{-1} . **d** Ground truth wavefront at spatial frequency 1.4 mm^{-1} . **e** Same as **b** but for 1.4 mm^{-1} . **f** Same as **c** but for 1.4 mm^{-1} . **g** Ground truth wavefront at spatial frequency 3.1 mm^{-1} . **h** Same as **b** but for 3.1 mm^{-1} . **i** Same as **c** but for 3.1 mm^{-1} .

The performance degradation of the 45° rotation method is attributed to algorithmic artifacts arising from the zero-padding necessary to fit the diamond-shaped valid region into a rectangular matrix for the solver. These padded boundaries introduce artificial gradient discontinuities that propagate through the reconstruction and degrade overall accuracy, an effect confirmed by the superior performance of the ideal rotated sampling without zero-padding. For a more intuitive comparison, Supplementary Figure 5 presents reconstruction results at three representative spatial frequencies (0.7 , 1.4 , and 3.1 mm^{-1}), showing ground truth wavefronts, reconstructions from both methods, corresponding

residuals, and phase profiles along $y = 0$. These visual comparisons corroborate the quantitative findings.

In summary, whether the sensor employs the current rectangular checkerboard layout or a rotated 45° arrangement, the physical wavefront information collected remains identical. The choice of local 4-neighbor interpolation is motivated by the current hardware geometry and the observation that this algorithmic treatment introduces relatively less error compared to the zero-padding artifacts associated with post-acquisition rotation.

This analysis also suggests an intriguing direction for future hardware design. Implementing a 45° rotation directly at the hardware level, where the orientation and spacing of the DH-PSF are rearranged into a rectangular uniform sampling grid instead of the current checkerboard layout, would eliminate the need for post-processing rotation in the algorithm. This approach avoids errors from zero padding the diamond-shaped valid region while preserving the underlying physical sampling information, thereby achieving comparable reconstruction accuracy with a simplified pipeline.

Supplementary Note 5. Quantitative Analysis of DH-PSF Decoding Robustness

The robustness of our dual-encoding strategy, which combines four orientation angles with two distinct lobe spacings to enable simultaneous dual-channel operation, was systematically evaluated against potential sources of decoding errors including noise, aberrations, and fabrication imperfections.

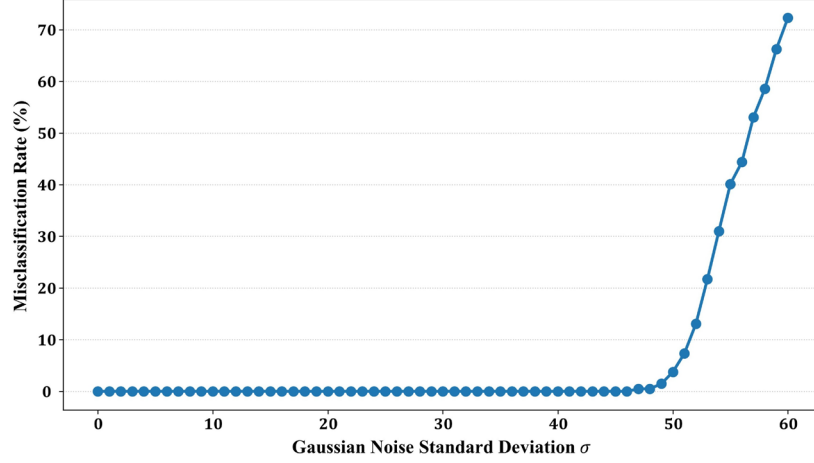
To quantify polarization leakage between channels under noisy conditions, an experimental image containing all eight DH-PSF patterns was corrupted with increasing levels of Gaussian noise. Specifically, for each pixel at coordinate (x, y) , the noisy intensity is generated as:

$$I_{noisy}(x, y) = I_{origin}(x, y) + N(x, y), \quad (6)$$

where $N(x, y) \sim \mathcal{N}(0, \sigma^2)$ is an independent zero-mean Gaussian random variable with standard deviation σ . The resulting noisy image is then clipped (truncated) to the standard 8-bit dynamic range $[0, 255]$, yielding the final pixel intensity:

$$I_{final}(x, y) = \max(0, \min(255, I_{noisy}(x, y))). \quad (7)$$

Then, it was processed through the complete decoding pipeline incorporating template matching with $\pm 22.5^\circ$ rotation tolerance. The misclassification rate as a function of noise standard deviation σ is plotted in Supplementary Figure 6, demonstrating that the error rate remains essentially zero for $\sigma \leq 50$ (in units of image gray levels), confirming robust channel discrimination under realistic noise conditions. The error rate rises sharply only above this threshold, establishing the noise floor of the current implementation. More complex aberrations that may affect both spacing and orientation were further investigated through a systematic Zernike-based analysis.



Supplementary Figure 6. Polarization channel misclassification rate as a function of added Gaussian noise.

Since the first three Zernike terms either carry no phase information (piston, Z1) or correspond to the quantities we aim to measure (tip/tilt, Z2/Z3 are encoded as PSF displacement), the analysis focused on the next four primary aberrations: defocus (Z4), astigmatism (Z5/Z6), coma (Z7/Z8), and trefoil (Z9/Z10). Following the indexing and normalization convention defined by Noll [6]. Given that the eight DH-PSF patterns already cover two orthogonal orientation sets ($0^\circ/90^\circ$ and $45^\circ/135^\circ$) for both lobe spacings, one representative orientation was selected for each aberration type: defocus (Z4: $\sqrt{3}(2\rho^2-1)$), astigmatism (Z5: $\sqrt{6}\rho^2\sin 2\phi$), coma (Z7: $\sqrt{8}(3\rho^3-2\rho)\sin \phi$), and trefoil (Z9: $\sqrt{8}\rho^3\sin 3\phi$), where (ρ, ϕ) are polar coordinates in the pupil plane with ρ denoting the normalized radial coordinate ($0 \leq \rho \leq 1$). For each case, a single sub-aperture containing all eight PSF patterns was simulated with increasing Zernike coefficient magnitudes. These failure thresholds are summarized in Table 1, providing quantitative guidelines for allowable aberration levels in practical applications.

