

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

Corresponding Author: Professor Tao Li

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors report a chip-integrated metasurface Shack–Hartmann wavefront sensor (SHWFS) that addresses the fundamental trade-off between dynamic range and spatial sampling density. By utilizing a dual-encoding strategy—orientation-encoded double-helix point spread functions (DH-PSFs) and polarization-multiplexed channels—they demonstrate a significant increase in dynamic range (9.4um) and sampling density in an alignment-free, CMOS-bonded architecture. The experimental demonstrations, including 3D localization and video-rate droplet monitoring, are compelling. This work represents a notable step toward miniaturized metrology. I recommend its publication in Nature Communications once the authors clarify certain theoretical bounds and practical limitations.

1. While the dual-encoding strategy is effective, the authors should quantify the performance penalty when switching to the "vectorial mode." As noted in the Supplementary Information, lateral resolution drops by half in this mode; a more rigorous analysis of how increased DH-PSF complexity (higher vortex charge N) affects centroid localization precision under noise would strengthen the paper.
2. The current design is optimized for 532 nm. Given the dispersive nature of metasurfaces, the authors should discuss the system's susceptibility to chromatic aberrations or provide a roadmap for broadband operation (e.g., via achromatic metalens designs?).
3. The metasurface is bonded to the CMOS sensor via an optically transparent adhesive (OCA). The authors should comment on (or ideally provide data for) the thermal and mechanical stability of this bond. Specifically, how do temperature fluctuations affect the PSF orientation or the precise focal distance required for high-accuracy reconstruction?
4. The demonstration shows video-rate monitoring, but it appears to be processed offline. For applications in adaptive optics, real-time feedback is crucial. A brief discussion on the computational complexity of the TV-regularized Poisson integration and potential for hardware acceleration would be beneficial.
5. The manuscript would benefit from a more comprehensive discussion of the recent landscape in metasurface and nonlocal-filter-based wavefront sensing (Laser & Photonics Reviews 17, 2300426 (2023), Science Advances 10, eadl0501 (2024), Nature Photonics 19, 1315 (2025), National Science Review nwaf561 (2025)).

Reviewer #2

(Remarks to the Author)

In this work, authors propose a chip-integrated metasurface Shack–Hartmann-type wavefront sensor without external relay optics, claiming to overcome the dynamic-range vs sampling-density vs accuracy trade-off. The approach uses DH-PSFs with multiple orientation encodings to mitigate spot ambiguity under large slopes, and LCP/RCP polarization channels with distinct lobe separation to enable interleaved sampling that doubles effective sampling density. The manuscript reports $\pm 9.4^\circ$ dynamic range per polarization channel via tilt-plane-wave calibration, and demonstrates wavefront reconstruction using a pipeline involving PSF localization, gradient completion via neighborhood interpolation, and TV-regularized Poisson integration.

The core idea of orientation-encoded double-helix PSF and polarization-based interleaving to expand dynamic range while maintaining high sampling density is technically interesting and potentially impactful. However, several key claims are currently overstated or insufficiently supported, particularly regarding below:

1. To avoid ambiguity between adjacent spots under large wavefront tilts, the measurable tilt is restricted to half the

microlens diameter, resulting in a dynamic range of only about $\pm 1^\circ$. (from introduction)

As noted, this requires explicit assumptions (e.g., no advanced spot-matching, no multi-frame methods, particular pitch/focal length/NA regime). In general, the tilt limit scales with lenslet NA / f and detector allocation, so presenting the dynamic range of conventional SH sensor as a blanket constraint seems misleading.

For example, the commercially available SH sensor from OPTOCRAFT provides dynamic range up to 10 degrees.

2. "Stress-induced birefringence" is claimed but not convincingly demonstrated experimentally

The abstract and near end of the manuscript, it is implying to include "stress-induced birefringence."

The Supplement shows polarization-resolved wavefront reconstruction workflow (LCP vs RCP) and residual maps, but this reads as capability demonstration, not a clear stress-birefringence measurement with a stressed sample and a validated ground truth.

If described in abstract, I propose that you should either (i) add a direct experiment (stressed glass/plastic with known birefringence patterns, compare to polarimetry/known model), or (ii) tone down to "supports polarization-resolved wavefront sensing".

3. Concerns on reconstruction pipeline

The Supplement describes gradient completion by local 4-neighbor averaging to fill missing nodes due to interleaved sampling.

Given that the method's value proposition is quantitative metrology, you need to show why simpler coordinate transforms are not adequate in this sampling geometry (your Fig. S1 concern). Please quantify bias vs ground truth (especially for high spatial frequencies) and show impact on final phase error after TV-Poisson integration. For example, it looks to me if the data is 45 degree rotated, no interpolation is necessary.

4. Figures/residual maps are not interpretable enough

Some of the figures (especially for Fig. 3i,j , 4b) are visually unclear (colorbar/saturation), making them meaningless.

Especially for Fig. 3i case, I propose to adjust the contrast or choosing different focal length to show clear differences.

Please ensure wavefront/gradient maps have units, scale bars.

5. minors

a. The paper uses four principal orientation angles (0° , 90° , 45° , 135°) for advanced dynamic range. Right now, the manuscript asserts DH-PSF provides unambiguous features at large gradients, but does not quantify or explain why DH-PSF is the right PSF family and how performance scales with the number of distinct encodings. A compact but convincing addition would be: ambiguity rate / crosstalk vs 2 or 8 orientations, including SNR and localization penalty.

b. Related to above, since both polarization channels are measured simultaneously and separated by lobe spacing, you might quantify polarization leakage / misclassification rate under noise and wavefront gradients.

c. How sensitive is DH-PSF orientation decoding to defocus, astigmatism, or fabrication-induced phase errors?

Reviewer #3

(Remarks to the Author)

The authors present a chip-integrated metasurface Shack–Hartmann sensor operating directly on a CMOS image sensor without external optics, using a dual-encoding strategy. The key innovation lies in the use of orientation-encoded double-helix point spread functions (DH-PSF) to extend the dynamic range while preserving localization accuracy. The work is commendable and suitable for Nature Communications. However, I suggest addressing the following points before publication:

1. The manuscript should clarify whether the simultaneous control of LCP and RCP is achieved via a single structure that combines propagation and structural phase, or through interleaved sets of structures that independently control the two polarization states.

2. Since the inter-vortex separations for LCP and RCP incidence are crucial for doubling the sampling density without sacrificing dynamic range, the relationship between the inter-vortex separation d and phase mapping in DH-PSF tuning should be provided in the equations.

3. The paper uses a 2×2 superlattice. Would using additional point spread functions with different rotational angles (e.g., 3×3 or 4×4 superlattices) further enhance the dynamic range while maintaining sampling density? What challenges might arise from using a larger number of PSFs? Additionally, what is the smallest resolvable phase gradient detectable by the system?

4. In the "3D localization of a point source" section, is the axial position determined through wavefront fitting? Some studies use the rotation angle of the DH-PSF to measure axial position (e.g., Adv. Photonics 1, 036001 (2019)). Does the DH-PSF employed here have this capability? Are there limitations on the axial distance range that can be measured? What is the vertical position resolution achievable with this method?

5. The dynamic wavefront measurement is intriguing. I recommend that the authors evaluate its accuracy. While the water droplet shape changes, could internal stresses lead to refractive index non-uniformities that affect measurement accuracy? Has any research been published describing the behavior of water droplet surface changes, or could simulation be used to verify this?

6. The mapping figures should include coordinates or scale bars for clarity. This applies to Figure 1d, e, Figure 2, Figure 3c-h (and insets), and Figure 4b.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I have read the revised manuscript, as well as the reply to the other referees. The authors' response to my comments (and those of the other referees) are satisfactory and I recommend publication.

Reviewer #2

(Remarks to the Author)

Overall, the revised manuscript is significantly strengthened and the central technical contribution—orientation-encoded DH-PSFs combined with polarization-interleaved sampling in a CMOS-bonded architecture—now appears well supported. I therefore recommend publication in Nature Communications after minor revisions, primarily limited to final polishing and consistency checks.

Although the current sensor is a prototype, it nevertheless achieves the intrinsic limit corresponding to a $\pm 9.4^\circ$ dynamic range (including the result in Fig. 2h), which would not be feasible without employing a metasurface. It may be helpful to emphasize this point in the Discussion as a record-breaking result, while also noting that there remains room for further improvement through algorithmic enhancements and related approaches.

In addition, double-helix PSFs are most commonly used for depth sensing, whereas this work leverages their distinctive features to reduce crosstalk. It would strengthen the manuscript to highlight this repurposing more explicitly. Finally, while the current design is optimized for a single wavelength, it would be useful to emphasize the broader potential of the platform: by changing the material platform and redesigning the metasurface, the approach must be adaptable to essentially any wavelength band.

Reviewer #3

(Remarks to the Author)

The authors have addressed my concerns thoroughly and in great detail. I recommend this manuscript for publication in Nature Communications.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Re: Manuscript ID: NCOMMS-25-98695

Title: Chip-Integrated Metasurface with Double-Helix PSF for Ultra-Large Dynamic Range Wavefront Sensing

Dear Editors,

We sincerely thank you for your efforts and help in dealing with our manuscript. We have carefully considered each comment and have revised the manuscript accordingly. Below, we provide a point-by-point response to all reviewer comments. All changes in the revised manuscript are highlighted in yellow for ease of review.

Response to Reviewer #1

Comment: The authors report a chip-integrated metasurface Shack–Hartmann wavefront sensor (SHWFS) that addresses the fundamental trade-off between dynamic range and spatial sampling density. By utilizing a dual-encoding strategy—orientation-encoded double-helix point spread functions (DH-PSFs) and polarization-multiplexed channels—they demonstrate a significant increase in dynamic range (9.4 μm) and sampling density in an alignment-free, CMOS-bonded architecture. The experimental demonstrations, including 3D localization and video-rate droplet monitoring, are compelling. This work represents a notable step toward miniaturized metrology. I recommend its publication in Nature Communications once the authors clarify certain theoretical bounds and practical limitations.

Author's Response: We sincerely thank you for the constructive and insightful comments on our manuscript. We are particularly grateful to the reviewers for their overall positive assessment of our work, which highlights the significance and novelty of our approach.

Comment1: While the dual-encoding strategy is effective, the authors should quantify the performance penalty when switching to the "vectorial mode." As noted in the Supplementary Information, lateral resolution drops by half in this mode; a more rigorous analysis of how increased DH-PSF complexity (higher vortex charge N) affects centroid localization precision under noise would strengthen the paper.

Author's Response:

We thank the reviewer for these thoughtful questions regarding the performance characteristics of our system. Regarding the spatial sampling density reduction when switching to vectorial mode, we would like to provide a more precise clarification. When operating in vectorial mode, each polarization channel 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 $40\sqrt{2} \mu\text{m}$, corresponding to a linear sampling density improvement by a factor of $\sqrt{2}$ relative to the 80 μm 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 rigorously quantify the performance penalty when switching from scalar mode to vectorial mode, we simulated the reconstruction of two-dimensional sinusoidal wavefronts generated as $\phi(x,y) = A \cos(2\pi f_x x) \cos(2\pi f_y y)$ with amplitude $A=140$ rad and spatial frequency f ranging from 0.1 to 4.7 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 under both operating modes. Figure. 1 presents the root-mean-square error (RMSE) of the reconstructed wavefronts as a function of spatial frequency. The red curve corresponds to scalar mode, while the blue curve corresponds to vectorial mode.

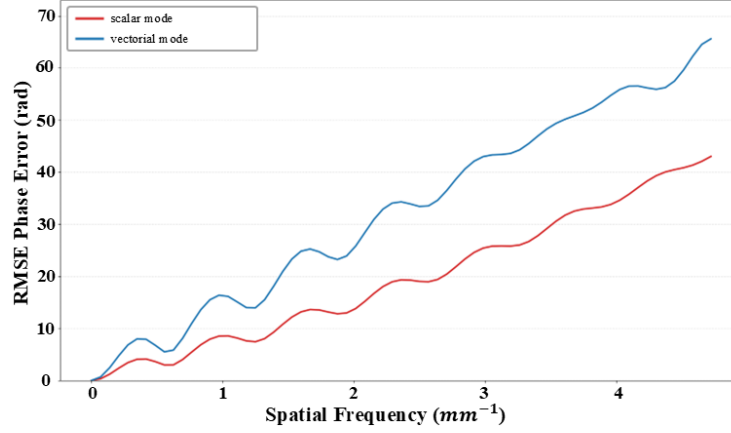


Fig. 1 **Root-mean-square error (RMSE) of reconstructed wavefronts as a function of spatial frequency.** The red curve corresponds to scalar mode, while the blue curve corresponds to vectorial mode.

