Supplemental Document



Complex-field meta-microscopy with high spatial and phase-gradient resolution: supplement

Qi Liu,[1] Wenwei Liu,[1,5] Qiang Li,[2] Duk-Yong Choi,[3] Zhancheng Li,[1] Yongliang Liu,[1] Bo Yu,[1] Yifei Xu,[1] Hua Cheng,[1,6] and Shuqi Chen[1,2,4,7,]

[1]*The Key Laboratory of Weak Light Nonlinear Photonics, Ministry of Education, School of Physics and TEDA Institute of Applied Physics,Nankai University, Tianjin 300071, China*
[2]*School of Materials Science and Engineering, Nankai University, Tianjin 300350, China*
[3]*Laser Physics Centre, Research School of Physics, Australian National University, Canberra, ACT 2601, Australia*
[4]*The Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*
[5]*wliu@nankai.edu.cn*
[6]*hcheng@nankai.edu.cn*
[7]*schen@nankai.edu.cn*

This supplement published with Optica Publishing Group on 17 November 2026 by The Authors under the terms of the Creative Commons Attribution 4.0 License in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: https://doi.org/10.6084/m9.figshare.34789212

Parent Article DOI: https://doi.org/10.1364/OPTICA.612791

COMPLEX-FIELDS META-MICROSCOPY WITH HIGH SPATIAL AND PHASE-GRADIENT RESOLUTION: SUPPLEMENTAL DOCUMENT

QI LIU[1], WENWEI LIU[1,5], QIANG LI[2], DUK-YONG CHOI[3], ZHANCHENG LI[1], YONGLIANG LIU[1], BO YU[1], YIFEI XU[1], HUA CHENG[1,6], SHUQI CHEN[1, 2, 4,7]

[1]*The Key Laboratory of Weak Light Nonlinear Photonics, Ministry of Education, School of Physics and TEDA Institute of Applied Physics, Nankai University, Tianjin 300071, China*

[2]*School of Materials Science and Engineering, Nankai University, Tianjin 300350, China*

[3]*Laser Physics Centre, Research School of Physics, Australian National University, Canberra, ACT 2601, Australia*

[4]*The Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*

[5]*email: wliu@nankai.edu.cn*

[6]*email: hcheng@nankai.edu.cn*

[7]*email: schen@nankai.edu.cn*

S1: Principle of Interference and Reconstruction in Metasurfaces

For an incident wave with a complex amplitude given by $A_0e^{i\varphi(x,\,y)}$, modulation via the metasurface transforms the wavefront into four sheared versions: $A_0e^{i\varphi(x+\Delta d/2,\,y+\Delta d/2)}$, $A_0e^{i\varphi(x+\Delta d/2,\,y-\Delta d/2)}$, $A_0e^{i\varphi(x-\Delta d/2,\,y+\Delta d/2)}$, $A_0e^{i\varphi(x-\Delta d/2,\,y-\Delta d/2)}$, where Δd denotes the shear distance. The resulting interference intensity distribution is expressed as:

$$\begin{aligned} I(x,y) = A_0^2[1 &+ \frac{1}{2}\cos\left(\Delta\varphi_x + 2k_{shear}x\right) + \frac{1}{2}\cos\left(\Delta\varphi_y + 2k_{shear}y\right) \\ &+ \frac{1}{2}\cos\left(\Delta\varphi_{x+y} + 2k_{shear}\left(x+y\right)\right) + \frac{1}{2}\cos\left(\Delta\varphi_{x-y} + 2k_{shear}\left(x-y\right)\right)] \end{aligned} \tag{S1}$$

in which $\Delta\varphi_{x/y/x\pm y}$ represent the phase differences corresponding to partial derivatives in the x, y and $x \pm y$ directions, respectively. Expanding the phase $\varphi(x, y)$ using a Taylor series:

$$\begin{aligned} \varphi\left(x \pm \frac{\Delta d}{2}, y\right) &= \varphi(x,y) \pm \frac{\partial\varphi(x,y)}{\partial x}\frac{\Delta d}{2} + \ldots \\ \varphi\left(x, y \pm \frac{\Delta d}{2}\right) &= \varphi(x,y) \pm \frac{\partial\varphi(x,y)}{\partial y}\frac{\Delta d}{2} + \ldots \end{aligned} \tag{S2}$$

when the shear distance is negligible compared with the characteristic size of the phase distribution, it can be approximated as

$$\begin{aligned} \Delta\varphi_x &= \varphi\left(x + \frac{\Delta d}{2}, y\right) - \varphi\left(x - \frac{\Delta d}{2}, y\right) \\ &\approx \frac{\partial\varphi(x,y)}{\partial x}\Delta d \end{aligned} \tag{S3}$$

Therefore, by substituting the phase difference approximations in each direction into Eq. (S1), the interference intensity distribution in Eq. (1) can be obtained.

To reconstruct the phase from its gradients in the x- and y-directions, we employ the higher-order finite-difference-based least-squares integration (HFLI) method. Compared to lower-order techniques—such as traditional finite-difference-based least-squares integration (TFLI)[1], cosine transform integration (CTI)[2], and radial basis function integration (RBFI)[3]—the HFLI method offers improved accuracy and computational efficiency. This advantage arises from its fifth-order error convergence rate

O(Δd^5), which allows finer phase details to be captured more precisely[4]. For a phase function φ defined on an M×M grid, the relationship between the phase and its gradient g is given by:

$$\begin{cases} \dfrac{\varphi_{m,n+1}-\varphi_{m,n}}{\Delta d}=\dfrac{g^x_{m,n+1}+g^x_{m,n}}{2}, & \begin{matrix} m=1,2,\ldots,M \\ n=1,2,\ldots,M-1 \end{matrix} \\ \dfrac{\varphi_{m+1,n}-\varphi_{m,n}}{\Delta d}=\dfrac{g^y_{m+1,n}+g^y_{m,n}}{2}, & \begin{matrix} m=1,2,\ldots,M-1 \\ n=1,2,\ldots,M \end{matrix} \end{cases} \tag{S4}$$

where $\varphi_{m,n}$ denotes the phase at coordinates (m, n), Δd is the shear distance, and $g^x_{m,n}$ and $g^y_{m,n}$ represent the phase gradients in the x and y directions, respectively. By expressing $\varphi_{m,n}$ and $g_{m,n}$ in matrix form and applying the least squares method, the reconstructed phase can be obtained as follows:

$$\begin{bmatrix} \varphi_{1,1} \\ \varphi_{2,1} \\ \vdots \\ \varphi_{M,M} \end{bmatrix} = \left(\mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{G} \tag{S5}$$

where D is a coefficient matrix and can be written as:

$$\mathbf{D}=\begin{bmatrix} D^x \\ D^y \end{bmatrix}=\begin{bmatrix} -1 & 0 & \ldots & 0 & 1 & 0 & \ldots & \ldots & 0 \\ 0 & -1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \ldots & 0 & -1 & 0 & \ldots & 0 & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ -1 & 1 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & -1 & 1 & 0 & \ldots & \ldots & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & -1 & 1 \end{bmatrix} \tag{S6}$$

and the vector G is the phase gradient matrix:

$$\mathbf{G}=\begin{bmatrix} \mathbf{G}^x \\ \mathbf{G}^y \end{bmatrix}=\frac{1}{24}\begin{bmatrix} 12(g^x_{1,2}+g^x_{1,1})\Delta d \\ (-g^x_{1,4}+13g^x_{1,3}+13g^x_{1,2}-g^x_{1,1})\Delta d \\ (-g^x_{1,5}+13g^x_{1,4}+13g^x_{1,3}-g^x_{1,2})\Delta d \\ \vdots \\ 12(g^x_{M,M}+g^x_{M,M-1})\Delta d \\ \cdots\cdots\cdots\cdots \\ 12(g^y_{2,1}+g^y_{1,1})\Delta d \\ (-g^y_{4,1}+13g^y_{3,1}+13g^y_{2,1}-g^y_{1,1})\Delta d \\ (-g^y_{5,1}+13g^y_{4,1}+13g^y_{3,1}-g^y_{2,1})\Delta d \\ \vdots \\ 12(g^y_{M,M}+g^y_{M-1,M})\Delta d \end{bmatrix} \tag{S7}$$

Based on the aforementioned principle, the interference process of the metasurface can be calculated. The simulation reconstruction results are compared with the ground truth as illustrated in Fig. S1.

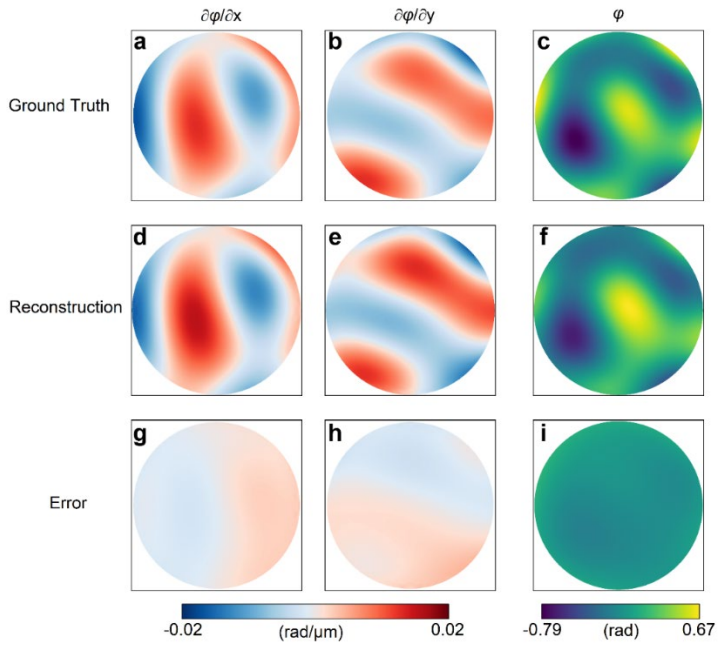


Fig. S1. Comparison of simulation-based phase gradient and phase reconstruction. (a) Gradient of the input phase along the *x*-axis. (b) Gradient of the input phase along the y-axis. (c) Input phase. (d) Reconstructed gradient along the *x*-axis. (e) Reconstructed gradient along the *y*-axis. (f) Reconstructed phase. (g) Reconstruction error of the phase gradient along the *x*-axis. (h) Reconstruction error of the phase gradient along the *y*-axis. (i) Phase reconstruction error.