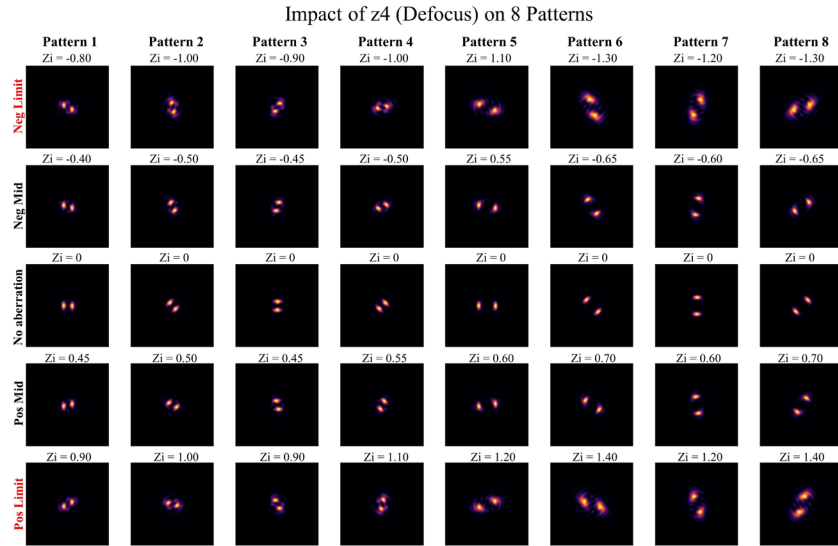
Table 1. Tolerance ranges of Zernike coefficients for the eight DH-PSF patterns.

Pattern	Defocus (z4)	Astigmatism (z5)	Coma (z7)	Trefoil (z9)
Pattern 1	-0.8 ~ 0.9	-0.6 ~ 0.5	-1.0 ~ 1.3	-0.9 ~ 0.9
Pattern 2	-1.0 ~ 1.0	-1.0 ~ 1.1	-0.8 ~ 0.8	-0.9 ~ 0.9
Pattern 3	-0.9 ~ 0.9	-0.5 ~ 0.6	-0.7 ~ 0.7	-0.8 ~ 0.8
Pattern 4	-1.0 ~ 1.1	-1.2 ~ 1.0	-0.9 ~ 0.9	-0.8 ~ 0.8
Pattern 5	-1.1 ~ 1.2	-0.8 ~ 0.7	-1.1 ~ 1.1	-1.3 ~ 1.3
Pattern 6	-1.3 ~ 1.4	-1.3 ~ 1.4	-1.2 ~ 1.9	-1.3 ~ 1.3
Pattern 7	-1.2 ~ 1.2	-0.7 ~ 0.8	-1.0 ~ 1.0	-1.2 ~ 1.0
Pattern 8	-1.3 ~ 1.4	-1.4 ~ 1.3	-2.0 ~ 2.0	-1.2 ~ 1.2

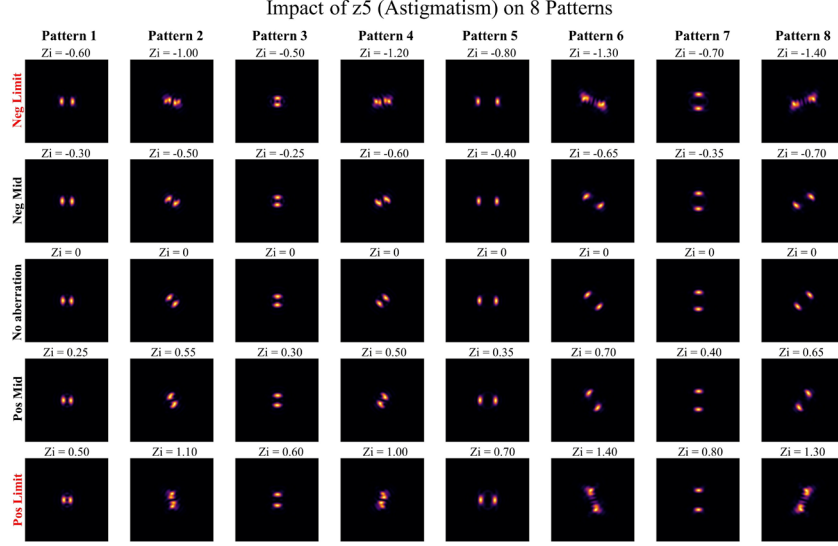
Supplementary Figure 7, Supplementary Figure 8, Supplementary Figure 9, Supplementary Figure 10 illustrates the effect of defocus, astigmatism, coma, and trefoil, respectively, on the intensity profiles of all eight PSF patterns. In each figure, every column corresponds to one PSF pattern, and each row represents a different Zernike coefficient magnitude. The middle row shows the aberration-free profile, the top and bottom rows correspond to the negative and positive failure thresholds respectively, and the second and

fourth rows display intermediate aberration levels. From these images, distinct failure mechanisms can be observed across aberration types: defocus primarily influences the orientation angle; astigmatism mainly affects lobe spacing for patterns 2, 4, 6, and 8 while impacting both orientation and spacing for patterns 1, 3, 5, and 7; coma and trefoil have more pronounced effects on overall lobe morphology. Because these effects are often coupled, particularly for higher-order aberrations, unified empirical thresholds were adopted rather than attempting to disentangle individual mechanisms.

