

Synthetic aperture waveguide holography for compact mixed-reality displays with large étendue

In the format provided by the
authors and unedited

This PDF file includes:

- Supplementary Note 1: System design and implementation
- Supplementary Note 2: The theory of partial coherence
- Supplementary Note 3: Waveguide models
- Supplementary Note 4: Partially coherent wave propagation models
- Supplementary Note 5: Synthetic aperture holography
- Supplementary Note 6: Additional simulation results
- Supplementary Note 7: Additional captured results
- Supplementary Note 8: Space–bandwidth product, bandwidth, and effective étendue
- Supplementary References

Other supplementary material for this manuscript includes the following:

- Supplementary Video

Supplementary Note 1: System design and implementation

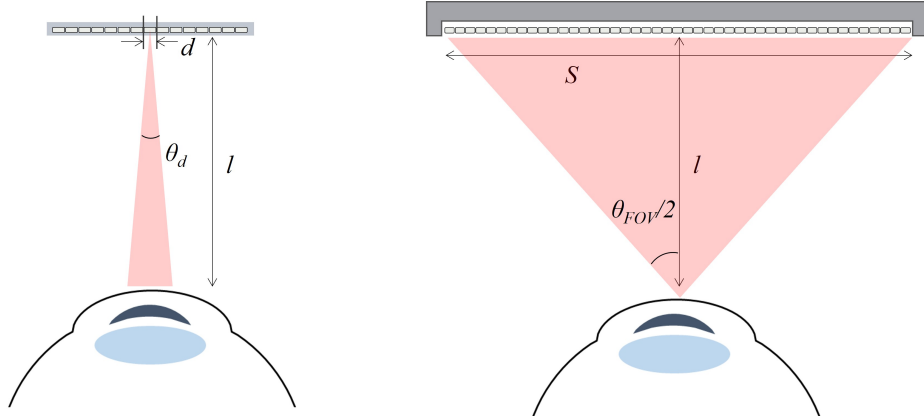
Supplementary Fig. S1 illustrates the intrinsic eyebox E , resolution, and field of view (FOV) of view of the system, without applying the synthetic aperture holography method as:

$$E = a \sin \left(\frac{\lambda}{d} \right) \times l, \quad (\text{S1})$$

$$FOV = 2 \times \text{atan} \left(\frac{S}{2l} \right), \quad (\text{S2})$$

where λ is the wavelength of light source, d is the pixel pitch of the SLM, l is the eye relief which is approximately same as the focal length of the eyepiece lens. For simplicity, the waveguide and lens are omitted in the illustration. The intrinsic eyebox size sets the diffraction-limited resolution of the system, which is determined by the pixel pitch of the SLM as:

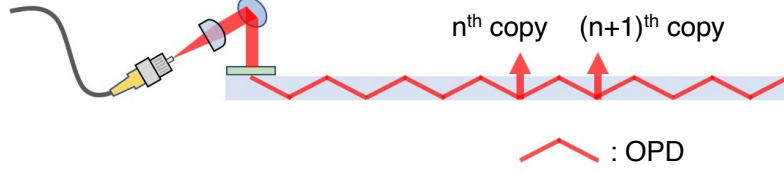
$$\delta\theta_{res} \approx \frac{\lambda}{E}. \quad (\text{S3})$$



Supplementary Fig. S1 Intrinsic eyebox and field of view.

A product of field of view and eyebox size is limited as a bandwidth of spatial light modulator that can be shown by combining Eq. S1 and Eq. S2, which is essentially the number of pixels in the SLM. Widely used SLMs have 2K (Full HD) resolution, which provides only about a 1.8 mm instantaneous eyebox to achieve a 38° degree diagonal FOV.

Light source. Holographic displays require a coherence light source to generate and modulate a wavefront. High coherency of light is important to achieve high resolution close to the diffraction limit, however, this also generates coherent artifacts such as speckle. Partially coherent light sources can alleviate the speckle issue, however this could degrade the maximum



Supplementary Fig. S2 Illustration of optical path length difference between adjacent replicated output beam.