For a more intuitive comparison, Fig. 2 presents the reconstruction results at three representative spatial frequencies. The top row of Fig. 2(a) shows, from left to right, the ground truth wavefront to be reconstructed in scalar mode at spatial frequency 0.7 mm^{-1} , its reconstructed counterpart, and the residual between reconstruction and ground truth. The bottom row of Fig. 2(a) displays the corresponding ground truth, reconstruction, and residual for vectorial mode at the same spatial frequency. Figure. 2(b) presents the phase profiles along $y = 0$ for the scalar mode reconstruction, vectorial mode reconstruction, and ground truth.

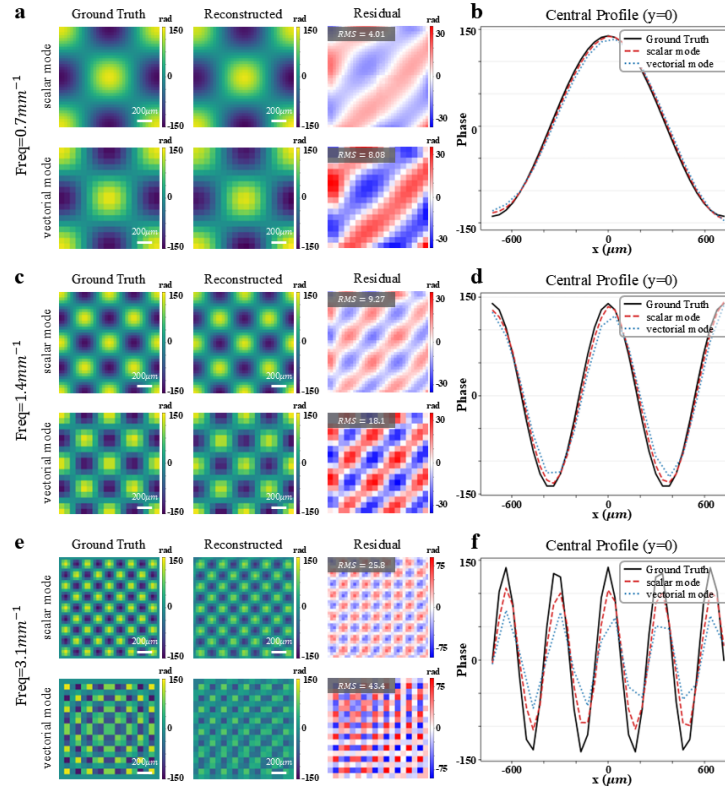


Fig. 2 **Comparison of scalar and vectorial mode reconstruction accuracy at different spatial frequencies.** **a** At spatial frequency 0.7 mm^{-1} , ground truth, reconstructed wavefront, and residual for scalar mode (top row) and vectorial mode (bottom row). **b** Phase profiles along $y = 0$ comparing 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} .

Figure. 2(c) shows the results at spatial frequency 1.4 mm^{-1} , with the top row displaying the ground truth, reconstruction, and residual for scalar mode, and the bottom row showing the same for vectorial mode. Figure. 2(d) provides the corresponding phase profiles along $y = 0$ for both modes alongside the ground truth. Figure. 2(e) extends the comparison to spatial frequency 3.1 mm^{-1} , again presenting scalar mode results in the top row (ground truth, reconstruction, residual) and vectorial mode results in the bottom row. Figure. 2(f) displays the phase profiles along $y = 0$ for this highest frequency.

As evident from these comparisons, 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.

Regarding how increasing the complexity of the DH-PSF (e.g., the vortex number N) affects localization precision under noisy, we conducted a rigorous analysis. The localization accuracy of the DH-PSF is primarily governed by three factors: focusing efficiency, main lobe shape, and main lobe spacing. Higher focusing efficiency improves the signal-to-noise ratio (SNR); a more regular lobe shape facilitates algorithm-based recognition; and greater distinguishability in lobe spacing helps avoid crosstalk between polarization channels. We comprehensively investigated how key design parameters of the DH-PSF, including the vortex number N (where $N=2M+1$), the inter-vortex separation d , and the vortex topological charge c , influence the focusing efficiency, main lobe spacing, and main lobe shape characteristics.

Specifically, Fig. 3(a) illustrates the efficiency of the DH-PSF as a function of the vortex number N (where $N=2M+1$), under fixed conditions of vortex separation $d=30 \text{ }\mu\text{m}$ 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 Fig. 3(a) reveals that the main lobe intensity profile exhibits an irregular shape when the number of vortices is small (e.g., $M=1$ or $M=3$). As M increases, the DH-PSF gradually assumes a regular and stable shape. Figure. 3(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 .

Figure. 3(c) and 3(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. Figure. 3(e) and 3(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 $d=30 \text{ }\mu\text{m}$. The results show that efficiency decreases with increasing c , whereas the main lobe separation increases approximately linearly with c .

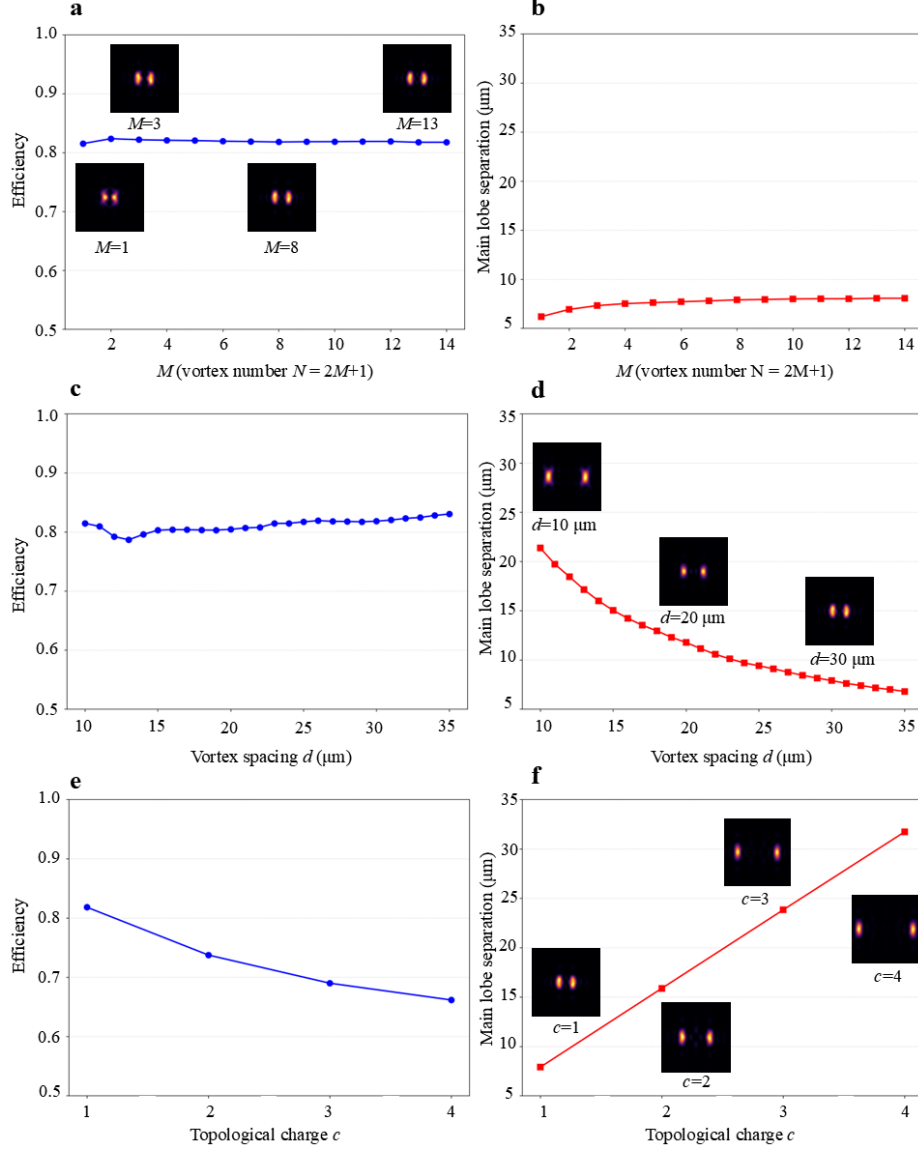


Fig. 3: **Design parameter dependence of double-helix point spread function characteristics** a-b: Efficiency and main lobe separation of the double helix point spread function as a function of M . c-d: Efficiency and main lobe separation of the DH-PSF as a function of the vortex separation d . e-f: 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 $(30 \times 10^{-6} \text{ m}, 8)$ and $(15 \times 10^{-6} \text{ m}, 10)$, respectively. The resulting efficiencies and main lobe separations were (81%, 7.9 μm) for the LCP channel and (80%, 15.4 μm) for the RCP channel. This parameter selection ensures sufficient separation between the DH-PSFs of the two polarization channels to prevent crosstalk, while maintaining regular main 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.

We have now supplemented Supplementary Material S1 (DH-PSF design) with a detailed analysis of how key design parameters influence the characteristics of the DH-PSF.

We also added the discussion regarding the performance penalty in vectorial mode to Supplementary Material S8 (Vectorial mode performance).

Comment2: The current design is optimized for 532 nm. Given the dispersive nature of metasurfaces, the authors should discuss the system's susceptibility to chromatic aberrations or provide a roadmap for broadband operation (e.g., via achromatic metalens designs?).

Author's Response:

We appreciate the reviewer's insightful comment regarding chromatic performance. The reviewer correctly notes that our present design is optimized for a single wavelength (532 nm). In such an integrated system where the metasurface-to-sensor distance is fixed, dispersion would indeed introduce a wavelength-dependent defocus term. This defocus, in turn, would manifest as a uniform rotation of the entire DH-PSF array with wavelength, potentially degrading performance if uncorrected. As the reviewer helpfully suggests, a direct and promising path to broadband operation lies in employing achromatic metalens designs, where the meta-atom group delay is engineered to ensure a consistent focal length across the target band. The feasibility of this approach has been recently demonstrated in an achromatic double-helix metalens operating across the 1200–1400 nm range (*ACS Photonic.* 5c02219 (2026) doi:10.1021/acsp Photonics.5c02219.), which successfully eliminates chromatic aberration-induced PSF rotation and achieves a 14.39-fold enhancement in energy utilization efficiency. An alternative strategy is to optimize the DH-PSF phase pattern for an extended depth-of-focus (DOF), which is crucial for maintaining high diffraction efficiency when the PSF rotates due to defocus across wavelengths (a concept supported by works such as (*Nat. Commun.* **14**, 1035 2023)). In either case, implementing broadband detection would additionally require a wavelength-specific recalibration of the DH-PSF orientation map. We note that our existing orientation decoding scheme, with its inherent $\pm 22.5^\circ$ tolerance, provides a valuable buffer that could enhance robustness against residual chromatic variations in a broadband system.

In response to this valuable suggestion, we have now included a brief discussion on broadband operation in the revised Discussion:

“While the current design is optimized for a single wavelength of 532 nm, future iterations could adopt achromatic metalens designs or depth-of-focus-optimized phase patterns to enable broadband operation, mitigating chromatic aberrations inherent to dispersive metasurfaces.”

Comment3: The metasurface is bonded to the CMOS sensor via an optically transparent adhesive (OCA). The authors should comment on (or ideally provide data for) the thermal and mechanical stability of this bond. Specifically, how do temperature fluctuations affect the PSF orientation or the precise focal distance required for high-accuracy reconstruction?

Author's Response:

We sincerely thank the reviewer for raising this important point regarding the thermal mechanical stability of the optical adhesive bonding. In response, we now specify the optical clear adhesive used (OCA8146-2 Thorlabs.) in the main text upon its first mention, and added a sentence directing readers to Supplementary Materials S2 for a detailed discussion on its thermal and mechanical stability and its impact on wavefront reconstruction accuracy. In Supplementary Materials S2, we provide the official temperature stability parameters of the adhesive (short-term

high-temperature resistance up to 177°C and a long-term operational range from −40°C to 85°C), which confirm its reliability under typical laboratory and environmental conditions.

Regarding the potential effect of temperature fluctuations on measurement accuracy, we acknowledge that thermal expansion or contraction of the CMOS cover glass and the adhesive layer could alter the optical path length between the metasurface and the detector. In principle, this might induce a slight rotation of the DH-PSF. However, our encoding scheme incorporates substantial tolerance: the four principal orientation angles of the DH-PSFs (0°, 45°, 90°, and 135°) are designed with an allowable decoding margin of $\pm 22.5^\circ$. Within this range, minor PSF rotation does not affect orientation identification or centroid localization accuracy. Consequently, the wavefront reconstruction precision remains uncompromised. This inherent design robustness ensures stable operation under realistic temperature variations.