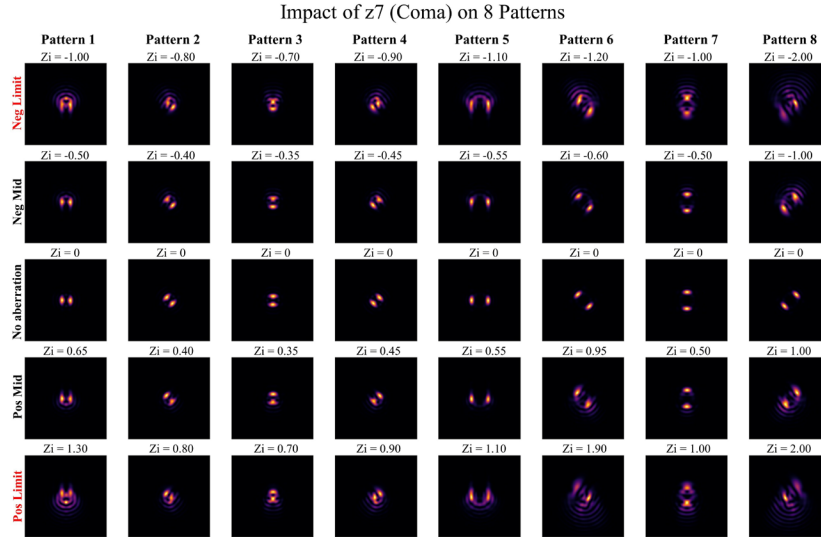
Regarding fabrication-induced phase errors, small deviations in nanopillar dimensions during electron-beam lithography can introduce either increased background noise or localized initial aberrations. As demonstrated in the noise robustness analysis, decoding accuracy remains unaffected as long as noise stays below the established threshold. For induced initial aberrations, a calibration step with plane-wave illumination is performed prior to each measurement to record reference PSF positions, effectively compensating for any systematic offsets. Combined with the strong tolerance to spacing and orientation variations demonstrated in the aberration analysis, fabrication errors within typical specifications (critical dimension variations below 10 nm) have negligible impact on decoding accuracy.



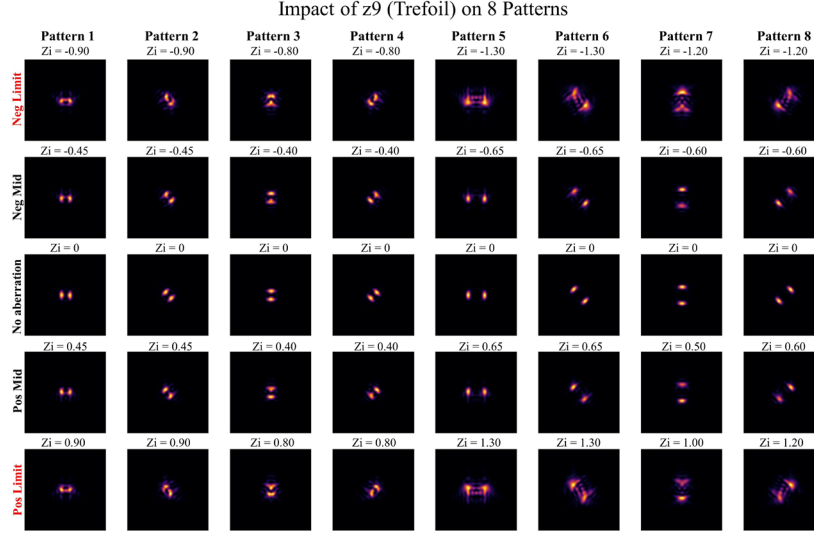
Supplementary Figure 7. Effect of defocus on the intensity profiles of eight DH-PSF patterns. Each column corresponds to one of the eight PSF patterns (four orientations $\times$ two lobe spacings). Each row represents a different Zernike coefficient magnitude for defocus (Z_4). The middle row shows the aberration-free profile, the top and bottom rows correspond to the negative and positive failure thresholds respectively, and the second and fourth rows display intermediate aberration levels.



Supplementary Figure 8. Effect of astigmatism on the intensity profiles of eight DH-PSF patterns. Each column corresponds to one of the eight PSF patterns (four orientations $\times$ two lobe spacings). Each row represents a different Zernike coefficient magnitude for astigmatism (Z_5). The middle row shows the aberration-free profile, the top and bottom rows correspond to the negative and positive failure thresholds respectively, and the second and fourth rows display intermediate aberration levels.



Supplementary Figure 9. Effect of coma on the intensity profiles of eight DH-PSF patterns. Each column corresponds to one of the eight PSF patterns (four orientations $\times$ two lobe spacings). Each row represents a different Zernike coefficient magnitude for coma at (Z_7). The middle row shows the aberration-free profile, the top and bottom rows correspond to the negative and positive failure thresholds respectively, and the second and fourth rows display intermediate aberration levels.