S2: Design and Optimization for Polarization-Multiplexed Metasurfaces

For the design of highly efficient polarization-multiplexed metasurfaces, we employ a precomputed database based on parameterized unit cells. The structural parameter space is defined as $\mathbf{P}=\left\{\left(L_x,L_y\right)\mid L_x\in\mathbf{L}_x,L_y\in\mathbf{L}_y\right\}$, where L_x and L_y denote the characteristic dimensions of the unit cell along the x and y directions, respectively. Through systematic scanning of this parameter space, a comprehensive optical response database is established. For any structural parameter combination $\mathbf{p}=\left(l_x,l_y\right)\in\mathbf{P}$, the energy transmittance (T_x, T_y) and transmission phase delay (φ_x, φ_y) under orthogonally and linearly polarized incident light are computed via full-wave electromagnetic simulation using the finite-difference time-domain (FDTD) method. The phase response $\varphi_{x,y}(\mathbf{p})\in[0,2\pi)$ and the transmittance $T_{x,y}(\mathbf{p})\in[0,1]$ are defined accordingly. This database establishes a mapping from structural parameters to optical responses in the parameter space, supporting the design of metasurfaces. To enlarge the overall optical efficiency of the metasurface device, we first select the database for high-efficiency structures. A transmission threshold of T_{limit} = 0.85 is imposed, requiring high transmission under both orthogonal polarizations:

$$\mathbf{S}=\left\{\mathbf{p}\in\mathbf{P}\mid T_x\left(\mathbf{p}\right)\geq T_{limit}\ \text{and}\ T_y\left(\mathbf{p}\right)\geq T_{limit}\right\} \tag{S8}$$

The resulting candidate set S comprises all parameter sets that meet the high-efficiency criterion.

The metasurface is designed to realize target phase profiles $\varphi_x^{\text{target}}$ and $\varphi_y^{\text{target}}$ for two orthogonal polarizations. We adopt absolute error as the metric function: $\Delta(\varphi_1,\varphi_2)=|\varphi_1-\varphi_2|$. Then, we construct the multi-objective phase error function for *x* and y polarizations:

$$\varepsilon(\mathbf{p};\alpha)=\alpha\Delta\left(\varphi_x^{\text{target}},\varphi_x(\mathbf{p})\right)+(1-\alpha)\Delta\left(\varphi_y^{\text{target}},\varphi_y(\mathbf{p})\right) \tag{S9}$$

Here, α is the weighting coefficient used to adjust the relative priority of phase optimization for *x*- and *y*-polarizations according to the design objectives. In this work, because crosstalk exists between polarization channels, lens focusing yields higher energy concentration than shearing interference. Thus, to minimize the crosstalk introduced by lens focusing during shearing interference in the *y*-polarization, α was set to less than 0.5. In the optimization process, we set α to 0.3. The optimal structural parameter $\mathbf{p}^*$ is obtained by minimizing the weighted phase error function:

$$\mathbf{p}^*=\arg\min_{\mathbf{p}\in\mathbf{S}}\varepsilon(\mathbf{p};\alpha) \tag{S10}$$

This optimization problem is solved in the discrete parameter space S and can be implemented via traversal search.

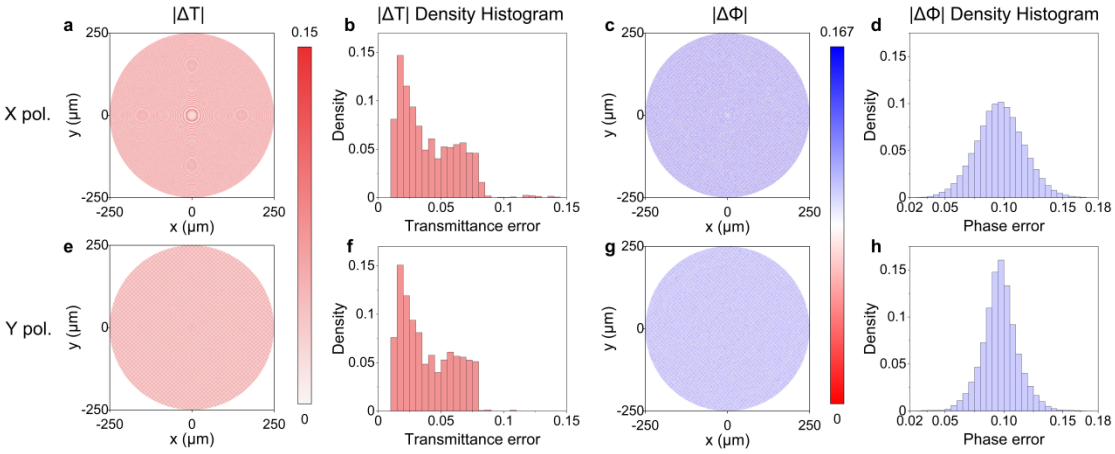


Fig. S2. Spatial error maps and statistical distributions of the metasurface. (a, c) Spatial maps of transmittance error (|ΔT|) and phase delay error (|ΔΦ|) for the x-polarization, alongside (b, d) their density histograms defined by the normalized amount the meta-atoms located in each error interval. (e-h) The corresponding error maps and histograms for the y-polarization channel.

The error maps of our optimized metasurface are shown in Fig. S2. These maps quantify the deviation of amplitude and phase delay caused by the discrete meta-atoms library. Furthermore, simulations based on the actual metasurface design reveal that the crosstalk from the *x*-polarization (focusing) to the *y*-polarization (shearing interference) channel is below –34 dB, while the reverse crosstalk is below –25 dB. We emphasize that this level of crosstalk does not degrade the final reconstruction due to the distinct spatial energy distributions of the two channels. The *x*-polarization channel forms a highly localized focal spot (on the order of λ/NA), whereas the *y*-polarization channel generates a non-local interference pattern distributed across the entire aperture. Consequently, the intensity contrast of the localized spot can exceed 10^6. We also numerically evaluate the reconstruction performance under varying signal-to-noise ratios

(SNR). The results confirm that the experimental crosstalk has a negligible impact on the phase and image reconstruction accuracy.

To quantify the residual polarization crosstalk, we define the crosstalk level as $CT = 10\log_{10}(P_{ct}/P_{target})$, where P_{ct} and P_{target} are the optical powers from crosstalk and target channels, respectively. In the present design, the *y*-polarization channel generates the shearing interference pattern for phase reconstruction, whereas the *x*-polarization channel produces a diffraction-limited focusing spot. Therefore, leakage from the *x*-polarization channel to the *y*-polarization channel appears as a weak localized spot superimposed on the non-localized interference pattern. This crosstalk is mainly associated with the simultaneous encoding of two polarization-dependent phase profiles in the same anisotropic meta-atoms, while the residual crosstalk observed experimentally can be further affected by fabrication deviations.

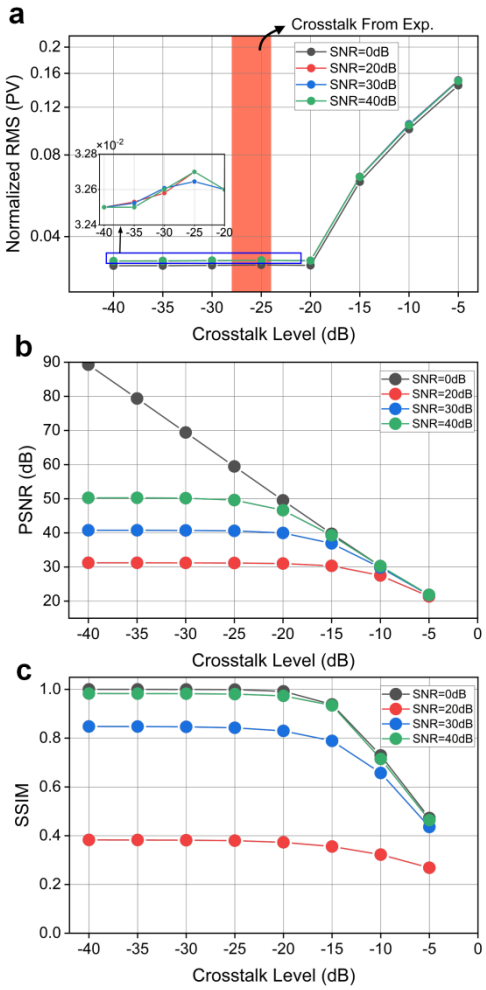


Fig. S3. Effects of crosstalk level on system performance under different SNRs. (a) Phase reconstruction accuracy (evaluated by normalized RMS) as a function of the crosstalk level under different SNRs. (b,c) Imaging quality versus crosstalk level under different SNRs, quantified by (b) PSNR and (c) SSIM, respectively.

We further evaluate the influence of crosstalk and detection noise on the reconstruction performance. The crosstalk level extracted from Figs. 2E and 2H is approximately –24 to –27 dB, and the typical SNR of the detection system is about 44 dB based on the Hamamatsu C13440-20CU sCMOS camera used in the experiments. As shown in Fig. S3, when the crosstalk level is below –20 dB, the relative root-mean-square error of the reconstructed phase remains around 3% PV. For the intensity-imaging channel, the

reconstruction quality was assessed using the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [5]. The calculated PSNR and SSIM under the experimentally relevant crosstalk level indicate that the residual crosstalk has a limited influence on the reconstructed intensity images (the definitions of the metrics can be found in S13).

S3: Sensitivity Analysis and Experimental Validation

We consider a scenario of shear interference along the x-direction. The intensity distribution of the shear interference is given as

$$I(x) = 2\left(1+\cos\left(2k_{shear}x + g_{\varphi}\Delta d\right)\right) \tag{S11}$$

where $g_{\varphi} = \Delta\varphi/\Delta d$ is the gradient of phase. Taking the variation of Eq. (S11)

$$\delta I(x) = -2\sin\left(\Delta\varphi\Delta d + 2k_{shear}x\right)\cdot\left(\Delta d\,\delta g_{\varphi}\right) \tag{S12}$$

and the δg_{φ} can be written as

$$\delta g_{\varphi} = \frac{\delta I(x)}{-2\Delta d\sin\left(g_{\varphi}\Delta d + 2k_{shear}x\right)} \tag{S13}$$

Note that when $|\sin(\cdot)|=1$, the phase gradient sensitivity reaches the maximum value. The minimal detectable intensity variation derived from the interference pattern correlates with the SNR of the imaging system. The SNR is defined as [6]:

$$\mathrm{SNR} = 20\lg\left(\frac{\mu_{sig}}{\sigma_{noi}}\right) \tag{S14}$$

with μ_{sig} representing the signal mean value and σ_{noi} standard deviation of noise. Thus, the maximum phase sensitivity (i.e., the minimum phase gradient resolution) can be obtained as:

$$g_{\mathrm{lim}} = \frac{10^{-\left(\frac{\mathrm{SNR}}{20}\right)}}{\Delta d} \tag{S15}$$

The resolution of phase gradient depends on the shear distance and the SNR. The shear distance Δd is given by $\Delta d = 2z\tan\theta$, where θ is the angle between the split beams and the optical axis, and z is the detection distance. Based on Eq. (S15) the relationship between θ and z is shown for various SNR values (Fig. S4a). Furthermore, the minimum resolvable phase, $\Delta\varphi_{\mathrm{lim}}$, can be determined by $\Delta\varphi_{\mathrm{lim}} = g_{\mathrm{lim}}\times\Delta d$. The variation of $\Delta\varphi_{\mathrm{min}}$ across different SNR levels is illustrated in Fig. S4b.