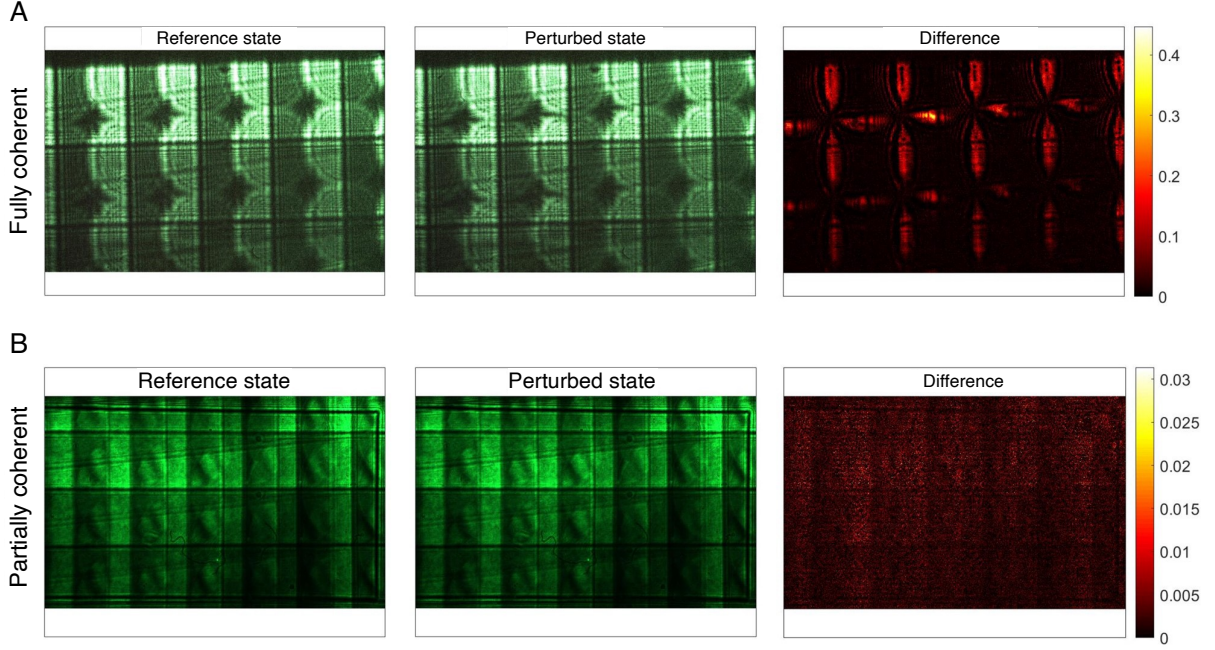
resolution. Therefore the light source should be chosen carefully regarding the architecture of holographic displays. Especially when the laser is combined with pupil-replicating waveguides, there are further unique coherent artifacts that are distinguished from generic free-space propagation cases. One of the major artifacts is the inter-pupil interference effect.¹ When the coherent wavefront is replicated and out-coupled from the waveguide, the out-coupled wavefront would interfere with adjacent replicated copies when the optical path length difference (OPD) is shorter than coherence length of the laser. Assuming that the spectral density follows a normal distribution, the coherence length of the laser and required bandwidth can be written as follows:

$$l_c = 0.44 \frac{\lambda^2}{n\delta\lambda} \quad (\text{S4})$$

$$\delta\lambda > 0.44 \frac{\lambda^2 \cos(\theta_{wg})}{2nt}, \quad (\text{S5})$$