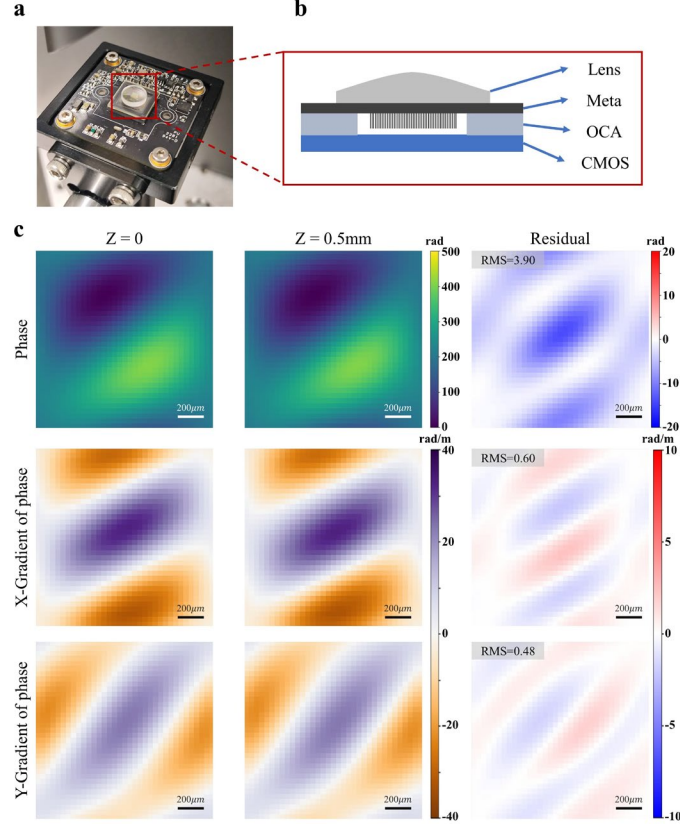
Supplementary Figure 10. Effect of trefoil on the intensity profiles of eight DH-PSF patterns. Each column corresponds to one of the eight PSF patterns (four orientations $\times$ two lobe spacings). Each row represents a different Zernike coefficient magnitude for trefoil at (Z_9). The middle row shows the aberration-free profile, the top and bottom rows correspond to the negative and positive failure thresholds respectively, and the second and fourth rows display intermediate aberration levels.

Supplementary Note 6. Substrate Propagation Correction

The metasurface is fabricated on a fused-silica substrate with a thickness of $500 \mu\text{m}$. Since the reconstructed phase is derived from the measured local gradients of the transmitted wavefront, the optical propagation through this finite-thickness substrate inevitably alters the retrieved phase distribution. This effect becomes particularly relevant when characterizing test lenses with known optical parameters, where accurate phase-to-height conversion is required.

Numerical simulations were performed using the angular spectrum propagation method [7] to analyze the propagation-induced deviation through the $500 \mu\text{m}$ substrate. A plane wave was propagated through a $500 \mu\text{m}$ layer using the angular spectrum method, and the phase and phase-gradient distributions were compared before and after transmission. The results, summarized in Supplementary Figure 11, reveal distinct discrepancies in both phase and gradient maps, demonstrating that free-space propagation through the substrate causes a non-negligible bias in the reconstructed wavefront.

To recover the intrinsic wavefront immediately at the metasurface plane, the experimentally measured phase was numerically back-propagated by $500 \mu\text{m}$ using the inverse angular spectrum propagation operator. This correction effectively compensates for the curvature and gradient distortion introduced by the substrate, yielding an accurate representation of the true exit wavefront from the test object.



Supplementary Figure 11. Substrate propagation effect and correction. **a** Photograph of the experimental setup for lens measurement. **b** Schematic illustration of the 500 μm substrate propagation and the corresponding numerical back-propagation. **c** Comparison of phase and phase-gradient distributions before and after 0.5 mm propagation: the 3×3 panels show phase (top), g_x (middle), g_y (bottom) at $z = 0$, $z = 0.5\text{ mm}$, and the corresponding residuals. Propagation induces noticeable deviations, which are eliminated by numerical back-propagation.

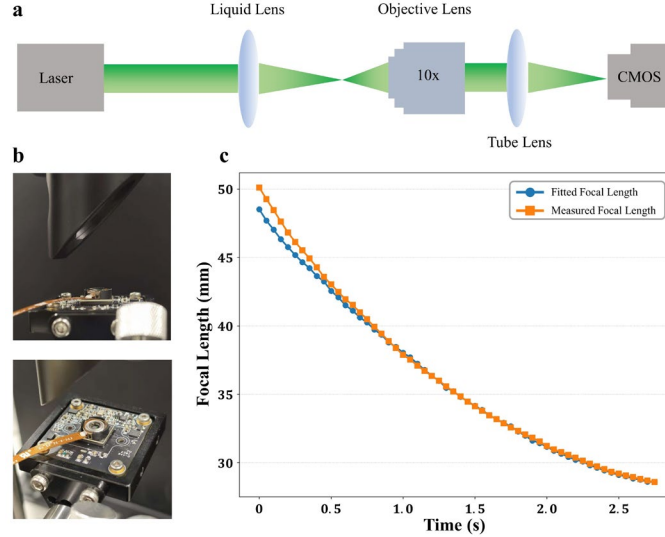
A quantitative comparison was carried out using a plano-convex lens as a reference sample. The curvature radius calculated from the uncorrected wavefront was consistently smaller than the nominal design value, while the result derived from the back-propagated phase showed excellent agreement. The numerical values are summarized in Table 2, confirming that the 500 μm back-propagation correction substantially improves the accuracy of curvature retrieval.