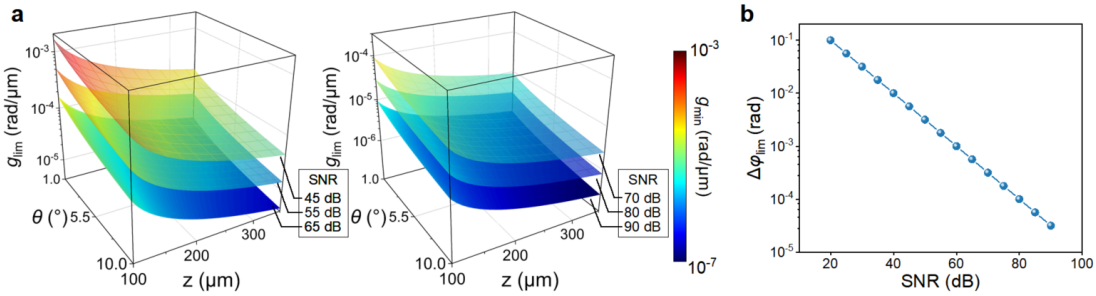


Fig. S4. Theoretical dependence of phase detection limits on system parameters and SNRs. (a) Minimum detectable phase gradient (g_{lim}) versus shear angles and propagation distances at varying SNRs. (b) Absolute phase sensitivity ($\Delta\varphi_{\mathrm{lim}}$) as a function of SNR.

To experimentally evaluate the SNR-dependent phase-gradient sensitivity, the programmed phase and camera exposure time were varied to produce different phase gradients and interferogram. For each experimental condition, N raw interferograms were recorded while the camera gain, illumination and optical configuration were fixed. Dark frames were acquired using the same exposure time and camera gain. The temporal mean and variance at pixel p were calculated as:

$$\bar{I}_p = \frac{1}{N}\sum_{k=1}^{N} I_{k,p} \tag{S16}$$

and

$$s_p^2 = \frac{1}{N-1}\sum_{k=1}^{N}\left(I_{k,p} - \bar{I}_p\right)^2 \tag{S17}$$

where $I_{k,p}$ is the intensity recorded at pixel p in frame k, the dark-corrected signal mean and temporal noise were calculated as

$$\mu_{\mathrm{sig}} = \frac{1}{M}\sum_{p}\left(\bar{I}_p - \bar{D}_p\right), \quad \sigma_{\mathrm{noi}} = \sqrt{\frac{1}{M}\sum_{p} s_p^2}. \tag{S18}$$

where $\bar{D}_p$ is the mean dark-frame value at pixel p. The effective SNR was calculated as:

$$\mathrm{SNR}_{\mathrm{eff}} = 20\log_{10}\left(\frac{\mu_{sig}}{\sigma_{noi}}\right) \tag{S19}$$

For each combination of programmed phase and exposure time, eight interferograms were acquired. Exposure times, ranging from 50 to 200 ms in 25 ms step, were used. The corresponding effective SNRs estimated using Eq. (S19) were 21, 29, 33, 36, 37, 39, and 40 dB, respectively. Figure S5 presents the reconstruction results at the calculated SNRs (columns) and different programmed phase PVs with their corresponding phase gradients (rows). A reconstruction was classified as failed when its normalized RMSE ($\mathrm{RMSE}_{\mathrm{recon}}/\mathrm{RMS}_{\mathrm{SLM}}$) was greater than 1[7], indicating that the reconstruction error had reached or exceeded the RMS magnitude of the target phase. Valid reconstructions with normalized RMSE ≤ 1 are outlined by red boxes.

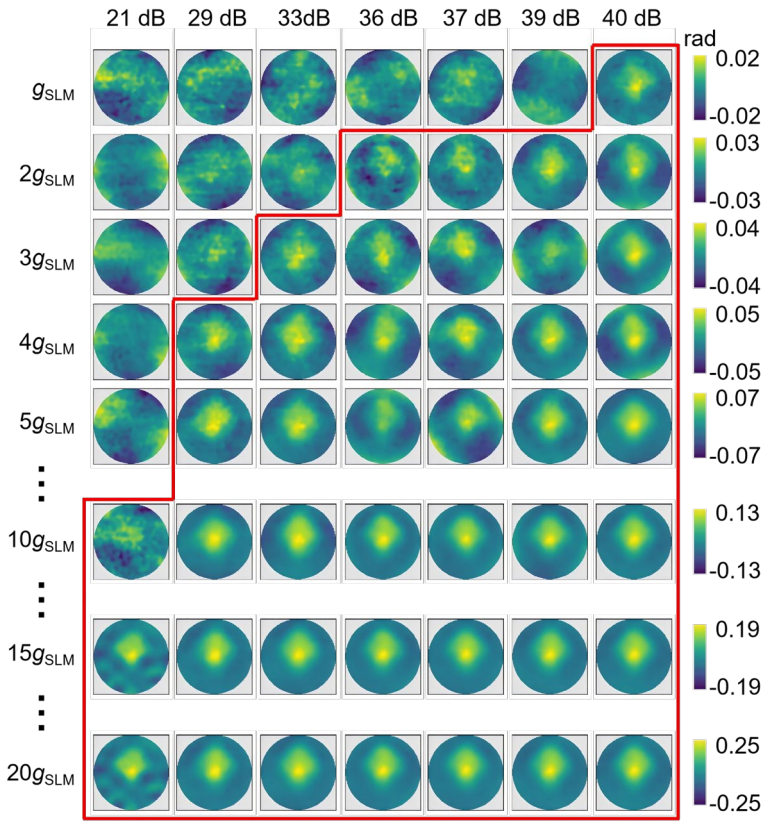


Fig. S5. Phase-gradient reconstruction at different programmed phase variations and exposure times. Rows correspond to different programmed PV values and reference phase gradients, whereas columns correspond to different exposure times and measured SNRs.

Note that the SNR is a system-level quantity determined by temporal intensity fluctuations from the camera, illumination source, optical setup, and experimental environment. For the CMOS detector, the noise includes photon shot noise, dark noise, and read noise. Fluctuations in the illumination source, optical path, and experimental environment can also perturb the phase and spatial coherence of the incident wavefront. Under a first-order approximation, a small perturbation $\delta(\Delta\varphi)$ in the phase difference $\Delta\varphi$ between the two sheared wavefronts induces a change in the interference intensity $I(x)$ given by:

$$\delta I(x) = -2\sin\left(2k_{shear}x + \Delta\varphi\right)\delta\left(\Delta\varphi\right) \tag{S20}$$

Since the sine term is bounded by [−1, 1], the intensity change satisfies $|\delta I| \le 2|\delta(\Delta\varphi)|$. The maximum sensitivity occurs when $|\sin(\cdot)| = 1$. Under this condition, the relationship simplifies to $\delta I = \pm 2\delta(\Delta\varphi)$.

Consequently, the phase gradient perturbation induced by a noise source that causes an intensity change δI can be expressed as:

$$\left|\delta g_{\varphi}\right| = \frac{\left|\delta(\Delta\varphi)\right|}{\Delta d} = \frac{|\delta I|}{2\Delta d} \tag{S21}$$

In shear interference, the phase difference between the two sheared wavefronts is $\Delta\varphi = \varphi(x) - \varphi(x - \Delta d)$. Phase noise results in the perturbation of $\Delta\varphi$, presented as

$\delta(\Delta\varphi) = \delta\varphi(x) - \delta\varphi(x-\Delta d)$. The noise exhibits an average value of $\langle\varphi\rangle = \varphi_n$. Since the addition of an arbitrary constant phase to the phase distribution does not alter the relative phase relationships, $\langle\varphi\rangle$ is negligible and is therefore set to zero. The variance of this perturbation is:

$$\begin{aligned}\langle\left(\delta(\Delta\varphi)\right)^2\rangle &= \langle\delta\varphi(x)^2\rangle + \langle\delta\varphi(x-\Delta d)^2\rangle - 2\langle\delta\varphi(x)\cdot\delta\varphi(x-\Delta d)\rangle \\ &= \sigma_\varphi^2 + \sigma_\varphi^2 - 2\langle\delta\varphi(x)\cdot\delta\varphi(x-\Delta d)\rangle \\ &= 2\sigma_\varphi^2\left[1-C\left(\Delta d\right)\right]\end{aligned} \tag{S22}$$

where σ_φ is the root mean square (RMS) of the noise. $C(\Delta d)$ is the normalized spatial autocorrelation function of the phase noise. Typically, phase noise arises from atmospheric turbulence or scattering by dust particles[8]. The autocorrelation term of the noise depends on both the shear distance and the correlation length L_c of the incident light. Consequently, $C(\Delta d)$ can be more generally expressed as $C(\Delta d, L_c)$. The specific form of $C(\Delta d, L_c)$ is related to the actual situation, such as the Gaussian type random phase distribution or the phase fluctuations caused by atmosphere turbulence. However, all these cases show the same trend and monotonicity: $C(\Delta d, L_c)$ increases as the coherence length L_c increases, and decreases as the shear distance Δd increases. The uncertainty in the phase gradient measurement originating solely from the incident phase noise is given by the standard deviation of δg_φ in Eq. (S21):

$$\begin{aligned}(g_{\mathrm{lim}})_{\mathrm{opt}} = \sigma_{\mathrm{noise},g} &= \frac{\sqrt{\left\langle\delta\left(\Delta\varphi\right)^2\right\rangle}}{\Delta d} \\ &= \frac{1}{\Delta d}\sqrt{2\sigma_\varphi^2\left[1-C\left(\Delta d, L_c\right)\right]}\end{aligned} \tag{S23}$$

Therefore, the noise of the phase gradient on the incident side correlates with the noise σ_φ of the phase, the shear distance Δd, and the coherence length of the incident light (L_c).

By combining the two kinds of phase sensitivity $(g_{\mathrm{lim}})_{\mathrm{CMOS}}$ and $(g_{\mathrm{lim}})_{\mathrm{opt}}$, and given that the noises in CMOS detection and in the optical setup are statistically independent, the total phase gradient resolution can be expressed as the root-sum-square (RSS) of their individual contributions:

$$\begin{aligned}(g_{\mathrm{lim}})_{\mathrm{total}} &= \sqrt{\left((g_{\mathrm{lim}})_{\mathrm{CMOS}}\right)^2 + \left((g_{\mathrm{lim}})_{\mathrm{opt}}\right)^2} \\ &= \frac{1}{\Delta d}\sqrt{2\sigma_\varphi^2\left[1-C\left(\Delta d, L_c\right)\right] + \frac{10^{-\left(\frac{\mathrm{SNR_{CMOS}}}{10}\right)}}{4}} \\ &= \frac{10^{-\left(\frac{\mathrm{SNR}}{20}\right)}}{\Delta d}\end{aligned} \tag{S24}$$

This expression demonstrates that the precision of phase gradient measurements is affected by both incident phase noises (σ_φ, L_c) and detector noises (related to the SNR of CMOS). Based on the analysis, phase gradient resolution can be enhanced by increasing the SNR of the CMOS sensor, improving the coherence of the incident light, and mitigating environmental disturbances.