where t is the thickness of the waveguide and θ_{wg} is the total internal reflection (TIR) angle in the waveguide. The inter-pupil interference could result in intensity fluctuation by the vibration of MEMS mirror while operating in a step scanning mode. In practice, the spectral density of a partially coherent source does not have an ideal Gaussian distribution, and thus the requirement may vary from Eq. S4. Therefore, we examined two light sources: A fully coherent diode-pumped solid-state laser with 532 nm center wavelength and 0.1 pm linewidth (Cobolt Samba 1500) and a partially coherent diode laser with 520 nm center wavelength and 0.8 nm bandwidth (Fisba Readybeam). When the fully coherent laser is used, a slight perturbation of input light, in our case, generated by MEMS mirror, caused an artifact at the output illumination as shown in Supplementary Fig. S3A. The fluctuation in illumination intensity varies up to 40% of the reference status. This artifact is likely to be visible by the user and difficult to be compensated since the source has uncontrolled, subtle vibration from the MEMS mirror. When using the partially coherent light source, as shown in Supplementary Fig. S3B, only a slight perturbation is observed as expected. Based on the experimental result, we select the partially coherent light source (FISBA READYBeam) to build the prototype.

Angle-encoded Holographic Waveguide. Our holographic waveguide consists of an in-coupler grating, an exit-pupil expanding grating, and an out-coupler grating, where each coupler is