Table 2. Comparison of retrieved curvature radii before and after 500 μm back-propagation.

Sample ID	Nominal radius (mm)	Retrieved radius without correction (mm)	Retrieved radius after correction (mm)	Relative error before (%)	Relative error after (%)
Lens #1	5.17	5.148	5.180	0.417	0.192
Lens #2	5.17	5.135	5.168	0.679	0.046
Lens #3	5.17	5.150	5.184	0.385	0.264
Lens #4	6.46	6.428	6.476	0.494	0.250
Lens #5	8.26	8.196	8.240	0.773	0.238
Lens #6	10.36	10.260	10.362	0.962	0.019
Lens #7	12.91	12.645	12.848	2.052	0.481

Supplementary Note 7. Dynamic Wavefront Reconstruction Demonstrations

The dynamic performance of the chip-integrated metasurface wavefront sensor was evaluated through two complementary approaches: a qualitative demonstration with an evolving water droplet (as presented in the main text) and a quantitative validation using a calibrated liquid lens.



Supplementary Figure 12. Experimental validation of dynamic wavefront measurement accuracy using a voltage-controlled liquid lens. **a** Schematic of the characterization setup for calibrating the relationship between applied voltage and focal length of the liquid lens. **b** Measurement configuration with the liquid lens placed directly on the sensor surface. **c** Comparison of focal lengths extracted from wavefront measurements (blue curve) and predicted from independent calibration (orange curve) as functions of time during a voltage sweep from 59 V to 70 V.

The water droplet experiment, shown in Fig. 5 of the main text, demonstrates the sensor's capability for capturing complex and unconstrained dynamic wavefronts. A deionized water droplet placed on the sensor surface was subjected to external vibration, inducing continuous deformation. The measurements were conducted in an offline imaging configuration, where the metasurface–CMOS assembly continuously captured DH-PSF frames at video frame rates. The recorded image sequences were processed using the reconstruction pipeline described in Section 4, including DH-PSF localization, gradient computation, and total-variation-regularized Poisson integration. For visualization purposes, the reconstructed wavefront maps, originally sampled on a 35×35 grid defined by the sub-aperture array, were spatially interpolated to a 350×350 grid. This upsampling facilitates smoother rendering of temporal phase variations without altering the underlying quantitative data. Supplementary Movies 1 shows the time-resolved reconstruction of the oscillating water droplet wavefront, illustrating consistent phase continuity across frames and faithful tracking of temporal wavefront evolution.

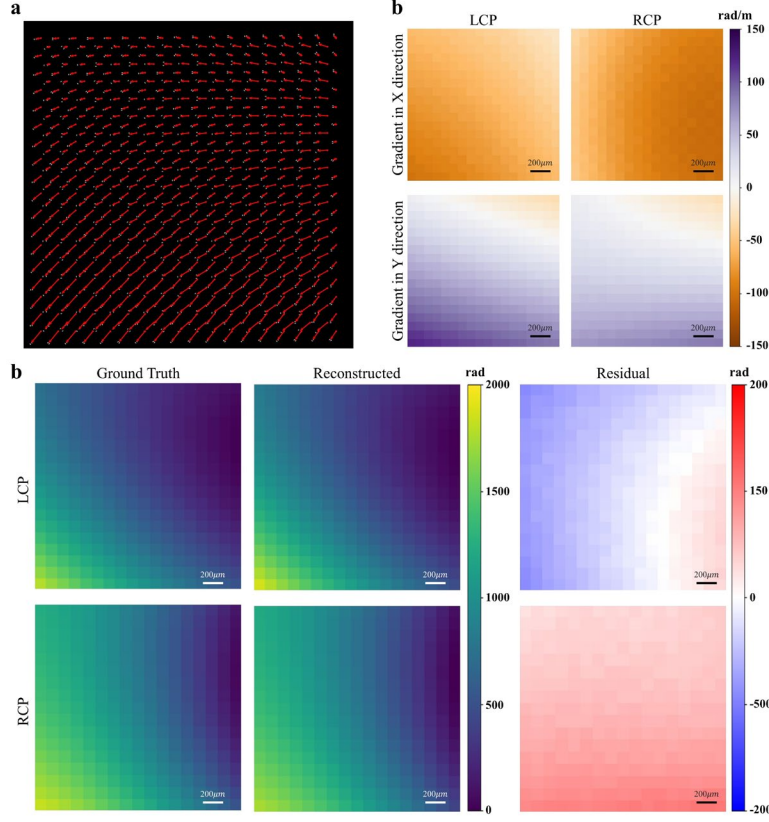
To complement this qualitative demonstration with a quantitative assessment of the sensor's accuracy under dynamic conditions involving a liquid-based optical element, an additional validation experiment was performed using a commercially available voltage-controlled liquid lens (Corning® Varioptic® A-25H1-D0-33). This lens operates on the electrowetting principle, where an applied voltage modifies the curvature of a liquid-liquid interface, enabling precise and repeatable focal length tuning without mechanical moving

parts. First, a dedicated characterization setup as shown in Supplementary Figure 13(a) was built to establish the relationship between applied voltage (from 59 V to 70 V) and the resulting focal length of the liquid lens. For each voltage increment, a motorized translation stage (Thorlabs, MTS50/M-Z8) synchronously translated the microscope objective, tube lens, and CMOS imaging sensor to locate the new focal plane position, thereby determining the corresponding focal length with high precision. This calibration curve served as the ground truth for subsequent comparison. Next, the liquid lens was placed directly on the surface of the sensor, replicating the same measurement configuration as in the water droplet experiment, as illustrated in Supplementary Figure 13(b). A linearly varying voltage from 59 V to 70 V was applied over time, while the transmitted wavefront was continuously recorded at 20 frames per second via the DH-PSF array. The captured image sequence was processed offline using the same reconstruction pipeline, yielding the time-evolving phase distribution, from which the instantaneous focal length was derived. The results are presented in Supplementary Figure 13(c). The blue curve shows the focal length extracted from the sensor's wavefront measurements as a function of time, while the orange curve represents the ideal focal length predicted by the independently calibrated voltage-to-focal-length relationship. Excellent agreement is observed throughout the entire voltage sweep. Supplementary Movies 2 presents the corresponding time-resolved wavefront reconstruction of the liquid lens during the voltage sweep, further illustrating the dynamic measurement capability.