In addition to the phase gradient resolution, the spatial resolution of the intensity image itself is fundamentally governed by diffraction limit. In this work, we perform a metalens in *x*-polarization. Its

resolution is limited by the diffraction limit, which is determined by the NA of the metalens and the wavelength λ. For a coherent imaging system, the minimum resolvable feature size is on the order of λ / NA. This diffraction-limited performance for intensity imaging serves as the foundation for the subsequent phase gradient super-resolution analysis.

S4: Eliminating System Phase Error to Reconstruct a More Accurate Phase

In theoretical calculations and simulations, the incident light is modeled as an ideal plane wave with superimposed phases from the incident source and the metasurface. However, in experimental conditions, the incident light exhibits aberration φ_{in}, the imaging system introduces phase error φ_{lens}, and the metasurface possesses phase error φ_{ms}. These components collectively form the system phase error φ_{err}. The direct reconstruction from an interferogram yields $\varphi_{\mathrm{recon}} = \varphi_{\mathrm{aim}} + \varphi_{\mathrm{err}}$. To eliminate the system phase error and enhance the reconstruction of the target phase, the metasurface-based complex amplitude imaging system must be calibrated prior to the reconstruction experiment. This calibration involves acquiring a single interferogram in the absence of the phase to be measured, which encapsulates the φ_{err} information. Perform step 1 and 2 of the reconstruction process outlined in Fig. 2i on the acquired interferogram to yield spectral information in the x and y directions, denoted as $S_{x_err}(k_x, k_y)$ and $S_{y_err}(k_x, k_y)$, which incorporate the system phase error. Then, apply the identical steps to the interferogram to which the phase under test has been added, obtaining $S_{x_all}(k_x, k_y)$ and $S_{y_all}(k_x, k_y)$ that encompass both the system phase error and the phase to test. Apply the inverse Fourier transform to the two acquired spectrograms. According to Eq. (2) in the maintext, we have

$$\mathrm{IFT}\left[S_{i_sys}\left(k_x, k_y\right)\right] = \frac{1}{2} I_0\left(x, y\right) e^{i\Delta d \frac{\partial \varphi_{i_err}}{\partial x}} \tag{S25}$$

$$\mathrm{IFT}\left[S_{i_all}\left(k_x, k_y\right)\right] = \frac{1}{2} I_0\left(x, y\right) e^{i\Delta d \frac{\partial \varphi_{i_all}}{\partial x}} \tag{S26}$$

where i corresponds to the spectrogram and phase gradient along the x and y directions. By combining Eqs. (S25) and (S26)

$$\frac{\mathrm{IFT}\left[S_{i_all}\left(k_x, k_y\right)\right]}{\mathrm{IFT}\left[S_{i_err}\left(k_x, k_y\right)\right]} = \exp\left(i\Delta d\left(\frac{\partial \varphi_{i_all}}{\partial x} - \frac{\partial \varphi_{i_err}}{\partial x}\right)\right) \tag{S27}$$

and then performing steps 3 and 4 in Fig. 2i to reconstruct φ_{aim}, this method eliminates system phase errors. As shown in Fig. S6, the systematic phase errors can be eliminated by the proposed method.

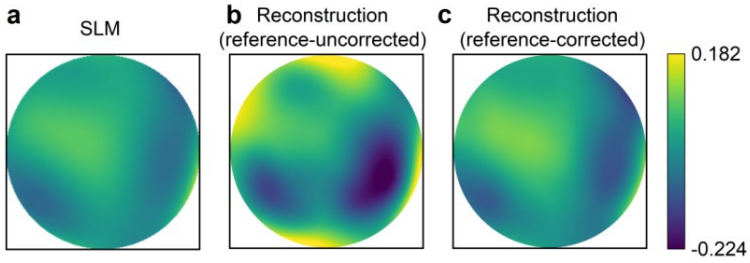


Fig. S6. Comparison of reconstruction results before and after systematic phase error correction. (a) Wavefront modulated by the SLM. (b) Uncorrected reconstruction result. (c) Corrected reconstruction result.

S5: Analysis of PGSR Network Architecture and Performance

We used a lightweight physics-based guided phase super-resolution network, termed PGSR, to combine the high-resolution (HR) intensity image and the low-resolution (LR) phase measurement. The intensity image provides high-spatial-frequency structural information, whereas the LR phase image provides the quantitative phase baseline.

The network follows a residual reconstruction strategy. Rather than directly predicting the full HR phase, it estimates a residual phase component $\Delta\varphi$, which is added to the upsampled (Up(·)) LR phase:

$$\varphi'_{\mathrm{HR}} = \mathrm{Up}(\varphi_{\mathrm{LR}}) + \Delta\varphi \tag{S28}$$

This formulation helps preserve the measured low-frequency phase information while recovering the missing high-frequency phase details.

PGSR consists of two input branches. The intensity branch extracts structural features from the HR intensity image, and the phase branch extracts phase features from the LR phase image and upsamples them to the HR grid. The two feature streams are fused by a local structure–phase modulation module. This module combines the intensity and phase features within a small 5×5 neighborhood, allowing the network to use local structural information from the intensity image to guide phase detail recovery[9].

The fused features are then passed to a reconstruction head to generate the residual phase. In the default setting, the network uses 64 feature channels, four residual blocks in each input branch, four residual blocks in the reconstruction head, and a 5×5 local modulation window. The network contains approximately 1.38 million trainable parameters.

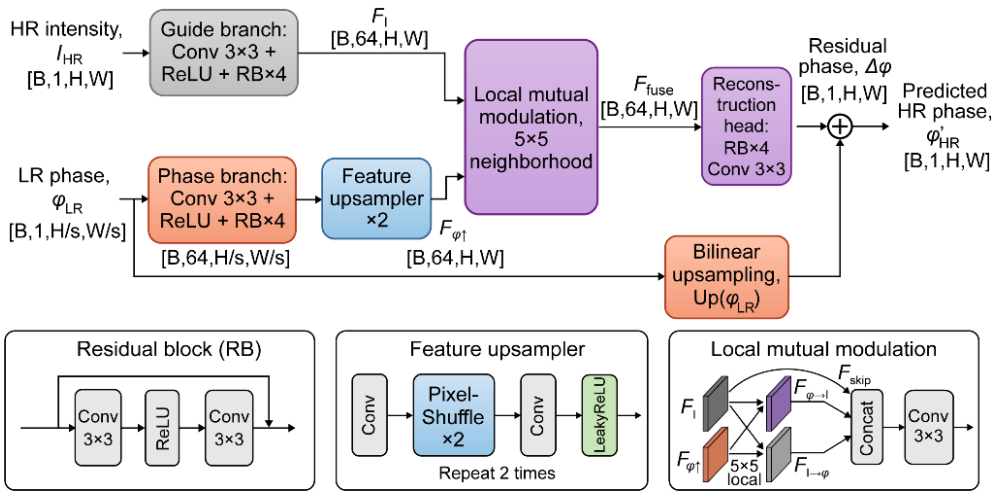


Fig. S7. Architecture of PGSR-Net.

The physical forward model is introduced to ensure that self-supervised training is consistent with the actual optical measurement. The measured LR phase is generated by the shear-interferometric acquisition and phase reconstruction process, rather than by simple digital downsampling. Therefore, conventional empirical degradation models, such as bicubic downsampling or Gaussian blurring, cannot fully represent the effects of shear displacement and Fourier-domain spatial filtering[10,11]. During training, the predicted HR phase is passed through a differentiable physical forward model H to generate a simulated LR phase:

$$\varphi'_{\mathrm{LR}} = H\left(\varphi'_{\mathrm{HR}}\right) \tag{S29}$$

The physical consistency loss is then defined as the L_1 distance (L_1 loss[12]), namely the mean absolute phase difference between the simulated and measured LR phase:

$$L_{\text{phys}} = \left\| \varphi'_{\text{LR}} - \varphi_{\text{LR}} \right\|_1 \tag{S30}$$

This term constrains the predicted phase in the measurement domain and enables self-supervised optimization without requiring HR phase labels. However, because the physical forward model maps the predicted HR phase to the LR measurement domain, this constraint mainly enforces consistency within the observable LR phase band. The missing high-frequency phase components remain underdetermined.

To examine whether the HR intensity image can provide structural information for the missing high-frequency phase components, we performed a frequency-dependent mutual information (MI) analysis[13,14] on an open-source FPM dataset [15], which provides paired HR intensity and phase images. MI measures the statistical dependence between two variables and is defined as

$$I(\mathbf{X};\mathbf{Y}) = \sum_{\mathbf{x}}\sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y}) \log \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \tag{S31}$$

where $p(\mathbf{x}, \mathbf{y})$ is the joint probability distribution and $p(\mathbf{x})$, $p(\mathbf{y})$ are the marginal distributions. The Fourier spectra of the intensity and phase images were divided into six concentric spatial-frequency rings with a ring width of $\Delta k = k_{\text{max}}/6$, where $k_{\text{max}} = 1/2\text{pix}$ (pix is the pixel size of the image), as shown in Fig. S8b. Each ring component was extracted by annular bandpass filtering and transformed back to the spatial domain. The MI between each intensity component and each phase component was then calculated to construct the frequency-dependent MI matrix.

As shown in Fig. S7a, the MI is strongest near the diagonal of the matrix, indicating that intensity and phase components at similar spatial-frequency bands share stronger statistical dependence. Importantly, the MI is still retained in the high-frequency region, suggesting that high-frequency intensity structures contain statistical information related to high-frequency phase structures. The spatial-frequency ring decomposition used in this analysis is schematically shown in Fig. S7b. These results support the use of the HR intensity image as a weak structural prior for localizing missing high-frequency phase details. However, the finite MI does not imply that the intensity image uniquely determines the phase. In our PGSR framework, the quantitative phase baseline remains constrained by the measured LR phase through the physical consistency loss, while the HR intensity image only provides auxiliary structural guidance.

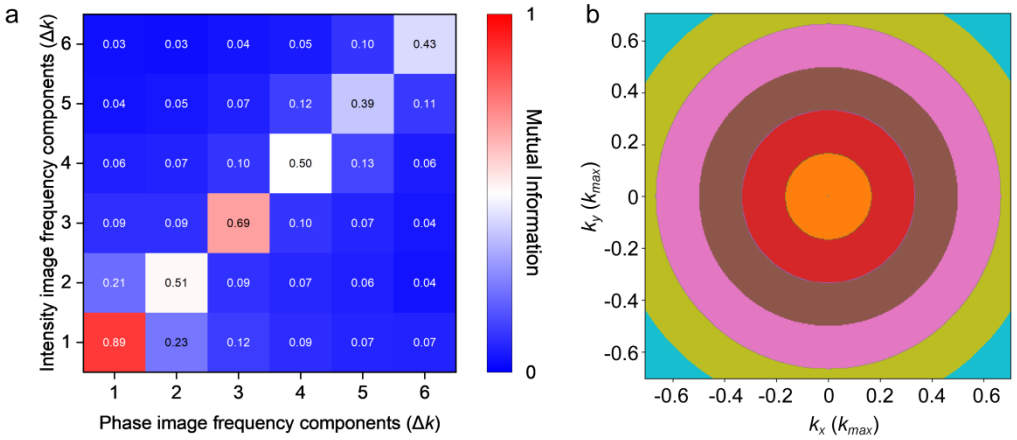


Fig. S8. Frequency-dependent MI between intensity and phase images. (a) MI matrix calculated between intensity and phase components decomposed into different spatial-frequency rings. (b) Schematic of the spatial-frequency ring decomposition used for the analysis. Different colors denote different frequency rings in the Fourier domain.