In response to this valuable suggestion, we have added an analysis of thermomechanical stability in Supplementary Material S2.

Comment4: The demonstration shows video-rate monitoring, but it appears to be processed offline. For applications in adaptive optics, real-time feedback is crucial. A brief discussion on the computational complexity of the TV-regularized Poisson integration and potential for hardware acceleration would be beneficial.

Author’s Response:

Thank you for the reviewer's valuable comments regarding real-time performance and computational complexity. We fully agree that real-time feedback capability is crucial for adaptive optics applications. To address this question comprehensively, the wavefront reconstruction pipeline has been analyzed in two distinct stages: gradient extraction (DH-PSF localization and slope computation) and phase retrieval (TV regularized Poisson integration). A detailed discussion of the computational characteristics and optimization pathways for each stage is provided below.

For the phase retrieval step, detailed benchmark tests were performed on a platform equipped with an AMD Ryzen 5 9600X 6-core processor, comparing three representative solvers. The results are summarized in Table 1. The key metrics for evaluating real-time performance are single-frame processing time, frame rate (FPS), speedup and theoretical complexity. Single-frame processing time (in milliseconds) directly determines whether video-rate or closed-loop control requirements can be met; FPS is the reciprocal of processing time; speedup is defined as the ratio of processing time between the baseline method and the optimized method; theoretical complexity reveals the fundamental speed differences from an algorithmic perspective. Additionally, reconstruction quality is assessed through residual maps comparing reconstructed and ground truth phases.

The CVXPY general-purpose convex solver, while offering straightforward implementation, employs interior-point methods with polynomial complexity. For 35×35 phase maps, it requires 59.43 ms per frame (approximately 16.8 FPS), adequate for offline analysis but insufficient for real-time feedback. To overcome this bottleneck, an Alternating Direction Method of Multipliers (ADMM) solver specifically tailored for the TV-regularized problem was implemented. By splitting variables and updating dual variables, ADMM decomposes the original problem into easily solvable subproblems, requiring only one fast DCT and one soft-thresholding operation per iteration, with overall complexity of $O(kN \log N)$ where k is the number of iterations. Under default parameters, ADMM reduces the average reconstruction time to 5.05 ms (approximately 198 FPS), achieving an $11.8\times$ speedup compared to CVXPY. Furthermore, the Discrete Cosine Transform (DCT)-based Poisson solver directly solves the unregularized equation in the frequency domain with a single DCT and inverse DCT, achieving $O(N \log N)$ complexity and a remarkable 0.07 ms per frame (over 13,000 FPS), corresponding to a $798.5\times$ speedup. Table 1 clearly demonstrates the progressive performance gains from a general-purpose solver to a specialized iterative algorithm

and finally to a direct spectral method, confirming that algorithmic optimization can dramatically accelerate phase retrieval.

Table 1. 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	$O(N\log N)$

Crucially, these speed improvements do not compromise accuracy. Figure. 4 presents a visual comparison of the three reconstruction methods. Figure. 4(a) shows the ground truth phase to be recovered. Figure. 4(b), 4(c), and 4(d) display the phases reconstructed using CVXPY, ADMM, and DCT, respectively. Figure. 4(e), 4(f), and 4(g) show the corresponding residual maps relative to the ground truth for CVXPY, ADMM, and DCT, respectively. The residual percentages for all three methods are relatively small, with no systematic bias, confirming that the high-speed algorithms do not sacrifice accuracy. Notably, although the DCT method lacks a regularization term, excellent reconstruction quality is maintained under the high SNR gradient data in this experiment, providing a reliable option for applications demanding high speed.

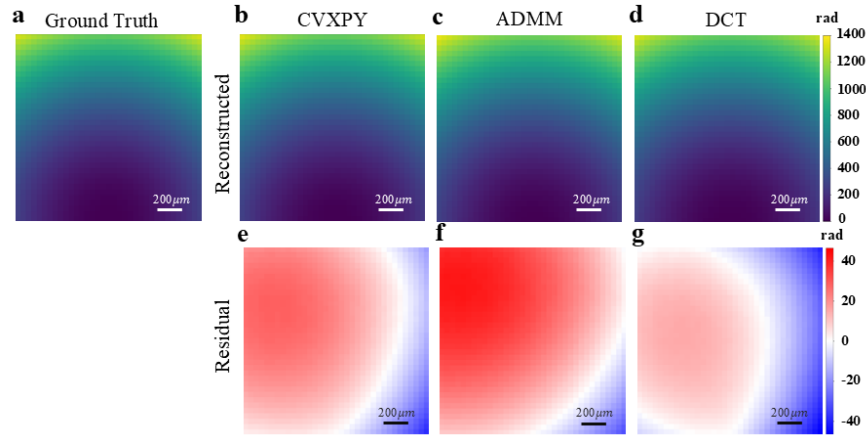


Fig. 4 Comparison of phase reconstruction accuracy across different solvers. **a** Ground truth phase. **b** Reconstruction using CVXPY. **c** Reconstruction using ADMM. **d** Reconstruction using DCT. **e** Residual map for CVXPY reconstruction relative to ground truth. **f** Residual map for ADMM reconstruction relative to ground truth. **g** Residual map for DCT reconstruction relative to ground truth.

For the gradient extraction step, which localizes spots by analyzing DH-PSF morphology and computes wavefront slopes, The initial implementation processed each sub-aperture serially, resulting in tens of seconds per frame and dominating system latency. Asynchronous multi-threading reduced this to approximately 0.4 seconds, yet the gradient extraction stage remains the primary computational constraint. It's worth noting that this stage exhibits inherent data parallelism: each sub-image is processed independently with no cross-dependencies. The core template matching operation, however, involves extensive sliding window searches that are computationally demanding on conventional 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. Such hardware acceleration would effectively eliminate the remaining bottleneck.

Based on the above analysis, the phase retrieval step has already achieved sub-millisecond processing through algorithmic optimization (ADMM or DCT), far exceeding standard video real-time (30 FPS) and even kHz-level closed-loop requirements. Although the gradient extraction step currently dominates the latency, its highly parallel nature makes it amenable to order-of-magnitude acceleration through hardware implementation. Therefore, the real-time feasibility of the entire wavefront reconstruction pipeline is not limited by fundamental principles or algorithmic frameworks, but rather by the extent to which parallel computing resources are utilized in engineering implementation. Hence, the gradient measurement and phase retrieval methods proposed in this study possess full potential for real-time adaptive optics systems.

In response to the reviewer's comment, we have added a discussion on real-time operation in the revised Discussion:

“Regarding real-time operation, our computational analysis (Supplementary Material S9) confirms that the wavefront reconstruction step can achieve sub-millisecond latencies with optimized algorithms, and the highly parallelizable gradient measurement process is amenable to hardware acceleration (e.g., GPUs or FPGAs) for demanding adaptive optics applications requiring kHz-level loop rates.”

Since ADMM achieves accuracy comparable to the previously used CVXPY solver while offering significant improvements in real-time performance, and is more widely adopted in industrial applications than the DCT-based method, we reprocessed all data using the ADMM solver and regenerated the corresponding figures. In addition, we unified the colormaps across all figures to ensure that wavefront maps, gradient maps, height maps, and residual maps each adopt a consistent colormap within their respective categories.

Additionally, we have included a new section, Supplementary Material S9 (Computational Complexity and Real-Time Implementation Analysis), in the Supplementary Materials.

Comment5: The manuscript would benefit from a more comprehensive discussion of the recent landscape in metasurface and nonlocal-filter-based wavefront sensing (Laser & Photonics Reviews 17, 2300426 (2023), Science Advances 10, eadl0501 (2024), Nature Photonics 19, 1315 (2025), National Science Review nwaf561 (2025)).

Author’s Response:

We thank the reviewer for this valuable suggestion to broaden the contextual discussion. We have carefully investigated the mentioned references and addressed them in the revised manuscript at appropriate locations, ensuring accurate representation of each work's contribution to the field.

Specifically, the review article on nonlocal thin-film optical filters (*Laser Photonics Rev.* **17**, 2300426 (2023)) has been included in the discussion of hardware-based dynamic range extension strategies in the second paragraph of the Introduction:

“adjusting the relative positions of the microlens array and image sensor¹⁶⁻¹⁸, integrating auxiliary sensing elements¹⁹⁻²¹, employing tunable lenses²²⁻²⁴, or utilizing nonlocal thin-film optical filters²⁵” This placement accurately reflects its contribution as a hardware-level method for extending sensor performance.

The three metasurface-based wavefront sensing works (*Sci. Adv.* **10**, eadl0501 (2024); *Nat. Photon.* **19**, 1315 (2025); *Natl. Sci. Rev.* nwaf561 (2025)) have been cited together in the third paragraph of the Introduction, where we now state:

“Several notable metasurface-based wavefront sensing schemes^{7,37-38} have been demonstrated recently, showcasing the potential of this approach for compact metrology.”

The reference numbering has been updated accordingly throughout the manuscript as follows: [7], [25], [37], [38].

Response to Reviewer #2

Comment: In this work, authors propose a chip-integrated metasurface Shack–Hartmann-type wavefront sensor without external relay optics, claiming to overcome the dynamic-range vs sampling-density vs accuracy trade-off. The approach uses DH-PSFs with multiple orientation encodings to mitigate spot ambiguity under large slopes, and LCP/RCP polarization channels with distinct lobe separation to enable interleaved sampling that doubles effective sampling density. The manuscript reports $\pm 9.4^\circ$ dynamic range per polarization channel via tilt-plane-wave calibration, and demonstrates wavefront reconstruction using a pipeline involving PSF localization, gradient completion via neighborhood interpolation, and TV-regularized Poisson integration.

The core idea of orientation-encoded double-helix PSF and polarization-based interleaving to expand dynamic range while maintaining high sampling density is technically interesting and potentially impactful. However, several key claims are currently overstated or insufficiently supported, particularly regarding below:

Author’s Response: We sincerely thank Reviewer 2 for their thoughtful and constructive comments, and for their positive assessment of our manuscript. We have diligently addressed each point, which has greatly improved the clarity of our work.

Comment 1: To avoid ambiguity between adjacent spots under large wavefront tilts, the measurable tilt is restricted to half the microlens diameter, resulting in a dynamic range of only about $\pm 1^\circ$. (from introduction) As noted, this requires explicit assumptions (e.g., no advanced spot-matching, no multi-frame methods, particular pitch/focal length/NA regime). In general, the tilt limit scales with lenslet NA / f and detector allocation, so presenting the dynamic range of conventional SH sensor as a blanket constraint seems misleading. For example, the commercially available SH sensor from OPTOCRAFT provides dynamic range up to 10 degrees.

Author’s Response:

We sincerely thank the reviewer for this insightful comment. We agree that the dynamic range of a Shack-Hartmann wavefront sensor (SHWFS) is not a fixed constant but is influenced by multiple factors, including advanced spot-matching algorithms, multi-frame method, and specific hardware design parameters (e.g., lenslet pitch and focal length). We appreciate the opportunity to clarify the context of our initial statement and improve the rigor of the manuscript.

The " $\pm 1^\circ$ " value mentioned in our Introduction was intended to represent the intrinsic geometric dynamic range ($\theta_{max} \approx \tan^{-1}(D/2f)$) of standard refractive microlens arrays (MLAs) typically used in SHWFS, as cited in recent literature [e.g., *Light Sci. Appl.* **13**, 187 (2024)]. This dynamic range is determined solely by the geometrical parameters of the MLAs, without incorporating algorithmic or hardware enhancements. Due to fabrication technology which imposes limitations on minimum feature size and maximum achievable curvature of MLAs, typical lenslet in MLAs have diameters around 100 μm and relatively small NA. As a result, the sampling density is limited to approximately 100 sub-apertures per square millimeter, and the maximum measurable angle ($\theta_{max} \approx \tan^{-1}(D/2f)$) is approximately 1° .