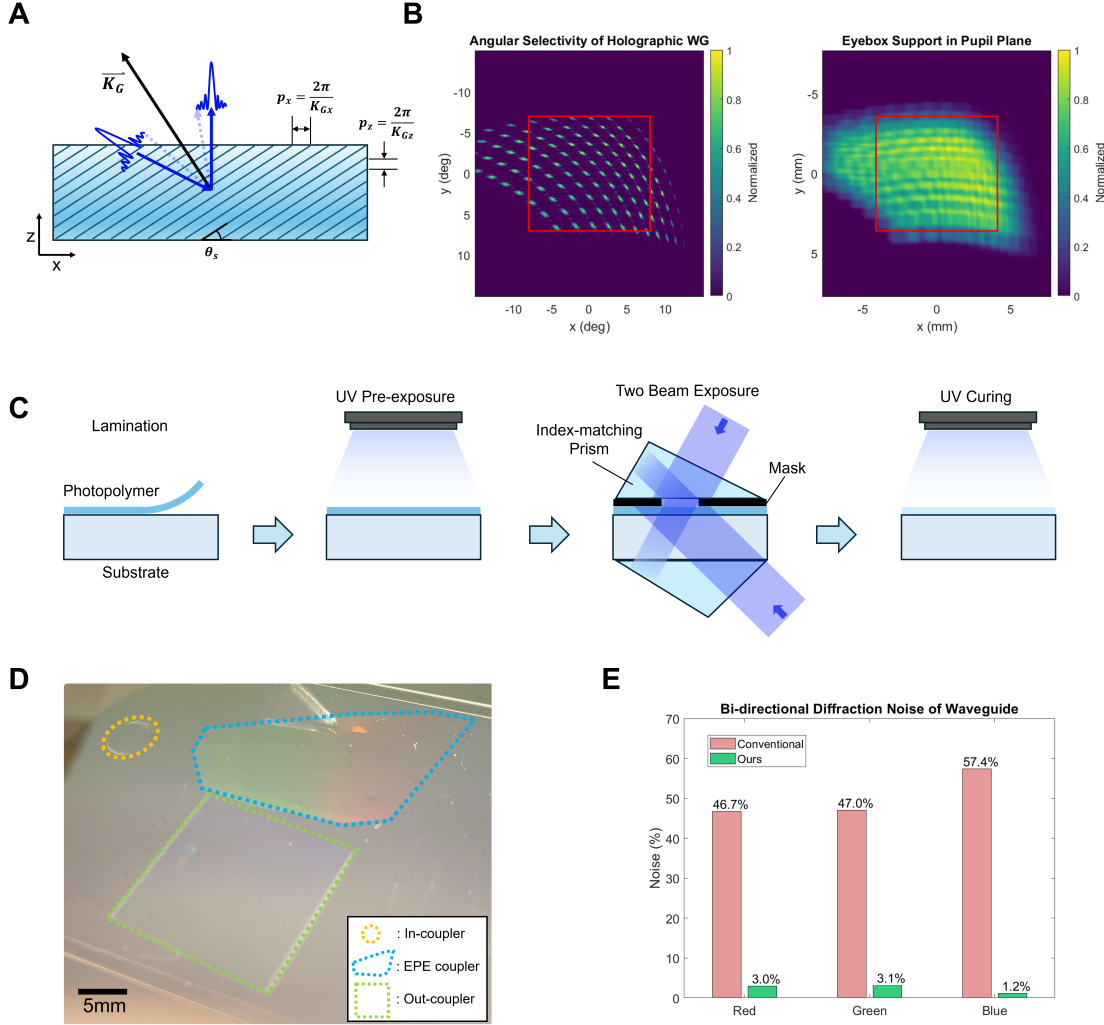


Supplementary Fig. S3 A. Captured images of output illumination intensity from the waveguide using a fully coherent light source as a reference status (*left*), the perturbation of wavefront caused by MEMS mirror vibration captured at the same position and arbitrary time (*middle*), and the visualization of the fluctuation of intensity normalized to the maximum intensity value of the reference status (*right*). The high contrast fringe is clearly observed at the overlapped area of adjacent copies. **B.** Results using a partially coherent light source, demonstrating a very subtle intensity fluctuation.

constructed of a uniquely designed volume Bragg grating (VBG). The total thickness of our waveguide is designed as 0.7 mm including glass substrate and photopolymer film layers. Our waveguide shares the pupil replication principle that surface-relief gratings (SRGs) use,² however, the VBG strategy provides unique benefits in high selectivity and multiplexing capability. Bragg diffraction in a volume grating occurs only for light meeting specific angle and wavelength conditions (known as Bragg condition^{3,4}), as illustrated in Supplementary Fig. S4 A. The Bragg grating can be specified by its grating vector $\vec{K}_G$, where each vector component is inversely proportional to the volume grating period in each direction. When the input beam satisfies the Bragg condition, the VBG can diffract the light into a single diffraction order (i.e., in one intended direction) with high diffraction efficiency.

Off-Bragg light, in contrast, passes through with no diffraction, as the volume grating remains nearly transparent to non-matching angles or wavelengths. This highly selective diffraction property enables light to be out-coupled predominantly into a single side of the waveguide, suppressing the DC noise, in contrast to conventional SRG waveguides which typically show

over 40% of world-side leakage.⁵



Supplementary Fig. S4 A. Visualization of a volume Bragg grating. **B.** Angular selectivity of our angle-encoded holographic waveguide design and its eyebox support in pupil plane selectivity. The red box indicates the target steering angle range of 8 degrees horizontal and 7 degrees vertical, and its corresponding region in the eyebox domain. **C.** Fabrication process of VBG using a two-beam exposure system. **D.** A photograph of our fabricated sample waveguide, with indicated layout of the couplers. **E.** Comparison of bi-directional diffraction noise of a conventional SRG waveguide and ours, experimentally measured for red, green, and blue, respectively.

We employ an angular encoding strategy that takes advantage of the VBG's angular selectivity. Different portions of the target steering angle bandwidth are encoded into each multiplexed