Together, the qualitative water droplet demonstration and the quantitative liquid lens validation confirm that the integrated sensing architecture supports high-fidelity spatiotemporal wavefront recovery, highlighting its robustness for dynamic optical field analysis and its potential for future real-time operation with hardware acceleration.

Supplementary Note 8. Compatible Dual-Channel Vector Wavefront Sampling

The metasurface wavefront sensor is intrinsically compatible with vectorial wavefront measurement through its two orthogonally polarized DH-PSF channels. In the standard scalar mode, both channels operate simultaneously in an interleaved configuration, which effectively doubles the spatial sampling density and allows reconstruction of a single scalar wavefront after interpolation. Beyond this operation mode, the same system also supports polarization-resolved wavefront analysis, in which the two channels respectively correspond to the left- and right-circularly polarized (LCP and RCP) components of the optical field.



Supplementary Figure 13. Compatible dual-channel vector wavefront measurement. **a** Localization results of DH-PSFs in the LCP and RCP channels. **b** Reconstructed phase-gradient maps (g_x, g_y) for LCP and RCP components. **c** Comparison between ground-truth, reconstructed, and residual wavefronts for the two polarization channels.

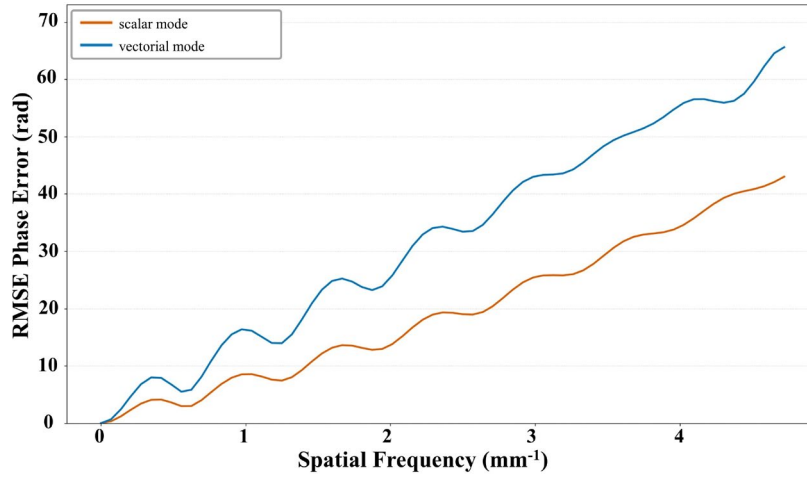
In the polarization-resolved measurement, the two channels are acquired and localized simultaneously from the same exposure, ensuring strict spatial correspondence between sub-apertures. The localization step jointly detects all DH-PSFs across both polarization channels, while the extracted gradient information is subsequently separated according to channel identity for independent phase reconstruction. Supplementary Figure 13 summarizes the complete process: Supplementary Figure 13(a) shows the DH-PSF localization results for both channels; Supplementary Figure 13(b) presents the corresponding phase-gradient maps (g_x, g_y) for LCP and RCP illumination, confirming that each polarization channel independently encodes the local wavefront slope; and Supplementary Figure 13(c) compares the ground-truth, reconstructed, and residual phase maps for LCP and RCP components, demonstrating accurate polarization-resolved reconstruction and quantitative agreement with the reference wavefronts.

When operated in this vectorial mode, each channel effectively occupies half of the spatial sampling positions, reducing the sampling areal density by a factor of two compared to the combined scalar configuration. Notably, if the interleaved sampling grid (comprising positions where both coordinates are even or both odd) is rotated by 45° , the resulting lattice exhibits a uniform spacing of $40 \mu\text{m}$, corresponding to a linear sampling density improvement by a factor of $\sqrt{2}$ relative to the $80 \mu\text{m}$ pitch of a single channel in vectorial mode. Through subsequent interpolation and reconstruction, the scalar mode achieves an

effective resolution improvement of approximately twofold over the raw vectorial mode sampling.