Regarding the Optocraft sensors mentioned by the reviewer, we have conducted further investigation and consulted their technical specifications. The two models that achieve a maximum tilt range of 10° , the parameters are summarized below:

Table 2. Performance specifications of two representative OPTOCRAFT Shack-Hartmann sensors

Performance	SHSCam HR3-130-GE	SHSCam HR3-075-GE
Detection area	$8.4 \times 7.0 \text{ mm}^2$	$8.4 \times 7.0 \text{ mm}^2$
Lateral resolution	$63 \times 53 \text{ microlenses}$	$111 \times 93 \text{ microlenses}$
Microlens-pitch	$0.13 \times 0.13 \text{ mm}^2$	$0.075 \times 0.075 \text{ mm}^2$

Focal length (typical)	3.3 mm	1.7 mm
Calibration wavelength	Default: 635 nm	Default: 635 nm
Wavelength range (high precision)	Calibration WL $\pm 10\%$	Calibration WL ± 50 nm
Usable wavelength range	355 – 1064 nm	355 – 800 nm
Maximum tilt (directly measurable)	$\pm 10^\circ$	$\pm 10^\circ$
Minimum local wavefront radius of curvature	10 mm	6 mm

As shown, the intrinsic dynamic range calculated from the sub-aperture geometry remains near 1° . We noted that the " 10° " maximum tilt specified in the manufacturer's datasheet typically refers to a global tilted plane wave (where only the Zernike tilt term is present). In this special case, all spots shift uniformly without crossing sub-aperture boundaries relative to each other, thus avoiding centroid confusion. This performance is indeed achievable for this special wavefront type, often in combination with dedicated algorithms, and can exceed the $\tan^{-1}(D/2f)$ bound. We acknowledge, as the reviewer correctly notes, that under certain wavefront conditions and with advanced processing, the effective dynamic range can be extended beyond the basic geometrical limit. However, for complex wavefronts with high local gradients, the conventional sub-aperture constraint still poses a significant limitation unless advanced algorithms or specialized hardware are employed.

In response to this comment, we have clarified the context of the $\pm 1^\circ$ value in both the Abstract and Introduction, we have explicitly qualified this value as the "*intrinsic geometric dynamic range*" to distinguish it from the effective dynamic range achievable with algorithmic or hardware enhancements. The revised Introduction now reads:

"To avoid ambiguity between adjacent spots under large wavefront tilts, the measurable tilt is restricted to half the microlens diameter, resulting in an intrinsic geometric dynamic range of only about $\pm 1^\circ$, when considering the microlens array alone without algorithmic or hardware enhancements¹⁵. This geometric limit can be expressed as $\theta_{\max} \approx \tan^{-1}(D/2f)$, where D is the microlens pitch and f is the focal length, corresponding to the maximum acceptance angle that can be unambiguously measured before a spot crosses into an adjacent sub-aperture's designated region."

We have moved the reference Go, G.-H. et al. (*Light Sci. Appl.* **13**, 187 (2024)) to directly support this discussion, as this recent work explicitly addresses the geometric limitations of conventional SHWFS. The reference now appears as citation [15] in the revised manuscript.

Comment 2: "Stress-induced birefringence" is claimed but not convincingly demonstrated experimentally. The abstract and near end of the manuscript, it is implying to include "stress-induced birefringence." The Supplement shows polarization-resolved wavefront reconstruction workflow (LCP vs RCP) and residual maps, but this reads as capability demonstration, not a clear stress-birefringence measurement with a stressed sample and a validated ground truth. If described in abstract, I propose that you should either (i) add a direct experiment (stressed glass/plastic with known birefringence patterns, compare to polarimetry/known model), or (ii) tone down to "supports polarization-resolved wavefront sensing".

Author's Response:

We thank the reviewer for this insightful observation. We acknowledge that the original phrasing regarding "stress-induced birefringence" could be interpreted as implying experimental validation, which was not our intention. In fact, the statement was intended to illustrate a potential future direction enabled by the system's polarization-resolved wavefront sensing capability, rather than to claim a demonstrated measurement of stress birefringence. As shown in Supplementary Material Fig. S13, our system indeed supports polarization-resolved wavefront sensing, enabling independent reconstruction of LCP and RCP wavefronts.

We also recognize that our current design, which prioritizes ultra-large dynamic range through orientation-encoded DH-PSFs and interleaved polarization channels, may not be optimally configured for high-precision birefringence measurements. Accurate stress-induced birefringence characterization typically requires high wavefront sensitivity that would necessitate a redesigned metasurface tailored to that particular application.

Therefore, we fully accept the reviewer's suggestion and have toned down the claims accordingly. The original wording in the Abstract, "*enabling reconstruction of polarization-phase distortions such as stress-induced birefringence*," has been revised to a more general statement that reflects the demonstrated capability: "*The architecture inherently supports vectorial wavefront sensing, enabling reconstruction of polarization-dependent phase distortions*." The corresponding sentence in the Discussion has also been updated to read: "*This capability opens up prospects for future applications in scenarios involving polarization-resolved wavefront sensing (see Supplementary Material S8)*." These modifications accurately reflect the demonstrated capabilities of our system while appropriately contextualizing future possibilities, ensuring readers understand the distinction between the system's demonstrated polarization-resolved sensing capability and its potential future applications.

Comment 3: Concerns on reconstruction pipeline. The Supplement describes gradient completion by local 4-neighbor averaging to fill missing nodes due to interleaved sampling. Given that the method's value proposition is quantitative metrology, you need to show why simpler coordinate transforms are not adequate in this sampling geometry (your Fig. S1 concern). Please quantify bias vs ground truth (especially for high spatial frequencies) and show impact on final phase error after TV-Poisson integration. For example, it looks to me if the data is 45 degree rotated, no interpolation is necessary.

Author's Response:

Thank you for this insightful comment regarding the gradient completion strategy in our reconstruction pipeline. We fully agree that for a method claiming quantitative metrology capability, a rigorous justification of the chosen interpolation approach is essential, particularly regarding its impact on reconstruction accuracy across different spatial frequencies.

To address this question, we conducted a comprehensive quantitative comparison of different strategies for handling the interleaved sampling geometry inherent to our dual-channel design. The goal was to evaluate how each approach affects the final phase reconstruction error, especially for high spatial frequencies, and to determine whether simpler coordinate transformations could eliminate the need for interpolation.

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. The field of view was set to 1440 μm × 1440 μm, matching the sensor dimensions, and the wavelength was 532 nm. The true gradient fields (g_x, g_y) were computed analytically and then sampled according to the actual sub-aperture locations of our sensor. Five different reconstruction scenarios were compared to isolate the effects of sampling pattern, interpolation, and coordinate transformation.

The first scenario corresponds to the interpolation method used in our manuscript. Starting from a full 35×35 Cartesian gradient grid, only the positions where both coordinates are even or both odd were retained, simulating the interleaved sampling of our dual-channel sensor, while all other positions were set to zero. The missing gradients were then filled by averaging the four nearest valid neighbors, and the completed gradient field was fed into the TV-regularized Poisson solver (ADMM) to reconstruct the phase.

The second scenario explores the 45° rotation method suggested by the reviewer. From the same interleaved data (i.e., retaining only the positions where both coordinates are even or both odd), a coordinate index transformation was applied: $u=(i+j)/2$, $v=(i-j)/2+(N-1)/2$ with $N=35$,

mapping the original grid to a rotated uniform grid of size $N \times N$. The gradient components were transformed accordingly as $g_u = (g_y + g_x) / \sqrt{2}$ and $g_v = (g_y - g_x) / \sqrt{2}$. In this rotated system, the valid data occupy a diamond-shaped region corresponding to the original field of view rotated by 45° , while the rest of the $N \times N$ matrix was padded with zeros as shown in Fig 5(a). Phase reconstruction was performed on this padded grid, and the result was transformed back to the original coordinates using the inverse transformation.

The remaining three scenarios serve as references. The ideal 35×35 full sampling uses the complete Cartesian grid without missing data, representing the best possible reconstruction given the native sampling density. The vector-mode single-channel sampling uses only one polarization channel on an 18×18 uniform grid, representing the reduced resolution case. Finally, to isolate the effect of zero-padding in the rotation method, an ideal rotated 25×25 sampling scenario was included where gradients were sampled on a uniform rectangular grid with spacing equal to the actual physical spacing after rotation ($40\sqrt{2} \mu\text{m}$), with the sampling area chosen large enough to cover the entire rotated field of view without missing points, thus avoiding any zero-padding.

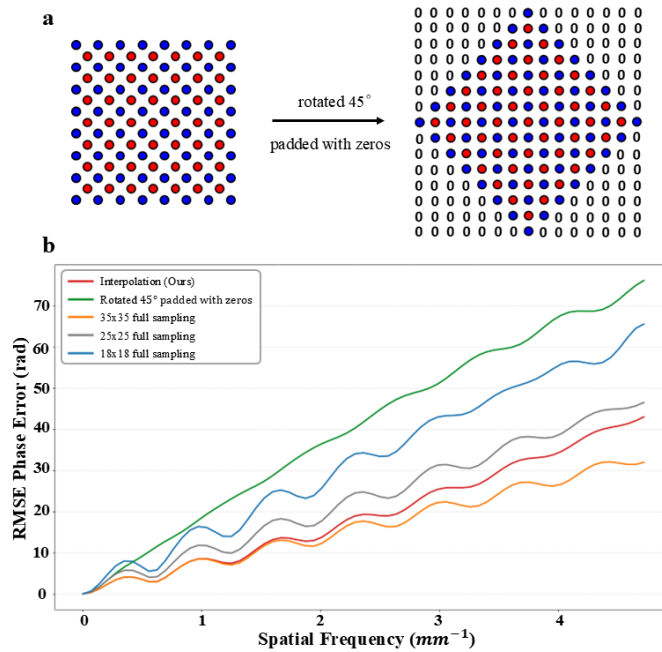


Fig. 5 Comparison of different gradient handling strategies for wavefront reconstruction. **a** Schematic illustration of the 45° rotation method with zero padding. **b** The root mean square errors (RMSE) of the reconstructed phase as a function of spatial frequency. The orange curve corresponds to the ideal 35×35 full sampling benchmark representing the native sampling density. The red 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 gray curve shows the ideal 25×25 rotated sampling with physical spacing of $40\sqrt{2} \mu\text{m}$ and without zero padding, which serves to isolate the effect of zero padding. The blue curve corresponds to the 18×18 single-channel sampling in vector mode, representing the reduced resolution case.

The root-mean-square errors (RMSE) of the reconstructed phase as a function of spatial frequency are plotted in Fig. 5(b), where the red curve corresponds to our interpolation method, the green curve represents the 45° rotation method with zero padding, the gray curve shows the 25×25 ideal rotated sampling, the orange curve represents the 35×35 full sampling benchmark, and the blue curve corresponds to the 18×18 single-channel sampling in vector mode.

In the low frequency region below 0.5 mm^{-1} , all methods exhibit very small errors that closely follow the 35×35 full sampling curve. As the frequency increases into the intermediate range from approximately 0.5 to 2 mm^{-1} , distinct behaviors emerge. The 18×18 single-channel sampling (blue

curve) and the 25×25 ideal rotated sampling (gray curve) show progressively increasing errors due to under-sampling. The 45° rotation method with zero-padding (green curve) exhibits a rapid error increase that even exceeds the 25×25 sampling curve (gray curve), despite theoretically containing the same physical information. In contrast, our interpolation method (red curve) maintains errors remarkably close to the 35×35 baseline and consistently below the 25×25 sampling curve. This demonstrates that local averaging effectively recovers much of the missing gradient information, enabling reconstruction quality approaching that of full sampling.

When the frequency continues to rise into the high region above 2 mm^{-1} , the interpolation method eventually begins to deviate from the 35×35 baseline, reflecting the inherent low-pass filtering effect of local averaging. However, its error remains substantially lower than both the 25×25 sampling curve and, more notably, far below the 18×18 single-channel reconstruction. This latter point is particularly significant, as it directly validates the advantage of our dual-channel polarization multiplexing strategy: by physically sampling more information through two interleaved channels, we achieve reconstruction fidelity that greatly exceeds what would be possible with a single channel of comparable pitch.

A particularly noteworthy finding concerns the performance of the 45° rotation method. Theoretically, this approach should utilize identical physical information to the 25×25 ideal rotated sampling, yet its reconstruction errors are consistently larger and even exceed the 25×25 sampling curve across most frequencies. Our analysis attributes this discrepancy 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.