grating, recorded with the same pitch but different slant angles. The pitch was designed to be fixed to ensure a continuously varying optical path for different steering angles as well as to prevent crosstalk between multiplexed gratings. Supplementary Fig.S4 B shows the simulation result of our designed waveguide for the green channel. The interval angle between adjacent gratings' Bragg conditions is set to be smaller than the intrinsic angular bandwidth generated from the SLM; 1.8 degrees in this case, so that it can generate a continuous eyebox. For example, for our green channel design, in-coupler and out-coupler have a surface pitch of 416 nm (indicated as p_x in Supplementary Fig.S4 A), slant angle varying from 26.83 degrees to 30.64 degrees (indicated as θ_s in Supplementary Fig.S4 A), and exit-pupil expansion (EPE) coupler has pitch of 294.2 nm, slant angle varying from 42.89 degrees to 54.18 degrees. The simulation result in Supplementary Fig.S4 B (left) shows that our waveguide collectively covers the target angular steering bandwidth of 8 by 7 degrees, horizontally and vertically respectively. Supplementary Fig.S4 B (right) shows the simulated eyebox support by calculating a convolution of angular selectivity with the intrinsic SLM bandwidth, and translated into the eyebox domain by multiplying with the focal length of the eyepiece lens. We note that Bragg selectivity ensures each color channel, i.e., red (638 nm), green (520 nm), and blue (460 nm), does not interfere with each other within the target steering angle range.

Supplementary Fig.S4 C illustrates the fabrication process of our multiplexed VBG. The process begins by laminating a photopolymer film onto a waveguide substrate (NBK7 glass). To prevent unwanted polymerization prior to exposure, the sample is prepared in a dark room. The recording process comprises two steps: pre-exposure and multiplexed exposures. To eliminate oxygen inhibition—which can quench free-radical polymerization⁶—a low-intensity UV pre-exposure is applied, rendering the film oxygen-depleted and ready for recording. Next, a two-beam interference setup is used to record the VBGs: a reference beam and a signal beam intersect within the film to generate an interference pattern that induces polymerization in high-intensity regions via a reaction-diffusion process, thus forming permanent volume Bragg gratings. We fabricated a mask to selectively expose areas designated for the in-coupler, EPE coupler, and out-coupler and block the rest of the area. Sequential recordings are then performed by realigning the optical setup to record gratings for different wavelengths or angles, resulting in multiple superimposed gratings in the photopolymer volume. The exposure energy for each recording is gradually increased during multiplexing because the photopolymer exhibits a nonlinear response between cumulative exposure energy and polymerization strength. The exposure recipe is iteratively tuned to produce uniform efficiencies among the multiplexed gratings; in our system, typical exposure conditions range from 0.2 mW/cm² to 5 mW/cm² with exposure times from 250 ms to 5 s. Finally, a UV flood cure⁶ is applied after multiplexed exposures are completed, polymerizing any remaining monomers and stabilizing the complete holographic element against environmental changes. After curing, the photopolymer film becomes optically transparent in the visible wavelength regime as shown in Supplementary Fig. S4 D. We demonstrate the noise reduction results of our waveguide in Supplementary Fig. S4 E, compared to a conventional off-the-shelf SRG waveguide (Dispelix). The noise is defined as the

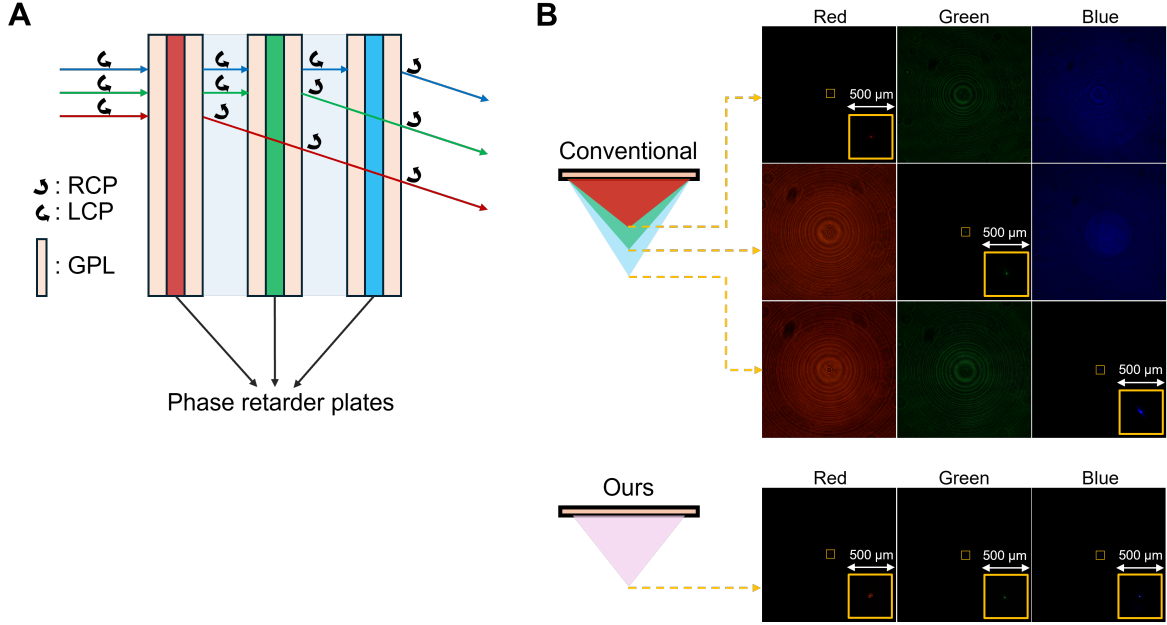
ratio of out-coupled energy diffracted in the unwanted direction to that in the desired direction, measured by a CMOS sensor at the same distance from the out-coupler on both sides.