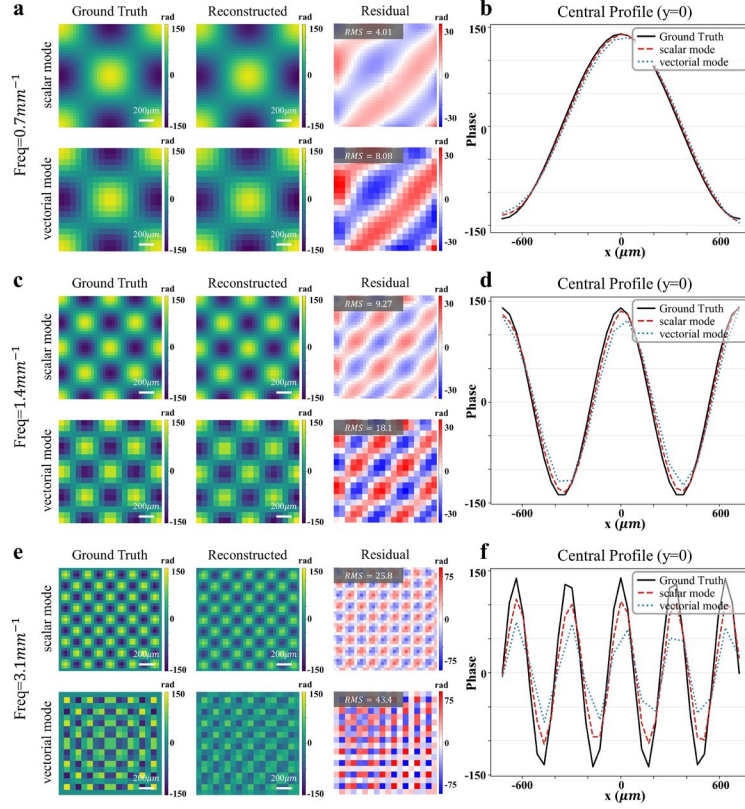
To quantitatively characterize the performance penalty when switching to the "vectorial mode.", a simulation study was conducted comparing wavefront reconstruction accuracy between scalar and vectorial modes across a range of spatial frequencies.

Two-dimensional sinusoidal phase objects were generated as $\phi(x,y)=A\cos(2\pi fx)\cos(2\pi fy)$ with amplitude $A=140$ rad and spatial frequency f ranging from 0.1 to 3.5 mm^{-1} . The field of view was set to $1440 \mu\text{m} \times 1440 \mu\text{m}$, matching the sensor dimensions, and the wavelength was 532 nm. The true gradient fields were sampled according to the actual sub-aperture locations of both operating modes: scalar mode (35×35 interleaved dual-channel sampling) and vectorial mode (18×18 single-channel sampling). Phase reconstruction was performed using the same TV-regularized Poisson solver employed throughout this work.



Supplementary Figure 14. Root-mean-square error (RMSE) of reconstructed wavefronts as a function of spatial frequency. The orange curve corresponds to scalar mode, while the blue curve corresponds to vectorial mode.

Supplementary Figure 14 presents the root-mean-square error (RMSE) of the reconstructed wavefronts as a function of spatial frequency. The orange curve corresponds to scalar mode, while the blue curve corresponds to vectorial mode. For a more intuitive comparison, Supplementary Figure 15 displays the reconstruction results at three representative spatial frequencies. At spatial frequency 0.7 mm^{-1} , Supplementary Figure 15(a) shows the ground truth, reconstructed wavefront, and residual for vectorial mode (top row) and scalar mode (bottom row), with Supplementary Figure 15(b) presenting the corresponding phase profiles along $y = 0$. Supplementary Figure 15(c) and 15(d) provide the same comparison at 1.4 mm^{-1} , while Supplementary Figure 15(e) and 15(f) extend the comparison to 3.1 mm^{-1} .



Supplementary Figure 15. Comparison of scalar and vectorial mode reconstruction accuracy at different spatial frequencies. **a** At spatial frequency 0.7 mm^{-1} , ground truth, reconstructed wavefront, and residual for scalar mode (top row) and vectorial mode (bottom row). **b** Phase profiles along $y = 0$ comparing scalar mode reconstruction, vectorial mode reconstruction, and ground truth at 0.7 mm^{-1} . **c** Same as **a** but for spatial frequency 1.4 mm^{-1} . **d** Corresponding phase profiles along $y = 0$ at 1.4 mm^{-1} . **e** Same as **a** but for spatial frequency 3.1 mm^{-1} . **f** Corresponding phase profiles along $y = 0$ at 3.1 mm^{-1} .

These results demonstrate that switching from scalar to vectorial mode incurs only a modest performance penalty at low spatial frequencies. At intermediate and high frequencies, however, the degradation becomes progressively more pronounced, consistent with the reduced sampling density in vectorial operation. This quantitative characterization provides a clear assessment of the trade-off involved when choosing between high-density scalar wavefront measurement and polarization-resolved vector sensing within the same unified metasurface architecture.

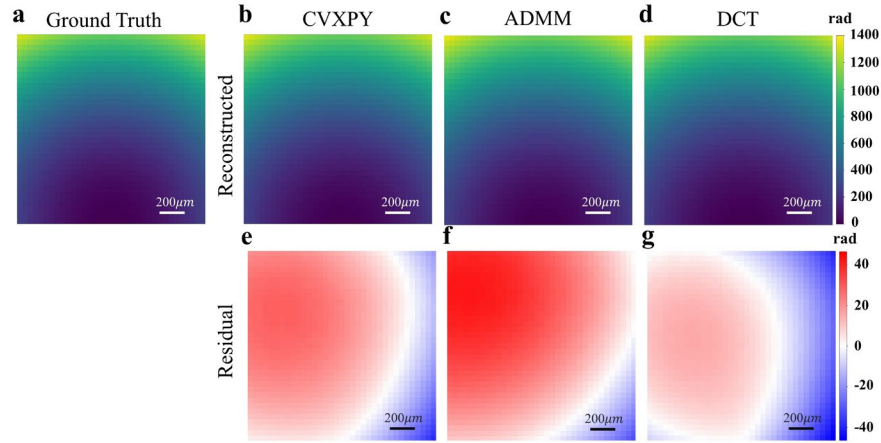
Supplementary Note 9. Computational Complexity and Real-Time Implementation Analysis

To assess the real-time potential of the proposed wavefront sensing method for adaptive optics applications, a detailed analysis of computational complexity and optimization pathways was conducted. The entire reconstruction pipeline is divided into two distinct stages: gradient extraction (DH-PSF localization and slope computation) and phase retrieval (TV-regularized Poisson integration). Several metrics are adopted to evaluate real-time capability, including single-frame processing time, frame rate, speedup relative to baseline implementations, and theoretical algorithmic complexity, while reconstruction quality is assessed through residual maps comparing recovered phases against ground truth.