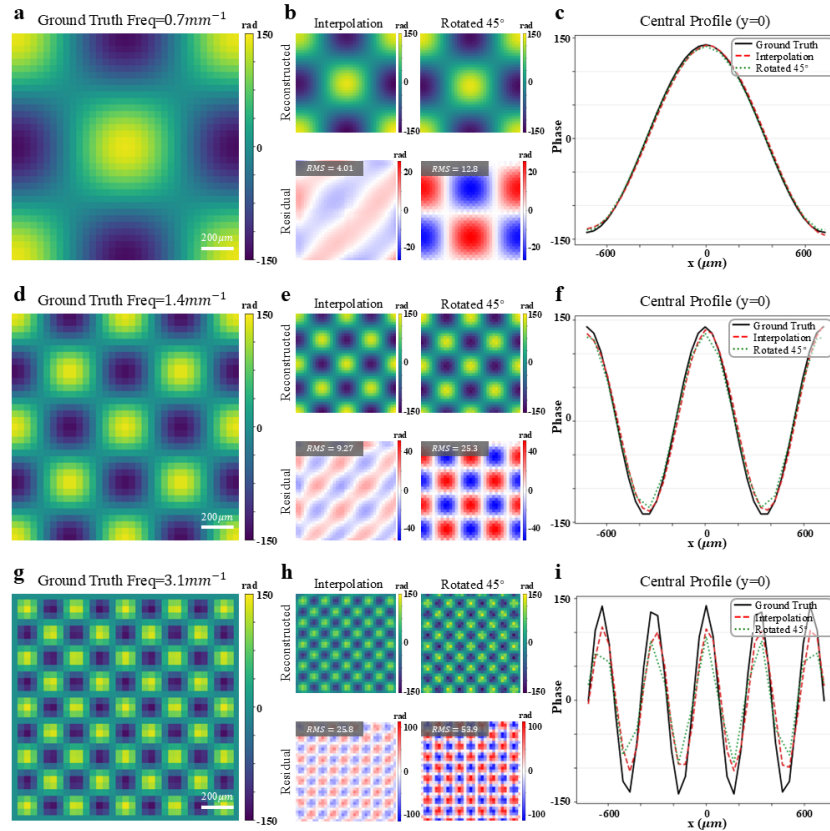


Fig. 6. Comparison of wavefront reconstruction accuracy between interpolation method and 45° rotation method at different spatial frequencies. **a** Ground truth wavefront at spatial frequency 0.7 mm^{-1} . **b** Top row: reconstructed phases using the interpolation method (left) and the 45° rotation 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} .

For a more intuitive comparison, Fig. 6 presents the reconstruction results at three representative spatial frequencies. Figure. 6(a) shows the ground truth wavefront at spatial frequency 0.7 mm^{-1} . In Fig. 6(b), the top row displays, from left to right, the phases reconstructed using the interpolation method employed in the main text and the 45° rotation coordinate transformation method, while the bottom row shows the corresponding residuals relative to the ground truth. Figure. 6(c) presents the phase profiles along $y = 0$ comparing the interpolation method reconstruction, the 45° rotation method reconstruction, and the ground truth. Figure. 6(d) shows the ground truth wavefront at spatial frequency 1.4 mm^{-1} . In Fig. 6(e), the top row displays the reconstructions from the interpolation method and the 45° rotation method, with the bottom row showing the corresponding residuals. Figure. 6(f) presents the phase profiles along $y = 0$ at 1.4 mm^{-1} for both methods alongside the ground truth. Figure. 6(g) shows the ground truth wavefront at spatial frequency 3.1 mm^{-1} . In Fig. 6(h), the top row displays the reconstructions from the interpolation method and the 45° rotation method, with the bottom row showing the corresponding residuals. Figure. 6(i) presents the phase profiles along $y = 0$ at 3.1 mm^{-1} for both methods alongside the ground truth.

In summary, whether the sensor employs the current rectangular checkerboard layout or a rotated 45° arrangement, the physical wavefront information collected remains identical. Our choice of local 4 neighbor interpolation is primarily 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. We are grateful to the reviewer for prompting this analysis, which also suggests an intriguing direction for future hardware design. By 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.

In response to this comment, the original content of Supplementary Material S4 (Wavefront Reconstruction Algorithm) has been reorganized and is now presented as 4.1 Algorithm description. Additionally, a new subsection 4.2 Validation of interpolation strategy has been added to provide quantitative justification for the chosen interpolation approach.

Comment 4: Figures/residual maps are not interpretable enough. Some of the figures (especially for Fig. 3i,j, 4b) are visually unclear (colorbar/saturation), making them meaningless. Especially for Fig. 3i case, I propose to adjust the contrast or choosing different focal length to show clear differences. Please ensure wavefront/gradient maps have units, scale bars.

Author's Response:

We thank the reviewer for the careful reading and for pointing out the insufficient interpretability of some figures. In response, we have made the following improvements to ensure all figures are clear and properly annotated.

For Fig. 3(i) and 3(j), we followed the reviewer's suggestion and adjusted the point source position for both figures to reduce the disparity in wavefront magnitude across different positions. Additionally, we have replaced the linear colormap with a nonlinear mapping to enhance the contrast of low-amplitude regions, making the fine structures more distinguishable while preserving the overall range. The updated figures now clearly reveal the previously less visible features. The revised versions of Figs. 3(i) and 3(j) in the main text are shown below in Fig. 7(a).

For Fig. 4(b), which presents the residual map, we now use a dedicated colorbar independent of those used for the ground truth and fitted height. This avoids the compression of the residual values due to a shared color scale and allows a more faithful representation of the error distribution. The revised versions of Figs. 4(b) in the main text are shown below in Fig. 7(b).

We have thoroughly checked all figures to ensure that units (e.g., μm , rad) and scale bars are explicitly included, following the reviewer's suggestion.

Additionally, we unified the colormap for each physical quantity: the same colormap is now used across all figures for wavefronts, gradients, heights, and residuals, respectively. This provides a consistent visual reference for the reader and facilitates cross-comparison among different panels.

We believe these improvements significantly enhance the clarity and interpretability of the figures. All figures have been updated accordingly in the revised manuscript.

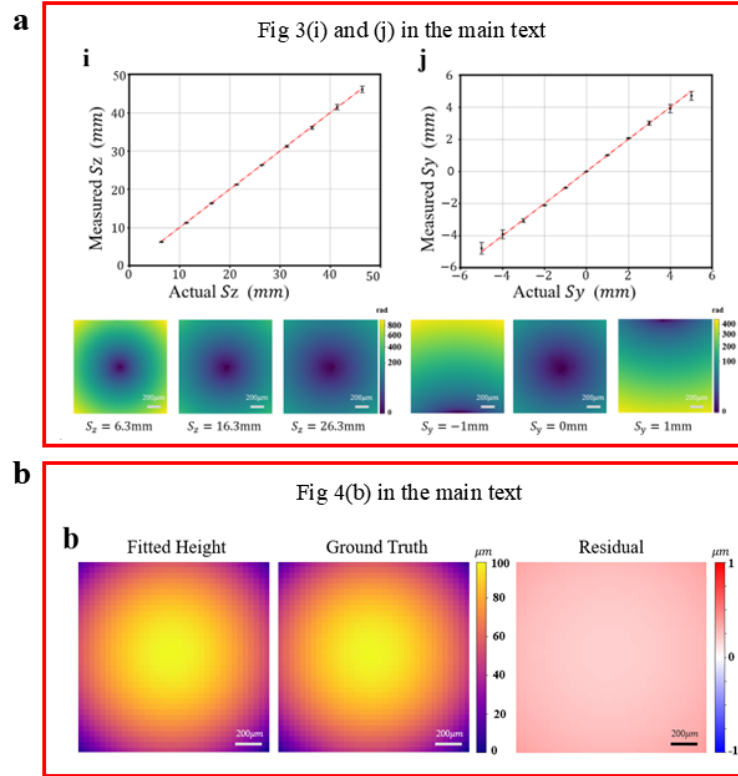


Fig. 7 **Revised figures from the main text.** **a** Corresponds to the revised Figs. 3(i) and 3(j). **b** Corresponds to the revised Fig. 4(b).

Comment 5a: The paper uses four principal orientation angles (0° , 90° , 45° , 135°) for advanced dynamic range. Right now, the manuscript asserts DH-PSF provides unambiguous features at large gradients, but does not quantify or explain why DH-PSF is the right PSF family and how performance scales with the number of distinct encodings. A compact but convincing addition would be: ambiguity rate/crosstalk vs 2 or 8 orientations, including SNR and localization penalty.

Author's Response:

We thank the reviewer for this thoughtful comment regarding the rationale for choosing the DH-PSF family and the scalability of orientation encoding.

The choice of DH-PSF is motivated by its favorable properties for encoding and localization. Conventional Airy disks, while providing high energy concentration in a single lobe, lack any intrinsic orientation information and thus cannot support orientation-based encoding schemes. Astigmatic PSFs, on the other hand, do exhibit orientation-dependent shape variations, but at the cost of energy spreading: the optical power is distributed over a larger area, reducing peak intensity

and signal-to-noise ratio (SNR). This energy dispersion makes centroid localization less accurate and orientation discrimination more challenging under noise-limited conditions.

The DH-PSF offers a compelling compromise. By concentrating optical energy into two well-defined lobes, it maintains high peak intensity, while the 180° rotational symmetry of the two-lobe structure provides sufficient orientational degrees of freedom (through lobe orientation) and an additional discriminative dimension (through lobe spacing) to implement robust encoding. This combination of energy concentration and structural information enables accurate centroid localization and reliable orientation decoding even under challenging signal conditions.

The specific choice of four orientation angles (0°, 45°, 90°, 135°) arises from our 2×2 superlattice design, where each basic periodic unit contains four sub-apertures with distinct orientations. When arranged into an array, this configuration doubles the dynamic range compared to conventional designs, with each orientation benefiting from a ±22.5° angular tolerance that ensures robust discrimination against fabrication misalignments and local wavefront aberrations. This four-orientation design is not a fundamental necessity but represents a deliberate trade-off between dynamic range expansion and decoding robustness tailored to our target application. In principle, using nine orientations in a 3×3 superlattice could triple the dynamic range, but would compress the angular tolerance to approximately ±10°, increasing misidentification risk under noisy or strongly aberrated conditions.

Regarding the broader PSF family, PSFs with three or four lobes could theoretically also support orientation and spacing encoding. However, such configurations distribute the available optical energy over more lobes, reducing peak intensity per lobe and degrading signal-to-noise ratio even further than astigmatic PSFs, which in turn compromises localization accuracy. The two-lobe DH-PSF thus strikes an optimal balance between encoding capacity and energy concentration for high-precision wavefront sensing.

In response to the reviewer’s discussion on the scalability of orientation encoding, we have added a brief trade-off statement in the Discussion section to clarify that employing a finer angular encoding scheme (e.g., nine orientations) could triple the dynamic range per channel, albeit at the cost of reducing the orientation tolerance from ±22.5° to approximately ±10°. The following sentence has been inserted:

“The dynamic range per channel could be tripled, rather than merely doubled, by employing a finer angular encoding scheme for the DH-PSFs, though this would reduce the orientation tolerance from ±22.5° to approximately ±10°.”

Comment 5b: Related to above, since both polarization channels are measured simultaneously and separated by lobe spacing, you might quantify polarization leakage/misclassification rate under noise and wavefront gradients.

Author’s Response:

We thank the reviewer for this important question regarding the robustness of our polarization channel discrimination under noise and wavefront gradients.

To quantify polarization leakage under noisy conditions, we took an experimental image containing all eight DH-PSF patterns (four orientations × two lobe spacings) and added increasing levels of Gaussian noise. Specifically, for each pixel at coordinate (x, y), the noisy intensity is generated as:

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

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:

$$I_{final}(x, y) = \max\left(0, \min\left(255, I_{noisy}(x, y)\right)\right)$$

We then ran the complete decoding pipeline (template matching with a $\pm 22.5^\circ$ rotation tolerance) on the noisy images and recorded the fraction of misclassified PSFs. The resulting misclassification rate as a function of noise standard deviation σ is plotted in Fig. 8, showing that the error rate remains essentially zero for $\sigma \leq 50$ (in units of image gray levels), confirming that the dual-channel simultaneous measurement is highly immune to realistic noise levels, with the error rate rising sharply only above this threshold indicating the noise floor of our current implementation.

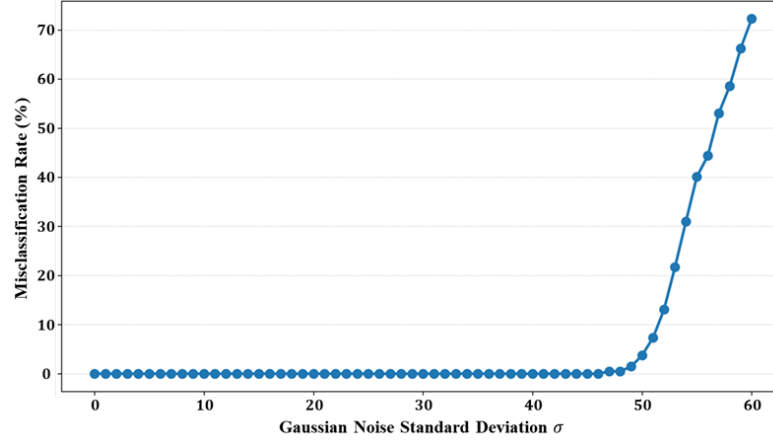


Fig. 8 Polarization channel misclassification rate as a function of added Gaussian noise.