Based on this observation, we included a high-frequency structural regularization term derived from the HR intensity image. The total training loss is written as

$$L_{\text{total}} = L_{\text{phys}} + \alpha L_{\text{struc}} \tag{S32}$$

where $L_{\text{struc}} = 1 - \text{Corr}(S(\varphi'_{\text{HR}}), S(I_{\text{HR}}))$ quantifies the structural consistency between the predicted phase and the measured intensity in the high-frequency domain. Here, $S(\cdot)$ denotes the high-frequency extraction operator, $\text{Corr}(\cdot, \cdot)$ denotes normalized correlation, and α is the weighting coefficient. Since the correlation is computed from the normalized magnitudes of the extracted high-frequency components, this term encourages spatially consistent structural features without imposing a direct quantitative correspondence between intensity and phase.

To evaluate the performance of PGSR in phase imaging tasks, a dataset was constructed using HR, large-field-of-view intensity and phase image pairs obtained from open-source Fourier ptychographic microscopy (FPM)[15]. As the FPM dataset consists of large-field-of-view cell images 2.1 × 1.8 mm^2, we randomly cropped the images into patches corresponding to the metasurface field of view and further augmented them by rotation to construct the validation dataset. Instead of directly downsampling the HR phase images, the corresponding LR phase images were generated by simulating the propagation of the shearing interferograms and subsequently reconstructing the phase according to Eq. (2). This procedure mimics the actual phase-extraction process in the *y*-polarized interferometric imaging channel. To further examine the contribution of each component in the PGSR framework, we performed an ablation study by comparing different degradation models and loss functions. Five methods were evaluated: direct upsampling of the LR phase image as the baseline, average degradation with an L_1 consistency loss, physics-based degradation with an L_1 consistency loss, average degradation with the proposed high-frequency structural prior, and physics-based degradation with the proposed structural prior, corresponding to the final PGSR model.

The average degradation model simulates the LR phase image by simple spatial averaging/downsampling, whereas the physics-based degradation model better accounts for the optical degradation process in the imaging system. As summarized in Table S1, incorporating the high-frequency structural prior consistently improves the reconstruction quality compared with using the L_1 loss alone. Under the average degradation model, adding the structural prior improves the PSNR from 23.1296 dB to 24.4522 dB and the SSIM from 0.8020 to 0.8531. Under the physics-based degradation model, the structural prior improves the PSNR from 23.7605 dB to 24.9022 dB and the SSIM from 0.8286 to 0.8664. The best performance is achieved by the full PGSR model, which combines the physics-based degradation model with the weak high-frequency structural regularization. These ablation results validate the two key design choices of PGSR: the physics-based degradation model better represents the actual *y*-polarized interferometric phase-extraction process, while the high-frequency structural prior derived from the intensity image improves structural fidelity without enforcing a one-to-one intensity–phase mapping. It is noted that the average-degradation model with only the L_1 consistency loss performs worse than direct upsampling because the self-supervised LR-domain constraint is insufficient to regularize the

underdetermined HR components when the degradation model is mismatched to the actual shear-interferometric measurement.

Representative PGSR results on the validation dataset are shown in Fig. S9, including the HR intensity image, ground-truth (GT) phase image, degraded LR phase image, and super-resolved phase image. The PSNR and SSIM values before and after PGSR processing are also compared, showing improved reconstruction accuracy and structural fidelity. The HR FPM phase images were used only as references for quantitative evaluation and were not used in the phase super-resolution reconstructions shown in Figs. 4c–4e and 4g–4i.

Method	PSNR (dB)	SSIM
Direct up-sampling baseline	23.9277	0.7820
Average degradation + L_1 loss	23.1296	0.8020
Physics-based degradation + L_1 loss	23.7605	0.8286
Average degradation + **L_1 loss with structural prior**	24.4522	0.8531
Physics-based degradation + L_1 loss with structural prior	**24.9022**	**0.8664**

Table S1. Ablation study of the degradation model and high-frequency structural prior in PGSR. Five reconstruction settings are compared on the dataset[15]. The direct upsampling baseline denotes bilinear upsampling of the simulated LR phase image without PGSR reconstruction. The average degradation model maps the predicted HR phase to the LR domain by conventional spatial averaging/downsampling[11]. In contrast, the physics-based degradation model simulates the optical degradation in the *y*-polarized channel, including shear displacement and the spatial filtering of the phase-reconstruction process, to generate the corresponding LR phase constraint. The L_1 loss denotes the mean absolute difference between the simulated and target LR phase maps[12]. The structural prior refers to the weak high-frequency structural prior derived from the HR intensity image. The final PGSR model combines the physics-based degradation model, L_1 consistency loss and structural prior, and is highlighted in red. PSNR and SSIM were computed between the reconstructed phase images and the GT.

In the actual implementation, $S(\cdot)$ was implemented using Sobel-gradient magnitude extraction[16,17]. Specifically, Sobel operators were applied to the predicted HR phase and the measured HR intensity image to obtain their gradient-magnitude maps

$$\begin{aligned} G_\varphi &= \sqrt{\left(K_x \otimes \varphi'_{\mathrm{HR}}\right)^2 + \left(K_y \otimes \varphi'_{\mathrm{HR}}\right)^2 + \varepsilon} \\ G_I &= \sqrt{\left(K_x \otimes I_{\mathrm{HR}}\right)^2 + \left(K_y \otimes I_{\mathrm{HR}}\right)^2 + \varepsilon} \end{aligned} \tag{S33}$$

where K_x and K_y are the horizontal and vertical Sobel kernels, respectively. $\otimes$ represents the convolution operation. Because the phase and intensity images have different physical units and dynamic ranges, the two gradient-magnitude maps were normalized by their spatial mean values $\widetilde{G_\varphi} = G_\varphi / \langle G_\varphi \rangle$, $\widetilde{G_I} = G_I / \langle G_I \rangle$. The high-frequency structural loss was then calculated using a local contrast-structure similarity metric:

$$L_{\text{struc}} = 1 - \left\langle \frac{2\sigma_{\varphi I} + C}{\sigma_{\varphi}^2 + \sigma_I^2 + C} \right\rangle \tag{S34}$$

where σ^2_{φ}, σ^2_I and $\sigma_{\varphi I}$ are the local variances and covariance of the normalized Sobel-gradient magnitude maps G_{φ} and G_I, respectively. The operator $\langle \bullet \rangle$ denotes spatial averaging over the image. We set C = 0.01 and $\varepsilon = 1\times10^{-8}$ for numerical stability. The PGSR network was optimized using the Adam optimizer with a learning rate of 0.001. We adopted an image-specific zero-shot optimization strategy, in which the network was trained independently for each input intensity-phase pair without using HR phase labels. For each input pair, the network was iteratively optimized for 1500 iterations on an NVIDIA GeForce RTX 3090 GPU with 24 GB memory, requiring approximately 104s to reconstruct one HR phase image.

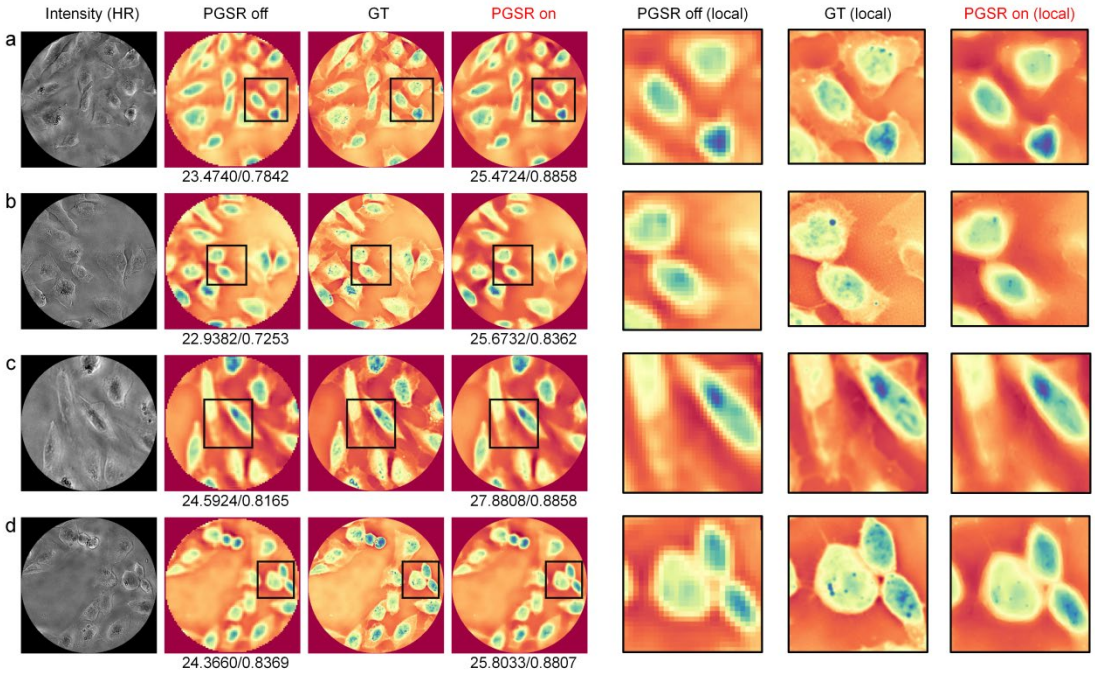


Fig. S9. The performance of the PGSR network on an independent and publicly available FPM dataset[15], with the ground truth (HR intensity, HR phase), LR phase with PGSR off, and SR phase with PGSR on. The value A/B under the phase image represents PSNR/SSIM. The images in black boxes are enlarged on the right three columns.

S6: Calibration and Measurement Process for Raw Interferograms

The obtained interferogram requires calibration to correct for tilt, offset and other anomalies in experiments. This calibration procedure is detailed in Fig. S10 and involves the following steps. First, the circular region of the interferogram is selected, and its radius and center are determined. The area outside the interferogram is then set to zero. Subsequently, a Fourier transform is applied to the interference pattern. The tilt angle θ_r is determined from the positions of the ±1st order peaks in the Fourier domain. The interference pattern is then rotated by θ_r to yield the aligned pattern. Finally, zero-padding is performed on the periphery of the interference pattern, and the resulting data undergoes the reconstruction process described in Fig. 2i.