Table 3. Performance Benchmark of Reconstruction Methods

Method	Time(ms)	Speedup (vs CVXPY)	Estimated FPS	Complexity Class
CVXPY (TV)	59.43	1.0x	~16.8	$O(N^3)$
ADMM (TV)	5.05	11.8x	~198.1	$O(kN\log N)$
DCT (Poisson)	0.07	798.5x	>13000	$O(N\log N)$

For the phase retrieval step, three representative solvers were benchmarked on a system equipped with an AMD Ryzen 5 9600X 6-core processor, with results summarized in Table 3. The CVXPY [8] general-purpose convex solver, despite its straightforward implementation, employs interior-point methods with polynomial complexity, requiring 59.43 ms per frame (approximately 16.8 FPS) for 35×35 phase maps. While adequate for offline analysis, this falls short of real-time feedback requirements. To overcome this limitation, an Alternating Direction Method of Multipliers (ADMM) solver specifically tailored for TV regularization was implemented. By decomposing the original problem into efficiently solvable subproblems, each iteration requires only one fast DCT and one soft-thresholding operation, yielding $O(kN\log N)$ complexity. Under standard parameters, ADMM achieves 5.05 ms per frame (approximately 198 FPS), representing an $11.8\times$ speedup relative to CVXPY. Further acceleration is achieved through a Discrete Cosine Transform (DCT) based Poisson solver [9] that directly inverts the gradient equation in the frequency domain. With a single DCT and inverse DCT ($O(N\log N)$ complexity), this approach completes reconstruction in 0.07 ms (over 13,000 FPS), corresponding to a remarkable $798.5\times$ speedup. Critically, these speed improvements do not compromise accuracy. As shown in Supplementary Figure 16, which compares reconstructions from all three methods against ground truth, the residual maps exhibit minimal errors with no systematic bias. Notably, even the DCT method maintains excellent fidelity under the high signal-to-noise conditions of the measured gradient data, establishing it as a viable option for applications prioritizing speed.



Supplementary Figure 16. Comparison of phase reconstruction accuracy across different solvers. **a** Ground truth phase. **b** Reconstruction using CVXPY. **c** Reconstruction using ADMM. **d** Reconstruction using DCT. **e** Residual map for CVXPY reconstruction relative to ground truth. **f** Residual map for ADMM reconstruction relative to ground truth. **g** Residual map for DCT reconstruction relative to ground truth.

The gradient extraction step, which localizes spots by analyzing DH-PSF morphology and computes wavefront slopes, presents different optimization challenges. The initial serial implementation processed each sub-aperture independently, resulting in tens of seconds per frame and dominating system latency. Although asynchronous multi-threading reduced this to approximately 0.4 seconds, this stage remains the primary computational constraint. Importantly, this step exhibits inherent data parallelism, as each sub-image is processed independently with no cross-dependencies. However, the core template matching operation involves extensive sliding window searches that are computationally demanding on conventional Central Processing Units (CPUs). Offloading these calculations to massively parallel architectures such as Graphics Processing Units (GPUs) or Field-Programmable Gate Arrays (FPGAs), combined with optimized memory access patterns through CUDA or OpenCL, offers a clear path toward millisecond-scale processing and would effectively eliminate the remaining bottleneck.

In summary, the phase retrieval stage has already surpassed real-time requirements through algorithmic advances, achieving sub-millisecond latencies that far exceed both conventional video rates (30 FPS) and demanding kHz-level closed-loop specifications. The gradient extraction stage, while currently rate-limiting, poses no fundamental obstacle given its highly parallelizable structure and the availability of established hardware acceleration techniques. The practical realization of real-time performance therefore depends primarily on engineering implementation rather than algorithmic feasibility, confirming that the proposed method possesses full potential for real-time adaptive optics systems.

Supplementary Reference

1. Primot J. Theoretical description of Shack–Hartmann wave-front sensor[J]. *Opt. Commun.*, **222**(1-6): 81-92 (2003).
2. Tyson R K, Frazier B W. Principles of adaptive optics[M]. CRC press, 2022.
3. Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. *Physica D*, **60**(1-4): 259-268 (1992).
4. Agrawal A, Raskar R, Chellappa R. What is the range of surface reconstructions from a gradient field?[C]//European conference on computer vision. Berlin, Heidelberg: Springer Berlin Heidelberg, 578-591 (2006).
5. Boyd, S. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *FNT in Machine Learning* 3, 1–122 (2010).
6. Noll, R. J. Zernike polynomials and atmospheric turbulence. *J. Opt. Soc. Am.* **66**, 207 (1976).
7. Goodman J W. Introduction to Fourier optics[M]. Roberts and Company publishers, 2005.
8. Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging[J]. *J. Math. Imaging Vis.*, **40**(1): 120-145 (2011).
9. Briggs, W. L., Henson, V. E. & McCormick, S. F. A Multigrid Tutorial 2nd edn (Society for Industrial and Applied Mathematics, 2000).