Regarding the effect of wavefront gradients, pure tilt only translates the entire PSF pattern laterally in the focal plane without altering the lobe spacing or orientation, and therefore causes no misclassification between polarization channels. However, more complex aberrations such as defocus, astigmatism, and coma can potentially affect both lobe spacing and orientation, and we have systematically investigated these effects in response to Comment 5c. That analysis provides quantitative thresholds for each aberration type beyond which reliable decoding is compromised, with the results confirming that our dual-encoding scheme remains robust within typical aberration levels of well-corrected optical systems. We believe these additions can address the reviewer's concern regarding polarization leakage under noise and wavefront gradients.

Comment 5c: How sensitive is DH-PSF orientation decoding to defocus, astigmatism, or fabrication-induced phase errors?

Author's Response:

We thank the reviewer for this important question regarding the robustness of DH-PSF orientation decoding to various aberrations and fabrication errors. In response, we have conducted a systematic analysis of how primary aberrations affect the correct identification of all eight DH-PSF patterns (four orientations $\times$ two lobe spacings) used in our dual-encoding scheme.

Since the first three Zernike terms either carry no phase information (piston, Z1) or are precisely the quantities we aim to measure (tip/tilt, Z2/Z3 correspond to wavefront gradients encoded as PSF displacement), we 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 [*J. Opt. Soc. Am.* **66**, 207 (1976)].

Because our eight DH-PSF patterns already cover two orthogonal orientation sets ($0^\circ/90^\circ$ and $45^\circ/135^\circ$) for both lobe spacings, it suffices to consider one orientation for each aberration type. Therefore, we selected 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: $\sqrt{8}\rho^3\sin 3\phi$) as representative cases where (ρ, ϕ) or detailed investigation, where (ρ, ϕ) are polar coordinates in the pupil plane with ρ denoting the normalized radial coordinate ($0 \leq \rho \leq 1$).

For each aberration type, we simulated a single sub-aperture containing all eight DH-PSF patterns with increasing Zernike coefficient magnitudes and evaluated the threshold at which pattern recognition (template matching with $\pm 22.5^\circ$ rotation tolerance) failed. The failure thresholds for each aberration type are summarized in Table 3, providing quantitative guidelines for allowable aberration levels in practical applications.

Table 3. 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

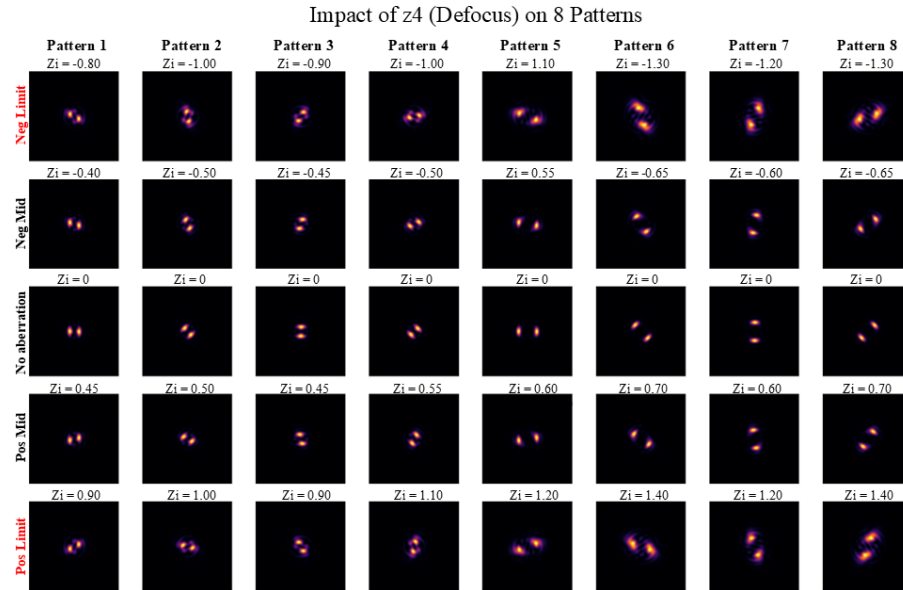


Fig. 9 Effect of defocus on the intensity profiles of eight DH-PSF patterns. Each column corresponds to one of the eight PSF patterns (four orientations $\times$ two lobe spacings). Each row represents a different Zernike coefficient magnitude for defocus (Z_4). The middle row shows the aberration-free 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.

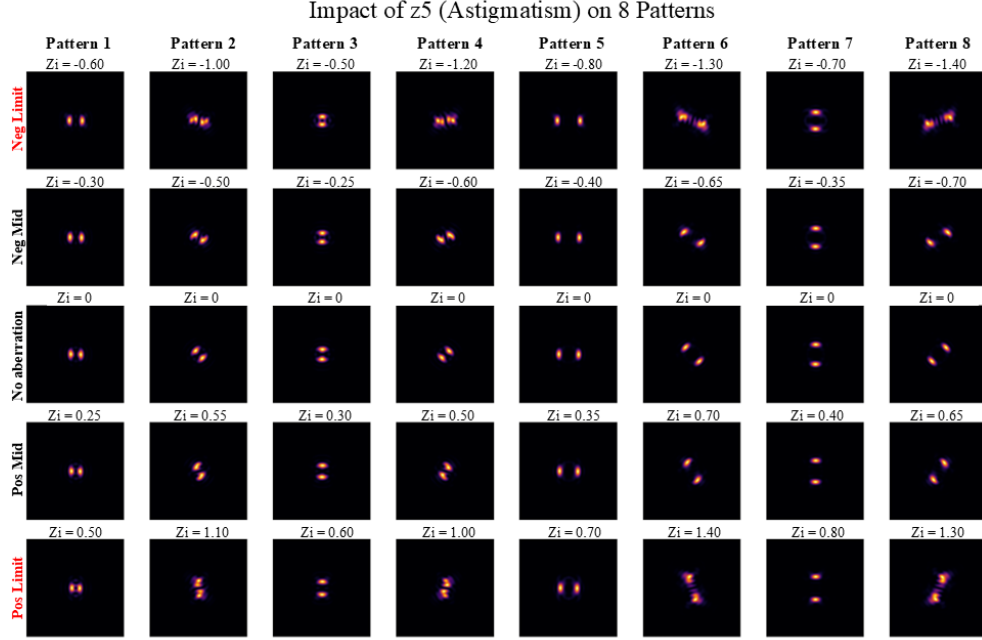


Fig. 10 Effect of astigmatism on the intensity profiles of eight DH-PSF patterns. Each column corresponds to one of the eight PSF patterns (four orientations $\times$ two lobe spacings). Each row represents a different Zernike coefficient magnitude for astigmatism (Z_5). The middle row shows the aberration-free 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.

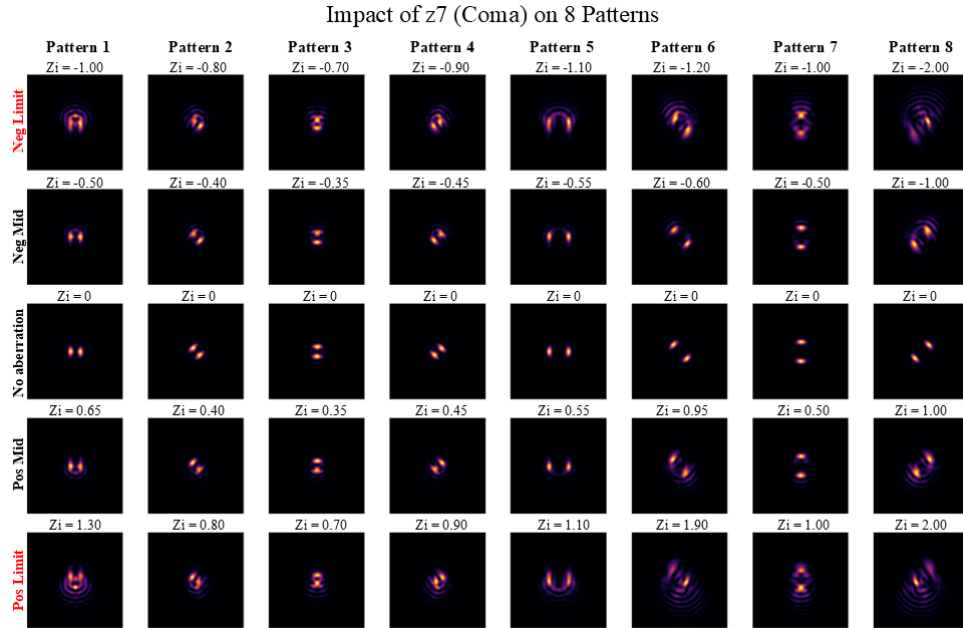


Fig. 11 Effect of coma on the intensity profiles of eight DH-PSF patterns. Each column corresponds to one of the eight PSF patterns (four orientations $\times$ two lobe spacings). Each row represents a different Zernike coefficient magnitude for coma at (Z_7). The middle row shows the aberration-free 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.

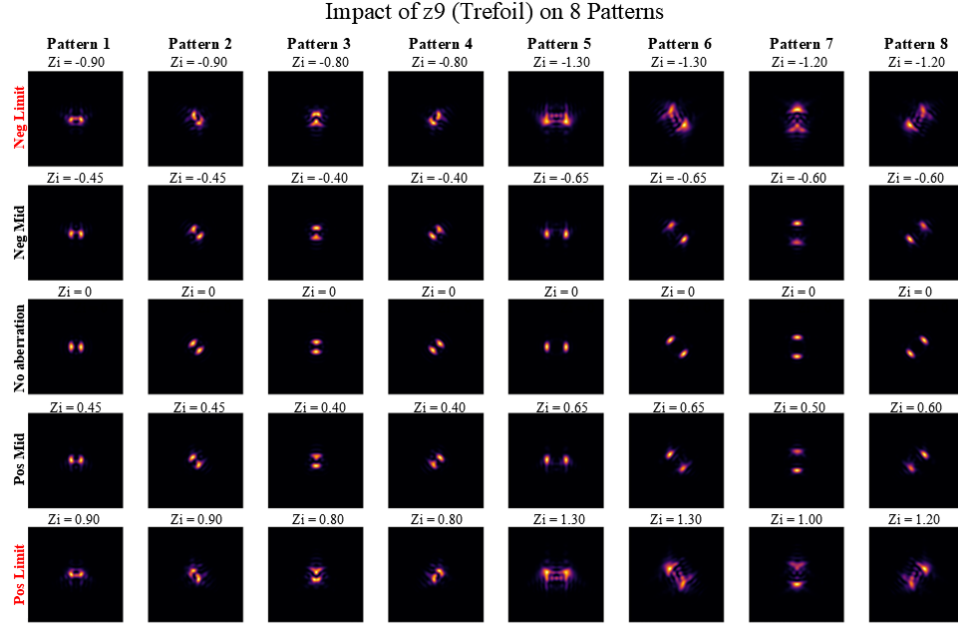


Fig. 12 **Effect of trefoil on the intensity profiles of eight DH-PSF patterns.** Each column corresponds to one of the eight PSF patterns (four orientations $\times$ two lobe spacings). Each row represents a different Zernike coefficient magnitude for trefoil at (Z_9). The middle row shows the aberration-free 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.

To intuitively illustrate the impact of different aberrations on the DH-PSF, Fig. 9 presents the effect of defocus on the intensity profiles of all eight PSF patterns. In this figure, each column corresponds to one PSF pattern, and each row represents a different Zernike coefficient magnitude. The middle row shows the aberration-free intensity profile, the top row corresponds to the negative-direction failure threshold, and the bottom row corresponds to the positive-direction failure threshold. The second and fourth rows display intensity distributions at intermediate aberration levels. Similarly, Fig. 10, Fig. 11, and Fig. 12 illustrate the influence of astigmatism, coma, and trefoil, respectively, on the DH-PSF intensity patterns.

From these images, it can be observed that different aberrations affect the double-helix point spread function through distinct mechanisms, leading to recognition failure. For example, defocus primarily influences the orientation angle. Astigmatism mainly affects the lobe spacing for patterns 2, 4, 6, and 8, while for patterns 1, 3, 5, and 7 it impacts both orientation and spacing. Coma and trefoil have more pronounced effects on the overall lobe morphology. Because these effects are often coupled, particularly for higher-order aberrations, we did not attempt to disentangle individual mechanisms. Instead, we investigated the intensity threshold at which each aberration type causes recognition failure.