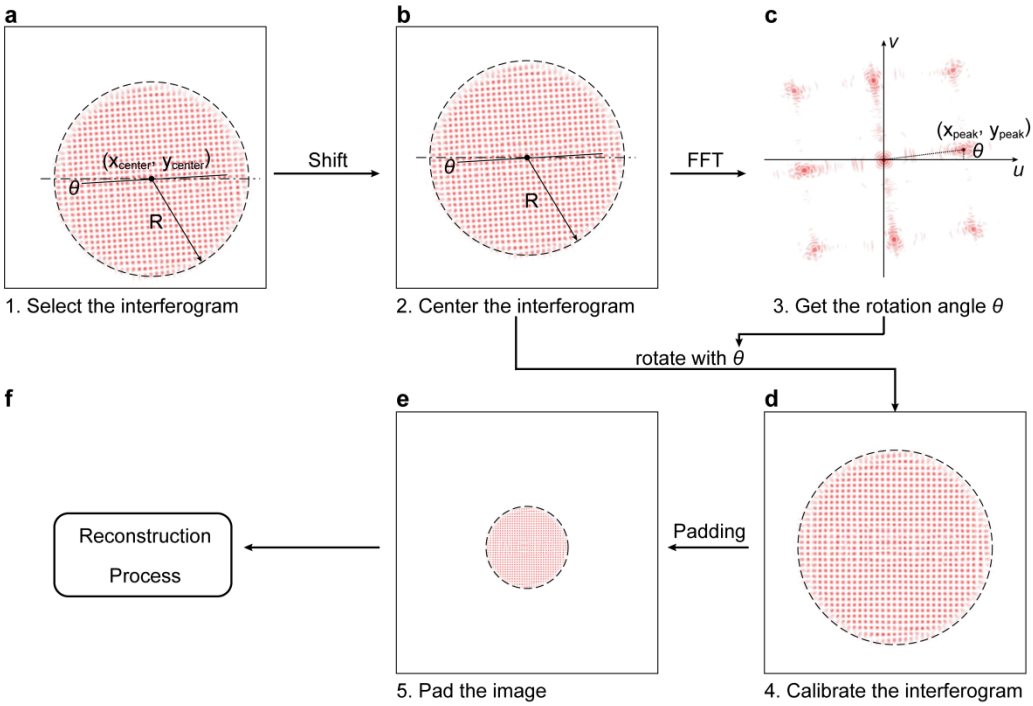


Fig. S10. Schematic calibration process for interferograms. (a) Raw data. (b) Centered interferogram. (c) Fourier spectrum of the interferogram. (d) Rotationally calibrated interferogram. (e) Zero-padded interferogram. (f) Reconstruction process in Fig. 2i.

S7: Experimental Setup for Testing the Resolution Limit of Phase Gradient Using a SLM

To overcome the inherent resolution limit of the spatial light modulator (SLM) in generating fine phase gradients, we implemented the experimental setup illustrated in Fig. S11. Note that high spatial-frequency phase targets with sharp step transitions, such as etched USAF targets, exhibit steep local gradients and are therefore unsuitable for establishing the lower detection limit of phase-gradient measurement. Therefore, we used an SLM to generate the minimum phase variation, followed by spatial filtering to suppress pixelation-induced high-frequency artifacts (Fig. S11). The SLM used in this study (HOLOEYE LETO II) has a pixel size of 6.5 μm and provides phase modulation with 256 discrete levels from 0 to 2π, resulting in a minimum native phase gradient of g_{slm} = (2π/256)/6.5 rad/μm ≈ 3.8×10^{-3} rad/μm. An imaging lens with magnification of M_{mag} = 5 is applied to stretch the phase profile, resulting in a reduced phase gradient to $g_{real} = g_{slm} / M_{mag} \approx 7.5\times10^{-4}$ rad/μm. This enables the generation of a phase gradient finer than the SLM's native hardware limit.

A critical challenge for liquid crystal SLMs stems from the inherent inter-pixel gaps—regions devoid of liquid crystal that occupy approximately 7% of the pixel area and are incapable of phase modulation. These gaps introduce a rapid and erratic phase variation, corresponding to relative high spatial frequencies (denoted as k_{other}), as depicted in Fig. S11b. To eliminate this detrimental component, an iris diaphragm placed immediately behind the lens is employed to control the numerical aperture of the imaging system. By adjusting the size of the iris, the spatial frequency cutoff of the system—which scales approximately as the inverse of the F-number (~1/(Fλ))—is determined. The target phase pattern exhibits a relatively low spatial frequency (k_{aim}), whereas the noise originating from the gaps has a much higher frequency, $k_{other} >> k_{aim}$. Therefore, by properly tuning the iris, the low-frequency signal from the modulated pixels is transmitted, while the high-frequency noise from the gaps is effectively suppressed.

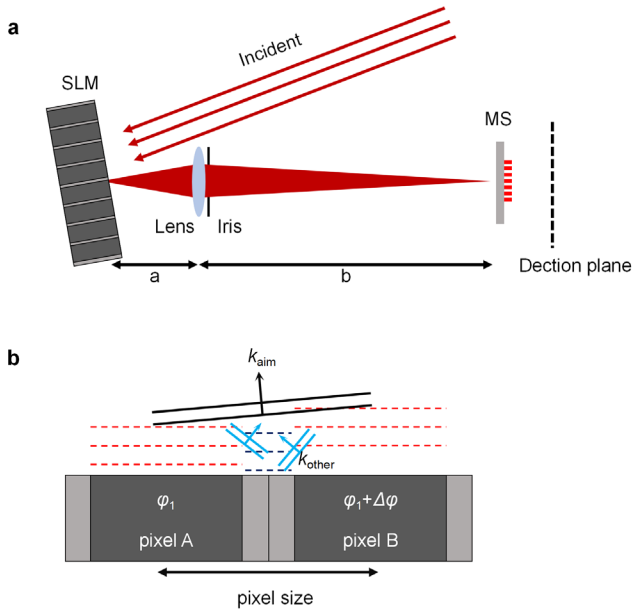


Fig. S11. Experimental setup for testing the phase resolution limit in Fig 3c. (a) The incident light is reflected off the SLM, transmits through a lens and an aperture iris to form a magnified image, and ultimately illuminates the metasurface. (b) The unfilled liquid crystal regions between SLM pixels influence the test; in addition to the programmed phase, the phase contribution from these unfilled regions must be considered.

The entire procedure—comprising both the scaling of the phase profile and the filtering of high-frequency noise—was verified through numerical simulations of wave propagation. The complex optical field modulated by the target phase φ_0 on the SLM was Fourier-transformed, multiplied by a low-pass filter function representing the aperture of the system, and subsequently inverse Fourier-transformed. The argument of the resulting complex field yields the final phase distribution, φ_{final}, presented in Fig. 3c, which confirms the successful generation of a high-fidelity sub-pixel phase gradient.

S8: Reconstruction Results of Interferograms at Different Detection Distances

The phase pattern was applied using the SLM, and the probe distance for the interference pattern was adjusted. Subsequently, the interference pattern was captured and the phase was reconstructed. The results are presented in Fig. S12.

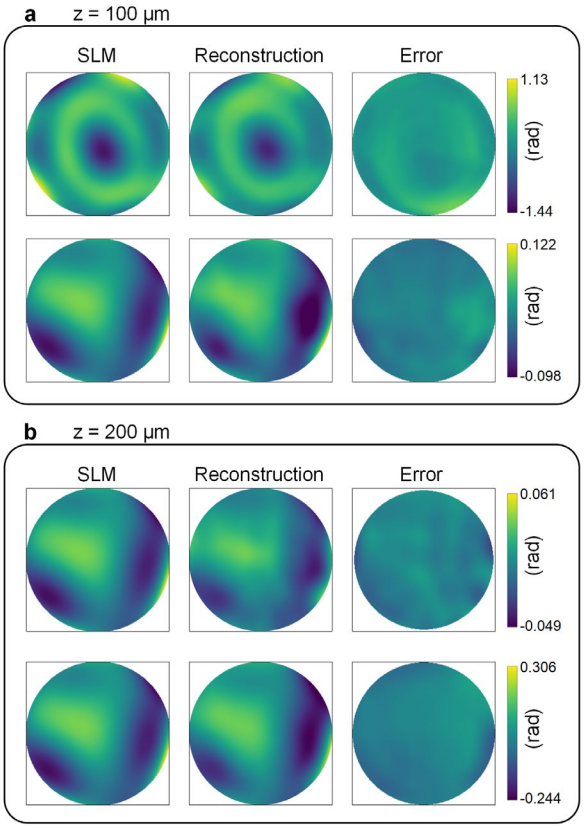


Fig. S12. Phase reconstruction results at different distances. (a) Results for various PV values at z = 100 μm. (b) Results for various PV values at z = 200 μm.

S9: Patterned Phase Reconstruction

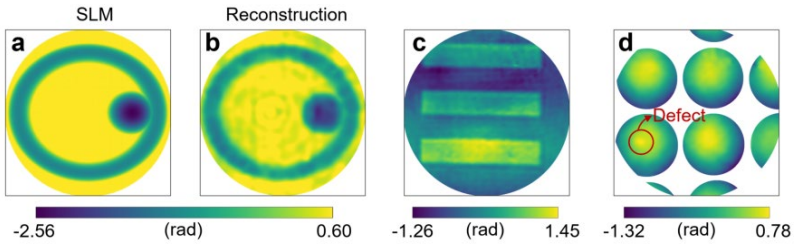


Fig. S13. Results of patterned phase reconstruction. (a, b) the phase loaded on the SLM and the reconstructed phase, respectively. (c) Reconstruction results of the USAF 1951 phase resolution test chart. (d) Reconstruction results for a defective microlens array.

The demonstration also involved testing the other patterned phase distribution. As depicted in Fig. S13a to d, the patterned phase was loaded using a SLM, with reconstruction results presented in Fig. S13a and b. Additionally, the USAF 1951 phase resolution test chart was evaluated, and the results are shown in Fig. S13c. Fig. S13d display the phase distribution reconstruction for the microlens array, representing the intact and defective regions.

S10: Extended Design of Metasurfaces for Phase Reconstruction in Shear Interferometry

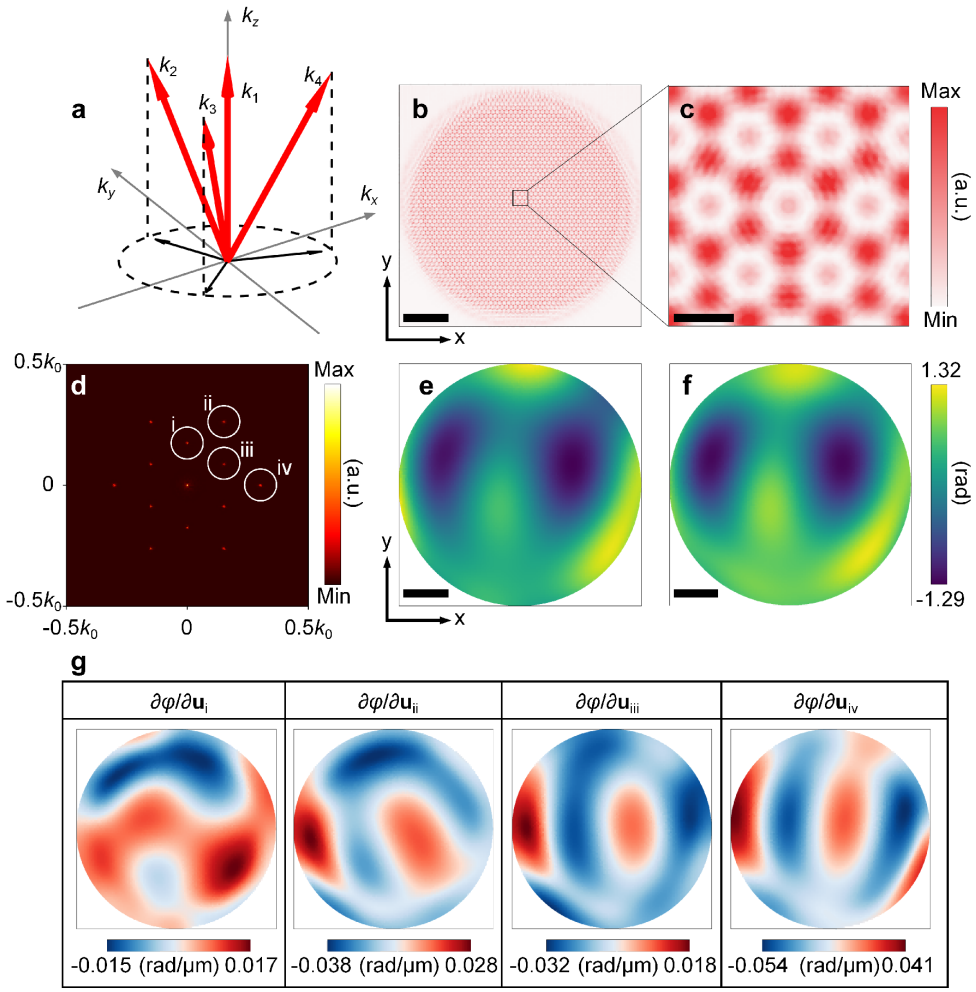


Fig. S14. Schematic illustration of shearing interferometry with enhanced design freedom. (a) Propagation of the outgoing beam in momentum space along directions k_1 to k_4. (b) Simulated interference pattern based on the metasurface design from (a). Scale bar: 100 μm. (c) Magnified view of the boxed region in (b), showing the interference pattern. Scale bar: 10 μm. (d) Spectral distribution of the interference pattern. (e,f) Simulated input phase and reconstructed phase, respectively. (g) Phase gradients in different directions reconstructed from the interference pattern in (d).

Compared to traditional diffractive optical elements, metasurfaces—benefiting from their subwavelength-scale features—offer significantly enhanced efficiency and flexibility in wavefront manipulation. Building upon these advantages, we further explore their potential for advanced interferometric applications, particularly in multi-directional shear interferometry for incident wavefront reconstruction.

As illustrated in Fig. S14a, the designed metasurface splits the transmitted incident light into four beams: one continues along the original optical axis, while the other three are symmetrically deflected at identical angles relative to the axis. The projections of these three tilted beams onto a plane perpendicular to the optical axis are oriented 120° apart from each other. Following the design process outlined in Fig. 1b, we performed numerical simulations, and the resulting interference pattern is presented in Fig. S14b, with a magnified view shown in Fig. S14c. Unlike the interference results discussed in the main text, this configuration yields a hexagonally periodic fringe pattern.

The Fourier transform of the interference pattern (Fig. S14d) reveals that the shear interference

effectively modulates the complex amplitude of the incident wavefront into distinct spatial frequency components. Specifically, the peaks labeled i, ii, iii, and iv correspond to the +1 order diffraction spots at 90°, 60°, 30°, and 0°, respectively. By applying the same processing method as described for Fig. 2i—involving peak shifting, filtering, inverse Fourier transformation, phase argument extraction, and phase unwrapping—we retrieved the phase gradient along each corresponding direction. Representative results are shown in Fig. S14g. This procedure allowed us to acquire phase gradient information along multiple azimuthal directions. Using a Fourier-based gradient integration algorithm, we successfully reconstructed the overall phase profile, as displayed in Fig. S14e and f.

These results demonstrate that metasurfaces not only simplify the design of compact and efficient shearing interferometers but also enable simultaneous acquisition of wavefront gradients along several directions. As supported by previous studies (e.g., Literature[18]), increasing the number of measured gradient directions significantly improves noise suppression and enhances the accuracy of phase reconstruction.

S11: Comparison with traditional QLSI technology.

Compared with conventional grating-based quadriwave lateral shearing interferometry (QLSI), our strategy differs in both physical mechanism and system performance. In conventional QLSI systems, binary cross-gratings with alternating 0 ~ π phase distributions are used to approximate the desired quadriwave splitting. This approximation inevitably introduces unwanted higher diffraction orders, such as the ±3 and ±5 orders, which cause spectral aliasing and require additional mask for spatial filtering. Such filtering reduces the photon budget and can degrade the SNR. In contrast, our metasurface supercell employs subwavelength nanostructures with local electromagnetic resonances to tailor the transmitted amplitude and phase, as shown in Figs. 2a and 2b. This allows optical power to be preferentially directed into the desired diffraction orders while suppressing parasitic higher orders (Figs. S14b and S14c). The propagation profile of our metasurface closely follows the ideal QLSI case, whereas the maskless cross-grating shows parasitic intensity modulations caused by higher-order diffraction (Fig. S15). The metasurface also enables a larger shearing angle and a finer interference period, leading to denser phase-gradient encoding and a much shorter propagation distance (100 – 300 μm in this work, compared to several millimeters for conventional QLSI). Moreover, the metasurface offers greater design flexibility. Although residual crosstalk may exist between the *x*- and *y*-polarization channels due to design overlap and fabrication imperfections, it can be reduced by feedback-based optimization (as discussed in Fig. 2 and Section S2 in Supplement 1). The interference pattern can also be customized to extract phase gradients along multiple directions, as shown in Supplement 1-S10, which introduces additional reconstruction constraints and improves noise robustness.

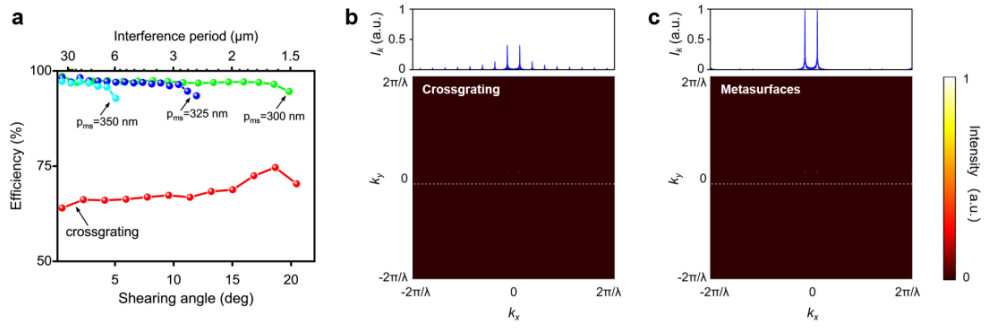


Fig. S15. Comparison of metasurfaces and cross-gratings for QLSI in the *y*-polarization channel. (a) Simulated energy efficiencies of the ±1st diffraction orders for metasurfaces (with various periods p_{ms})

compared with a conventional cross-grating. (b,c) Output energy distributions in the Fourier plane for (b) the cross-grating and (c) the designed metasurface at the same interference period, with top panels showing the 1D cross-sectional intensity profiles.

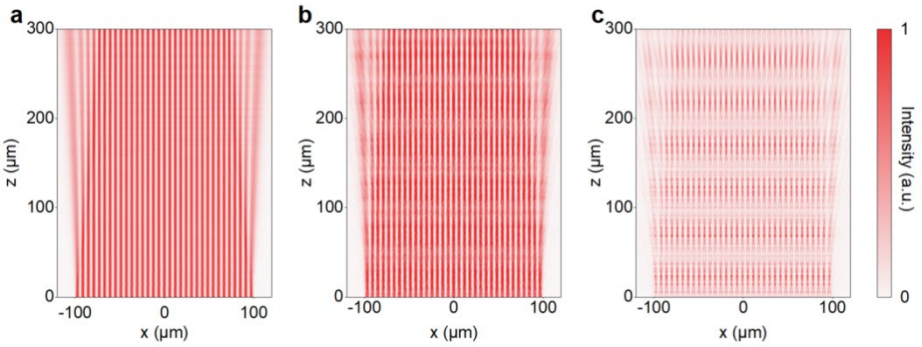


Fig. S16. Simulation of longitudinal intensity distributions for the y-polarized metasurface and the maskless cross-grating. (a) Ideal four-wave shearing interference. (b) The simulated longitudinal intensity distributions of the metasurface we proposed. (c) The simulated longitudinal intensity distributions of a maskless cross grating.

S12: Summary of reported resolution metrics in metasurface-based complex-amplitude imaging

Table S2. Summary of reported resolution metrics for related metasurface-based imaging techniques.

References	λ (nm)	Intensity spatial resolution (λ)	Phase spatial resolution (λ)	Phase gradient resolution (rad/μm)	Phase resolution (rad)
Li et al., Sci. Adv. (2024) [19]	**800**	[a] **5.48**	[a] **6.90**	[a] **0.042**	[a] **0.23**
Kwon et al., Nat. Photonics (2020) [20]	**850**	[a] **4.09**	[a] **3.25**	[a] **0.092**	[a] **0.207**
Li et al., Laser Photonics Rev. (2023) [21]	570	Phase only	[d] 12.11	[c] 0.0019	Not reported
Wu et al., Optica (2023) [22]	532	[a] 1.99	[a] 1.99	Not reported	Not reported
Yi et al., Nat. Commun. (2021) [23]	**532**	[d] **19.55**	[d] **19.55**	[c] **0.11**	**Not reported**
Go et al., Light Sci. Appl. (2024) [24]	**532**	[d] **24.34**	[d] **24.34**	[c] **0.051**	[c] **0.63**
Ji et al., Nat. Commun. (2022) [25]	630	[a] 1.24	[a] 1.24	Not reported	[a] 0.126
Shanker et al., Light Sci. Appl. (2024) [26]	530	[a] 2.08	[a] 2.08	Not reported	[a] 0.63
Wang et al., Light Sci. Appl. (2026) [27]	450	[a] 1.09	[a] 1.09	Not reported	[a] 0.38
Lee et al., Nat. Commun. (2026) [28]	532	[a] 1.64	[a, e] 1.64	Not reported	[a, e] 0.53
This Work	**750**	[a] **1.04**	[a, e] **11.69 (PGSR off)** [a, e] **2.85 (PGSR on)**	[b, e] **$(7.5 \pm 1.6)\times10^{-4}$**	[b, e] **0.049 ± 0.008**

a) resolution-target-based measurement; b) SLM-based calibration; c) theoretical value; d) local wavefront measurement by spatial sampling; e) algorithm-based reconstruction.