Regarding fabrication-induced phase errors, small deviations in nanopillar dimensions during fabrication can introduce two types of effects: they may increase background noise, slightly reducing the signal-to-noise ratio, or they may induce localized initial aberrations. As demonstrated in our response to Comment 5b, decoding accuracy remains unaffected as long as the noise level stays below the established threshold. For the case of induced initial aberrations, we perform a calibration step with plane-wave illumination prior to each measurement to record the reference PSF positions, effectively compensating for any systematic offsets. Moreover, as shown in our aberration robustness analysis above, the dual-encoding strategy exhibits strong tolerance to spacing and orientation variations. Consequently, under typical electron-beam lithography

fabrication errors (with critical dimension variations below 10 nm), the impact on decoding accuracy is negligible.

In response to this comment, we have included a new section, Supplementary Material S5 (Quantitative Analysis of DH-PSF Decoding Robustness), in the Supplementary Materials.

Response to Reviewer #3

Comment: The authors present a chip-integrated metasurface Shack–Hartmann sensor operating directly on a CMOS image sensor without external optics, using a dual-encoding strategy. The key innovation lies in the use of orientation-encoded double-helix point spread functions (DH-PSF) to extend the dynamic range while preserving localization accuracy. The work is commendable and suitable for Nature Communications. However, I suggest addressing the following points before publication:

Author's Response:

We thank the reviewer for the positive and encouraging assessment of our work. We are pleased that the reviewer finds our chip-integrated metasurface Shack–Hartmann sensor with dual-encoding strategy commendable and suitable for publication in Nature Communications. We sincerely appreciate the constructive comments provided, which will help us further strengthen the manuscript. In the following, we provide a point-by-point response to each of the specific points raised.

Comment 1: The manuscript should clarify whether the simultaneous control of LCP and RCP is achieved via a single structure that combines propagation and structural phase, or through interleaved sets of structures that independently control the two polarization states.

Author's Response:

We thank the reviewer for this clarifying question. In our design, the simultaneous control of left-circularly polarized (LCP) and right-circularly polarized (RCP) light is achieved within a single monolithic metasurface structure that combines both propagation phase and Pancharatnam–Berry (PB) phase. Each rectangular nanopillar is designed with specific geometric dimensions (length and width) to provide the desired propagation phase, while its in-plane orientation angle controls the PB phase. This dual-phase modulation mechanism enables independent and simultaneous phase profile engineering for the two orthogonal circular polarization states within the same physical structure, without requiring interleaved or spatially separated sets of nanopillars for each channel [Adv. Photon. 2, 066004 (2020)]. Specifically, for a given nanopillar, the transmitted light under LCP and RCP incidence acquires phase shifts of $\phi_{prop} + 2\theta$ and $\phi_{prop} - 2\theta$, respectively, where ϕ_{prop} is the propagation phase determined by the nanopillar dimensions and θ is its orientation angle. By appropriately designing the propagation phase and orientation angle distributions across the metasurface, we can independently prescribe the desired phase profiles for the LCP and RCP channels. This approach ensures that the two channels operate simultaneously without crosstalk while allowing independent control over their respective DH-PSF characteristics, including main-lobe separation and orientation encoding.

To clarify this point in the manuscript, we have inserted the following sentence in the section “Design of meta wavefront sensor (WFS)”:

"This simultaneous control of left circular polarization (LCP) and right circular polarization (RCP) light is achieved within a single monolithic metasurface by combining propagation phase with Pancharatnam–Berry (PB) phase"⁴⁸. The corresponding reference (Adv. Photon. 2, 066004 (2020).) has been added as citation [48], with subsequent reference numbers adjusted accordingly.

Comment 2: Since the inter-vortex separations for LCP and RCP incidence are crucial for doubling the sampling density without sacrificing dynamic range, the relationship between the inter-vortex separation d and phase mapping in DH-PSF tuning should be provided in the equations.

Author's Response:

We thank the reviewer for this insightful comment. The inter-vortex separation d indeed plays a critical role in determining the main-lobe separation of the double-helix point spread function, which is essential for distinguishing the two polarization channels and doubling the sampling density. To clarify this relationship, we have now explicitly incorporated the dependence on d into the phase mapping equation presented in the main text: *"In radial coordinates, the phase function of the metasurface for a single sub-aperture is given by:*

$$\Phi(r, \theta) = \arg \left(\prod_{j=-M}^M (r e^{i\theta} - r_j e^{i\theta_j}) \right) - \frac{kr^2}{2f},$$

where (r, θ) are the pupil plane coordinates, (r_j, θ_j) is the location of the j -th vortex, $\arg$ denotes the argument (phase) of a complex number and $M=(N-1)/2$, N is the number of vortices, k is the wave vector and f is the focal length, r_j and θ_j are respectively defined as follows:

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

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."

Comment 3: The paper uses a 2×2 superlattice. Would using additional point spread functions with different rotational angles (e.g., 3×3 or 4×4 superlattices) further enhance the dynamic range while maintaining sampling density? What challenges might arise from using a larger number of PSFs? Additionally, what is the smallest resolvable phase gradient detectable by the system?

Author's Response:

We thank the reviewer for raising these important points regarding the scalability of our orientation-encoding strategy and the sensitivity limits of the system. The reviewer correctly observes that our current implementation employs a 2×2 superlattice with four distinct PSF orientations (0° , 45° , 90° , and 135°). In principle, increasing the number of orientation-encoded PSFs within the same aperture layout, for example, using a 3×3 superlattice with nine distinct orientations, could further extend the dynamic range proportionally to the number of orientations while maintaining the spatial sampling density. This prospect aligns with the forward-looking perspective we have briefly outlined in the Discussion section. However, adopting a larger number of orientations would inevitably reduce the angular tolerance allocated to each orientation. In our current scheme, each orientation benefits from a $\pm 22.5^\circ$ recognition window, which provides robust discrimination against fabrication misalignments and local wavefront aberrations. A 3×3 superlattice would compress this window to approximately $\pm 10^\circ$, thereby increasing the risk of misidentification under noisy or strongly aberrated conditions. Thus, the choice of the number of orientations involves a fundamental trade-off between dynamic range expansion and decoding robustness, which can be optimized according to specific application requirements.

Regarding the smallest resolvable phase gradient, for our integrated sensor, the minimum resolvable angle is θ_{min} fundamentally limited by the focal length of the lenslet and the pixel pitch of the CMOS sensor as shown in Fig. 13. Under the small-angle approximation (valid for $\theta < 10^\circ$, which holds for all measurements in this work), The relation is given by $\theta_{min} = \arctan(\Delta/f)$, where ($\Delta = 1.67 \mu\text{m}$) is the pixel pitch and ($f = 470 \mu\text{m}$) is the focal length. This yields $\theta_{min} = 0.2^\circ$. 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 $4.2 \times 10^4 \text{ rad/m}$. Figure. 13 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 $2D$.

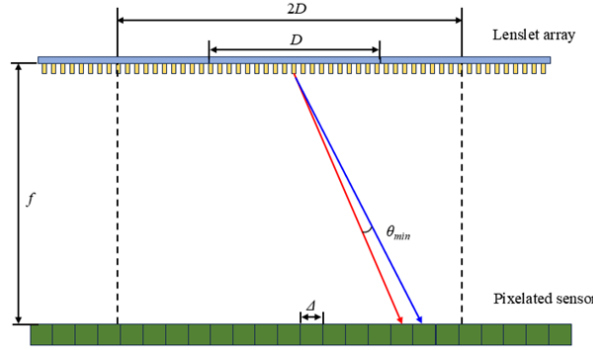


Fig. 13 **Side view schematic of the wavefront sensor.** f denotes the focal length of the lenslet, and D denotes the diameter of the lenslet. The black dashed lines indicate the allowed displacement region for the double-helix focal spot of each lenslet. θ_{min} represents the minimum resolvable angle, and Δ is the detector pixel pitch.

In response to the reviewer’s discussion on the scalability of orientation encoding, we have added a brief trade-off statement in the Discussion section to clarify that employing a finer angular encoding scheme (e.g., nine orientations) could triple the dynamic range per channel, albeit at the cost of reducing the orientation tolerance from $\pm 22.5^\circ$ to approximately $\pm 10^\circ$. The following sentence has been inserted:

“The dynamic range per channel could be tripled, rather than merely doubled, by employing a finer angular encoding scheme for the DH-PSFs, though this would reduce the orientation tolerance from $\pm 22.5^\circ$ to approximately $\pm 10^\circ$.”

We have also added a discussion on sensitivity limit in the “Characterization of meta wavefront sensor” section of the revised manuscript:

“Complementing the upper bound of dynamic range, we also evaluated the sensor’s sensitivity limit. The minimum detectable angular tilt is fundamentally limited by the pixel pitch of the CMOS sensor and the focal length of the metalens sub-apertures, yielding $\theta_{min} \approx 0.2^\circ$ for our system. This corresponds to a minimum resolvable phase gradient of approximately 4.2×10^4 rad/m at the operating wavelength of 532 nm. A detailed derivation of this sensitivity limit is provided in Supplementary Material S3.”

Additionally, we have also included a new section, Supplementary Material S3 (Wavefront Reconstruction Sensitivity Limit), in the Supplementary Materials.

Comment 4: In the “3D localization of a point source” section, is the axial position determined through wavefront fitting? Some studies use the rotation angle of the DH-PSF to measure axial position (e.g., Adv. Photonics 1, 036001 (2019)). Does the DH-PSF employed here have this capability? Are there limitations on the axial distance range that can be measured? What is the vertical position resolution achievable with this method?

Author’s Response:

We thank the reviewer for these insightful questions regarding the 3D localization capability of our system. The reviewer correctly notes that certain double-helix PSF designs encode axial position through the rotation angle of the two lobes, as demonstrated in the cited work (Adv. Photonics 1, 036001 (2019)). We confirm that our DH-PSF exhibits the same fundamental property: the orientation angle of the double-helix pattern rotates with defocus, which in principle could be used for direct depth encoding. However, in the present work, we employ a different approach for 3D point source localization. Rather than relying solely on PSF rotation, we reconstruct the full wavefront from the measured displacement field of the DH-PSF array and then perform inverse

fitting to determine the three-dimensional position of the point source. This wavefront-based method provides complementary advantages, particularly in scenarios where the wavefront carries additional information beyond simple defocus.

Regarding the axial range limitations, our method has both upper and lower bounds. When the point source is positioned too close to the sensor, the spherical wavefront incident on the metasurface exhibits excessively steep phase gradients at the aperture edges, which may exceed the sensor's dynamic range and lead to spot displacements beyond the unambiguous tracking region. Conversely, when the point source is too distant, the phase gradient becomes too shallow to produce a detectable displacement of the DH-PSFs, falling below the minimum resolvable gradient discussed in our response to Comment 3. Thus, the measurable axial range is bounded between these two extremes, which can be tailored by adjusting the sub-aperture focal length and the sensor's pixel pitch.

The axial resolution of our method is not a fixed value but depends strongly on the point source position. This position-dependent behavior arises from the nonlinear relationship between the axial displacement of the point source and the resulting curvature change of the spherical wavefront at the sensor plane. When the point source is near the sensor, a small axial movement produces a substantial change in wavefront curvature, leading to high axial resolution. When the point source is far from the sensor, the same axial movement results in a much smaller curvature change, yielding coarser resolution. Experimentally, we observed that at a source distance of 6.3 mm, the axial resolution is approximately 0.1 mm, while at 46 mm it degrades to about 4 mm.

Comment 5: The dynamic wavefront measurement is intriguing. I recommend that the authors evaluate its accuracy. While the water droplet shape changes, could internal stresses lead to refractive index non-uniformities that affect measurement accuracy? Has any research been published describing the behavior of water droplet surface changes, or could simulation be used to verify this?

Author's Response:

We sincerely thank the reviewer for finding our dynamic wavefront measurement demonstration "intriguing" and for providing thoughtful suggestions to further validate its accuracy.

As demonstrated in the main text through high-fidelity 3D point source localization (Fig. 3 in the main text) and precise lens profilometry (Fig. 4 in the main text), our wavefront reconstruction pipeline achieves high intrinsic accuracy under static conditions. Since the dynamic wavefront sequences were reconstructed offline using the same frame-by-frame algorithmic pipeline, the accuracy achieved in the dynamic scenario is inherently consistent with that demonstrated under static conditions. Our intention in presenting the water droplet experiment was primarily to showcase the sensor's capability for dynamic wavefront capture and its potential for monitoring time-varying optical fields, rather than to perform precise metrology of the droplet's surface itself.

We fully agree with the reviewer that accurately retrieving a droplet's 3D topography from its transmitted wavefront is highly complex. Factors such as the droplet's irregular three-dimensional shape leading to complex refraction, its subtle and time-varying morphological changes due to spreading, evaporation, or condensation, and even potential internal stress-induced refractive index non-uniformities, all significantly complicate any attempt at surface reconstruction. Indeed, specialized techniques have been developed for droplet profilometry (e.g., <https://doi.org/10.1063/5.0169887>), and commercial instruments such as the KRÜSS DSA100B are available for contact angle measurements. Our demonstration focuses on recovering the wavefront at the sensor plane, which inherently captures the cumulative optical effect of the droplet without needing to disentangle these physical complexities. Therefore, the factors mentioned by the reviewer, while critical for surface profilometry, do not compromise the validity of our wavefront measurement demonstration.

To provide a more rigorous quantification of the sensor's accuracy in a dynamic context involving a liquid-based optical element, we performed an additional validation experiment 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.

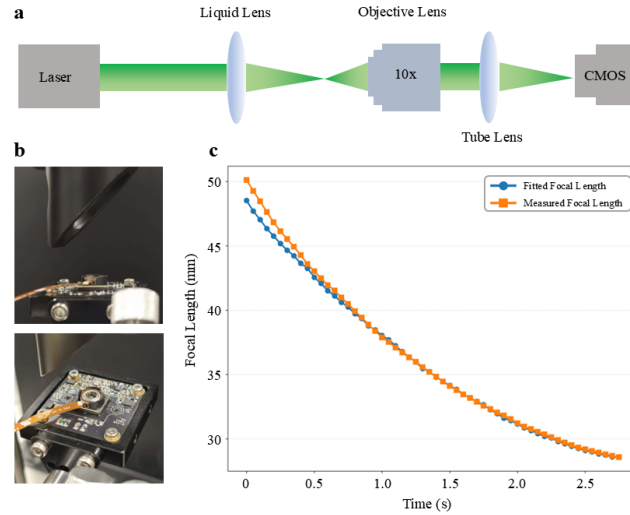


Fig. 14 Experimental validation of dynamic wavefront measurement accuracy using a voltage-controlled liquid lens. **a** Schematic of the characterization setup for calibrating the relationship between applied voltage and focal length of the liquid lens. **b** Measurement configuration with the liquid 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.

First, a dedicated characterization setup as shown in Fig.14(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) was used to synchronously translate 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, obtained with high precision, served as the ground truth for subsequent comparison. Next, the liquid lens was placed directly on the surface of our sensor, replicating the same measurement configuration as in the water droplet experiment, which is shown in Fig. 14(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 described in Supplementary S4, yielding the time-evolving phase distribution, from which the instantaneous focal length was derived.

The results are presented in Fig. 14(c). The blue curve shows the focal length extracted from our 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. This close correspondence confirms that our sensor maintains high accuracy even under dynamic conditions involving a liquid-based optical element. Together with the static accuracy benchmarks already demonstrated (3D point source localization and lens profilometry), this additional validation thoroughly addresses the reviewer's concern regarding measurement fidelity in dynamic demonstrations.

In response to the reviewer's comment, we have added the results of the liquid lens validation experiment to Supplementary Material S7 (Dynamic Wavefront Reconstruction Demonstrations).

Comment 6: The mapping figures should include coordinates or scale bars for clarity. This applies to Figure 1d, e, Figure 2, Figure 3c-h (and insets), and Figure 4b.

Author's Response:

We thank the reviewer for this valuable suggestion to improve the clarity of the figures. In response, we have added scale bars and coordinate information to all the figures mentioned (Fig. 1d, e; Fig. 2; Fig. 3c–h and their insets; and Fig. 4b) to indicate physical dimensions. We have thoroughly checked all figures to ensure that units (e.g., μm , rad) and scale bars are explicitly included, following the reviewer's suggestion. Additionally, we unified the colormap for each physical quantity: the same colormap is now used across all figures for wavefronts, gradients, heights, and residuals, respectively. This provides a consistent visual reference for the reader and facilitates cross-comparison among different panels. We believe these improvements significantly enhance the clarity and interpretability of the figures. All figures have been updated accordingly in the revised manuscript.

Conclusion

We are confident that the revisions made in direct response to the reviewers' comments have significantly enhanced the clarity, rigor, and overall impact of our manuscript. We deeply appreciate the feedback that has prompted these improvements and hope that the editors and reviewers find the revised version suitable for publication in Nature Communications.

List of major changes

1. In the **Abstract**, we have clarified that the $\pm 1^\circ$ dynamic range refers to the intrinsic geometric limit of conventional sensors, and removed the claim regarding “stress-induced birefringence” to accurately reflect demonstrated capabilities. (Responds to Reviewer #2, Comments 1 and 2)
2. In the **Introduction**, we have expanded the discussion on the geometric origin of the $\pm 1^\circ$ dynamic range by introducing an explicit formula to define this limit and adding explicit assumptions to ensure rigor, along with supporting citations. Additionally, we have incorporated recent references on metasurface-based wavefront sensing and nonlocal optical filters to better contextualize the work. (Responds to Reviewer #1, Comment 5; Reviewer #2, Comment 1)
3. In the section “**Design of meta wavefront sensor (WFS)**”, we have added explicit dependence of inter-vortex separation d in the phase mapping equation, and clarified that simultaneous LCP/RCP control is achieved via combined propagation and Pancharatnam–Berry phases. (Responds to Reviewer #3, Comments 1 and 2)
4. In the section “**Characterization of meta wavefront sensor**”, we have added a discussion on the sensitivity limit, including the minimum resolvable angle and phase gradient. (Responds to Reviewer #3, Comment 3)
5. In the **Discussion**, we have added brief discussions on broadband operation, scalability of orientation encoding (e.g., 3×3 superlattice trade-offs), and real-time computational feasibility. (Responds to Reviewer #1, Comments 2, and 4; Reviewer #2, Comments 5a; Reviewer #3, Comment 3)
6. In the **Supplementary Material S1** (DH-PSF design): we have added quantitative analysis of how key parameters (M , d , c) affect efficiency, lobe spacing, and shape, justifying the chosen parameters for LCP and RCP channels. (Responds to Reviewer #1, Comment 1)
7. In the **Supplementary Material S2** (Thermomechanical stability): we have added discussion on the thermal and mechanical stability of the optical adhesive, including tolerance to temperature-induced PSF rotation. (Responds to Reviewer #1, Comment 3)
8. In the **Supplementary Material S3** (Sensitivity limit): we have added derivation of the minimum resolvable angle and phase gradient. (Responds to Reviewer #3, Comment 3)
9. In the **Supplementary Material S4** (Wavefront reconstruction): we have reorganized algorithm description and added a new subsection validating the interpolation strategy through quantitative comparison with 45° rotation methods. (Responds to Reviewer #2, Comment 3)
10. In the **Supplementary Material S5** (DH-PSF decoding robustness): we have added systematic analysis of misclassification rates under noise, aberrations (defocus, astigmatism, coma, trefoil), and fabrication errors. (Responds to Reviewer #2, Comments 5a–5c)
11. In the **Supplementary Material S7** (Dynamic wavefront demonstrations): we have added quantitative validation using a voltage-controlled liquid lens to benchmark dynamic accuracy. (Responds to Reviewer #3, Comment 5)
12. In the **Supplementary Material S8** (Vectorial mode performance): we have added quantitative characterization of performance penalty when switching from scalar to vectorial mode, including RMSE analysis across spatial frequencies. (Responds to Reviewer #1, Comment 1)

13. In the **Supplementary Material S9** (Computational complexity): Added benchmarking of phase retrieval solvers (CVXPY, ADMM, DCT) and discussion on hardware acceleration for real-time operation. (Responds to Reviewer #1, Comment 4)
14. For **Figures**: we improved interpretability by adding scale bars, unified colormaps per quantity, and adjusted contrast where needed (e.g., Fig. 3i,j; Fig. 4b). (Responds to Reviewer #2, Comment 4; Reviewer #3, Comment 6)

Re: Manuscript ID: NCOMMS-25-98695A

Title: Chip-Integrated Metasurface with Double-Helix PSF for Ultra-Large Dynamic Range Wavefront Sensing

Dear Editors,

We thank the reviewers for their thoughtful comments and for recognizing the improvements made in the revised manuscript. Below we provide point-by-point responses to the remaining suggestions from Reviewer #2. Reviewer #1 and Reviewer #3 have recommended publication without further revisions, for which we are also grateful. All changes in the revised manuscript are highlighted in yellow for ease of review.

Response to Reviewer #2

Comment: Overall, the revised manuscript is significantly strengthened and the central technical contribution—orientation-encoded DH-PSFs combined with polarization-interleaved sampling in a CMOS-bonded architecture—now appears well supported. I therefore recommend publication in Nature Communications after minor revisions, primarily limited to final polishing and consistency checks.

Author's Response: We sincerely thank the reviewer for the positive and encouraging assessment. We greatly appreciate the recommendation for publication in Nature Communications subject to minor revisions. In the following, we address each of the remaining suggestions point by point.

Comment1: Although the current sensor is a prototype, it nevertheless achieves the intrinsic limit corresponding to a $\pm 9.4^\circ$ dynamic range (including the result in Fig. 2h), which would not be feasible without employing a metasurface. It may be helpful to emphasize this point in the Discussion as a record-breaking result, while also noting that there remains room for further improvement through algorithmic enhancements and related approaches.

Author's Response:

We thank the reviewer for this constructive suggestion. Following the recommendation, we have modified the first paragraph of the **Discussion** to explicitly highlight that the achieved $\pm 9.4^\circ$ dynamic range is a record-breaking result for a chip-integrated Shack–Hartmann wavefront sensor (SHWFS) realized purely by hardware-level encoding, while also noting that further improvement remains possible through algorithmic enhancements. The corresponding sentence now reads:

“...This decoupling of traditionally conflicting parameters represents a fundamental advance over both conventional and existing metasurface-based sensors, achieving, to the best of our knowledge, a record-breaking intrinsic geometric dynamic range ($\pm 9.4^\circ$) in a chip-integrated SHWFS purely by hardware-level encoding, while room remains for further improvement through algorithmic enhancements and related approaches.”

Comment2: In addition, double-helix PSFs are most commonly used for depth sensing, whereas this work leverages their distinctive features to reduce crosstalk. It would strengthen the manuscript to highlight this repurposing more explicitly.

Author's Response:

We thank the reviewer for this constructive suggestion. Following the recommendation, we have modified the last paragraph of the **Introduction** to explicitly highlight the repurposing of DH-PSFs. The revised text now reads:

“...Our design employs a dual-encoding strategy based on double-helix point spread functions^{43–47} (DH-PSF), which have been widely used in depth sensing and single-molecule localization. Here, we repurpose their characteristic lobe separation and orientation encoding to eliminate crosstalk under large wavefront tilts. Specifically, orientation-encoded DH-PSFs double the dynamic range per polarization channel without compromising sampling density, while orthogonal circular polarization channels with distinct lobe separations enable unambiguous interleaving of two sub-aperture arrays, doubling the sampling density without sacrificing dynamic range.”

Comment3: Finally, while the current design is optimized for a single wavelength, it would be useful to emphasize the broader potential of the platform: by changing the material platform and redesigning the metasurface, the approach must be adaptable to essentially any wavelength band.

Author's Response:

We thank the reviewer for this forward-looking comment. Following the suggestion, we have modified the second paragraph of the **Discussion** to emphasize the wavelength-adaptability of our platform. The revised text now reads:

“...While the current design is optimized for a single wavelength of 532 nm, by changing the material and redesigning the metasurface, this platform can be adapted to essentially any wavelength band. Moreover, future iterations could...

”

List of major changes

1. In the **Introduction**, we have explicitly stated the conventional applications of DH-PSFs (depth sensing and single-molecule localization) and how we repurpose their characteristic lobe separation and orientation encoding to eliminate crosstalk under large wavefront tilts. (Responds to Reviewer #2, Comment 2)
2. In the first paragraph of the **Discussion**, we have highlighted that the achieved $\pm 9.4^\circ$ dynamic range is a record-breaking result for a chip-integrated SHWFS realized purely by hardware-level encoding, while also noting that further improvement remains possible through algorithmic enhancements. (Responds to Reviewer #2, Comment 1)
3. In the second paragraph of the **Discussion**, we have added a sentence to emphasize that by changing the material and redesigning the metasurface, the platform can be adapted to essentially any wavelength band of interest. (Responds to Reviewer #2, Comment 3)