To provide a clearer view of the performance metrics reported in related metasurface-based complex-

amplitude imaging studies[19–28], we summarized representative values in Table S2. The listed metrics include operating wavelength, spatial resolution for intensity and phase, phase-gradient resolution, and absolute phase resolution. Superscripts indicate how the resolution-related values were obtained. For our work, the reported phase-gradient resolution corresponds to the minimum phase gradient programmable on the SLM under the experimental configuration. Values not explicitly reported in the cited studies are labeled as "Not reported." These annotations clarify the basis of each reference and provide a comparison among different imaging methods.

S13: Definition of PSNR and SSIM for image reconstruction quality metrics.

Peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) were used to quantitatively evaluate the reconstruction quality[29]. PSNR measures the pixel-wise reconstruction fidelity and is defined as:

$$\mathrm{PSNR} = 10\log\left(\frac{MAX_v^2}{MSE}\right) \tag{S35}$$

where MSE denotes the mean squared error between the ground-truth and reconstructed images, and MAX_v represents the maximum possible pixel value, e.g., 255 for 8-bit images or 1 for normalized images. A higher PSNR indicates a smaller pixel-wise reconstruction error. Therefore, PSNR is useful for quantifying overall numerical fidelity, but it does not fully describe whether the reconstructed image preserves local structural features.

SSIM was used to evaluate structural similarity between the reconstructed and ground-truth images. It is computed as:

$$\mathrm{SSIM}(I_a, I_b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{ab} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)} \tag{S36}$$

Where I_a and I_b represent corresponding local image patches from the original and reconstructed images, C_1 and C_2 are small constants to prevent division by zero, μ_a and μ_b denote the mean intensities of I_a and I_b, σ_a and σ_b are their standard deviations, and σ_{ab} is the covariance between I_a and I_b. SSIM evaluates the similarity between two images based on three components: luminance, contrast, and structure. SSIM ranges from 0 to 1, with a larger value indicating higher structural similarity. Values closer to 1 indicate better preservation of local morphology and structural features.

These two metrics provide quantitative evaluations of image similarity: PSNR reflects the overall pixel-wise reconstruction error, and SSIM characterizes the similarity of local structural patterns and morphological features.

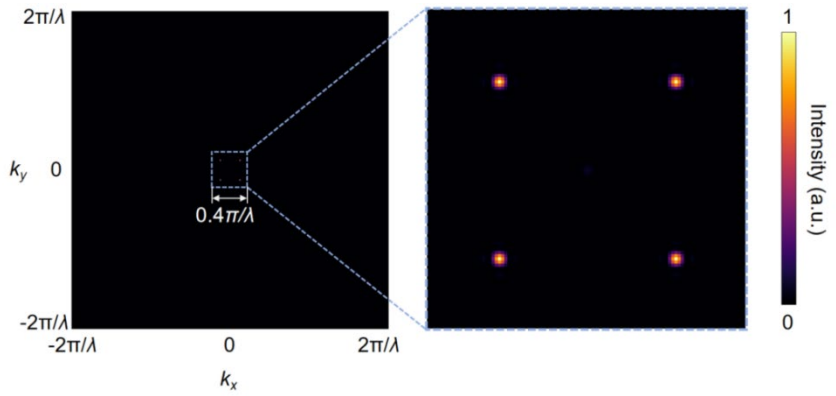


Fig. S17. Diffraction orders distribution of the metasurface under *y*-polarized incidence, simulated by FDTD.

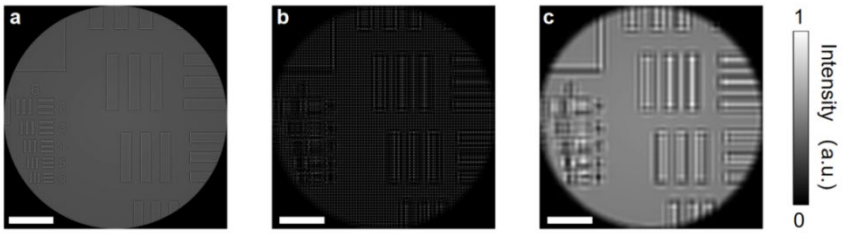


Fig. S18. Simulated comparison of intensity images obtained via direct *x*-polarized imaging and *y*-polarized interferometric reconstruction. (a) HR intensity image under *x*-polarization. (b) Simulated interferogram under *y*-polarization. (c) Intensity image reconstructed from the interferogram. Scale bar: 100 μm.

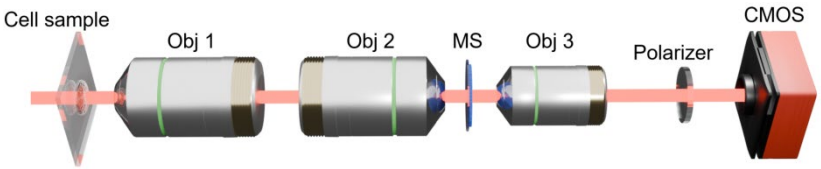


Fig. S19. Schematic of the complex-amplitude meta-microscopy system for label-free live-cell imaging. A paired-objective system projects a conjugate real image of the live HepG-2 cell specimen onto the operating plane of the metasurface (MS), which modulates the incident wavefront for direct intensity imaging and shearing interferometry. The resulting intensity image and interferogram are captured by an sCMOS camera through polarization switching.

Movie S1. Dynamic changes of the refractive-index distribution in the photoresist during UV exposure, corresponding to Figs. 5c-5h.

References:

1. W. H. Southwell, "Wave-front estimation from wave-front slope measurements," J. Opt. Soc. Am. **70**(8), 998–1006 (1980).
2. G. Li, Y. Li, K. Liu, *et al.*, "Improving wavefront reconstruction accuracy by using integration equations with higher-order truncation errors in the Southwell geometry," J. Opt. Soc. Am. A **30**(7), 1448–1459 (2013).
3. S. Ettl, J. Kaminski, M. C. Knauer, *et al.*, "Shape reconstruction from gradient data," Appl. Opt. **47**(12), 2091–2097 (2008).
4. L. Huang, M. Idir, C. Zuo, *et al.*, "Comparison of two-dimensional integration methods for shape reconstruction from gradient data," Opt. Lasers Eng. **64**, 1–11 (2015).
5. H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process. **15**(11), 3440–3451 (2006).
6. B. S. Carlson, "Comparison of modern CCD and CMOS image sensor technologies and systems for low resolution imaging," in *Proceedings of IEEE Sensors 2002* (IEEE, 2002), Vol. 1, pp. 171–176.
7. J. R. Fienup, "Invariant error metrics for image reconstruction," Appl. Opt. 36, 8352‑8357 (1997).
8. M. Nielsen, G. C. Larsen, and K. S. Hansen, "Simulation of inhomogeneous, non-stationary and non-Gaussian turbulent winds," J. Phys. Conf. Ser. **75**(1), 012060 (2007).
9. X. Dong, N. Yokoya, L. Wang, *et al.*, "Learning mutual modulation for self-supervised cross-modal super-resolution," in *Computer Vision – ECCV 2022*, S. Avidan, G. J. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., Vol. 13679 of *Lecture Notes in Computer Science* (Springer, 2022), pp. 1–18.
10. K. Zhang, J. Liang, L. Van Gool, *et al.*, "Designing a practical degradation model for deep blind image super-resolution," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, 2021), pp. 4771–4780.
11. A. Liu, Y. Liu, J. Gu, *et al.*, "Blind image super-resolution: a survey and beyond," IEEE Trans. Pattern Anal. Mach. Intell. **45**(5), 5461–5480 (2023).
12. H. Zhao, O. Gallo, I. Frosio, *et al.*, "Loss functions for image restoration with neural networks," IEEE Trans. Comput. Imaging **3**(1), 47–57 (2017).
13. Y. Tian, D. Krishnan, and P. Isola, "Contrastive multiview coding," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds., Vol. 12356 of *Lecture Notes in Computer Science* (Springer, 2020), pp. 776–794.
14. F. Maes, A. Collignon, D. Vandermeulen, *et al.*, "Multimodality image registration by maximization of mutual information," IEEE Trans. Med. Imaging **16**(2), 187–198 (1997).
15. L. Tian, X. Li, K. Ramchandran, *et al.*, "Multiplexed coded illumination for Fourier ptychography with an LED array microscope," Biomed. Opt. Express **5**(7), 2376–2389 (2014).
16. J. Jing, S. Liu, G. Wang, *et al.*, "Recent advances on image edge detection: a comprehensive review," Neurocomputing **503**, 259–271 (2022).
17. H. Farid and E. P. Simoncelli, "Differentiation of discrete multidimensional signals," IEEE Trans. Image Process. **13**(4), 496–508 (2004).
18. S. Velghe, J. Primot, N. Guérineau, *et al.*, "Wave-front reconstruction from multidirectional phase derivatives generated by multilateral shearing interferometers," Opt. Lett. **30**(3), 245–247 (2005).
19. L. Li, S. Wang, F. Zhao, *et al.*, "Single-shot deterministic complex amplitude imaging with a single-layer metalens," Sci. Adv. **10**(1), eadl0501 (2024).
20. H. Kwon, E. Arbabi, S. M. Kamali, *et al.*, "Single-shot quantitative phase gradient microscopy using a system of multifunctional metasurfaces," Nat. Photonics **14**(2), 109–114 (2020).
21. L. Li, W. Jia, C. Jin, *et al.*, "Single-shot wavefront sensing with nonlocal thin film optical filters," Laser Photonics Rev. **17**(12), 2300426 (2023).
22. Q. Wu, J. Zhou, X. Chen, *et al.*, "Single-shot quantitative amplitude and phase imaging based on a pair of all-dielectric metasurfaces," Optica **10**(5), 619–625 (2023).
23. S. Yi, J. Xiang, M. Zhou, *et al.*, "Angle-based wavefront sensing enabled by the near fields of flat optics," Nat. Commun. **12**(1), 6002 (2021).
24. G.-H. Go, D.-G. Lee, J. Oh, *et al.*, "Meta Shack–Hartmann wavefront sensor with large sampling density and large angular field of view: phase imaging of complex objects," Light Sci. Appl. **13**(1), 187 (2024).
25. A. Ji, J.-H. Song, Q. Li, *et al.*, "Quantitative phase contrast imaging with a nonlocal angle-selective metasurface," Nat. Commun. **13**(1), 7848 (2022).
26. A. Shanker, J. E. Fröch, S. Mukherjee, *et al.*, "Quantitative phase imaging endoscopy with a metalens," Light Sci. Appl. **13**(1), 305 (2024).
27. J. Wang, J. Sun, J. Li, *et al.*, "Plan meta-objective for sub-micron quantitative phase imaging," Light Sci. Appl. **15**(1), 71 (2026).
28. G.-Y. Lee, C. Kim, M. Gopakumar, *et al.*, "Neural phase microscopy with metasurface optics for real-time and nanoscale quantitative phase imaging," Nat. Commun. **17**(1), 1411 (2026).
29. A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th International Conference on Pattern Recognition* (IEEE, 2010), pp. 2366–2369.