Apochromatic holographic eyepiece lens The physical principle of our apochromatic holographic lens is based on geometric phase holograms, which is one category of Pancharatnam-Berry phase (PBP) optics. More specifically, it is fabricated with liquid crystal polymer, also known as reactive mesogen,⁷ where molecules align according to boundaries set by the surface and exhibit birefringence to control phase retardation. One can use this to encode and reproduce the desired wavefront for one polarization in principle of polarization holography. If a lens phase profile is encoded, the element is referred to as a geometric phase hologram lens (GPL). It is known that GPLs have severe chromatic dispersion when used to image broadband or polychromatic light, because it provides the same phase modulation regardless of wavelength.

To overcome chromatic dispersion and build apochromatic characteristics, we used 6 GPLs and 3 phase retarder waveplates^{8,9} as illustrated in Supplementary Fig.S5 A. The waveplates manage polarization states to ensure that each layer affects only its intended spectral band while allowing others to pass through unaffected. In our design, each phase retarder is designed to work as a color filter for the red, green and blue color band, so that each pair of GPLs provides optical power to only one color band, and passes the others. In Supplementary Fig.S5 B, we demonstrate the experimental results of apochromatic performance of our holographic lens compared to a conventional diffractive PBP lens. It can be clearly verified that our lens focuses at the same plane regardless of color band, showing no chromatic dispersion, whereas the conventional PBP lens exhibits severe dispersion. To ensure a fair comparison, we matched the numerical aperture sizes of both cases to be identical (0.18), as the conventional diffractive lens used in the experiment has a smaller numerical aperture than ours. Focus length of each R, G, and B is inverse-proportional to its wavelength, where green is 70 mm (ours is 28 mm).

The fabrication process begins with a liquid crystal alignment layer,⁹ where a transparent plastic substrate is coated with a photoalignment material and exposed to polarized UV light, imprinting a spatially varying orientation pattern that determines the lens phase profile. A liquid crystal monomer mixture is then deposited and polymerized under UV light, locking the geometric phase pattern into a solid polymer film. Multiple GPL lens layers are fabricated separately and then stacked with phase retarder waveplates in between. The layers are laminated using index-matched adhesives to minimize reflections and then cured into a unified optical element. The final stacked PBP lens provides chromatic aberration correction with high focusing efficiency across red, green, and blue wavelengths, maintaining uniform optical performance across the full aperture.

Schematic diagram of the prototype. Supplementary Figure S6 shows the overall schematic diagram of the prototype. The SLM patterns (holograms) generated for specific steering angles are sent to the SLM in sync with the MEMS mirror and the laser in color sequential mode. The

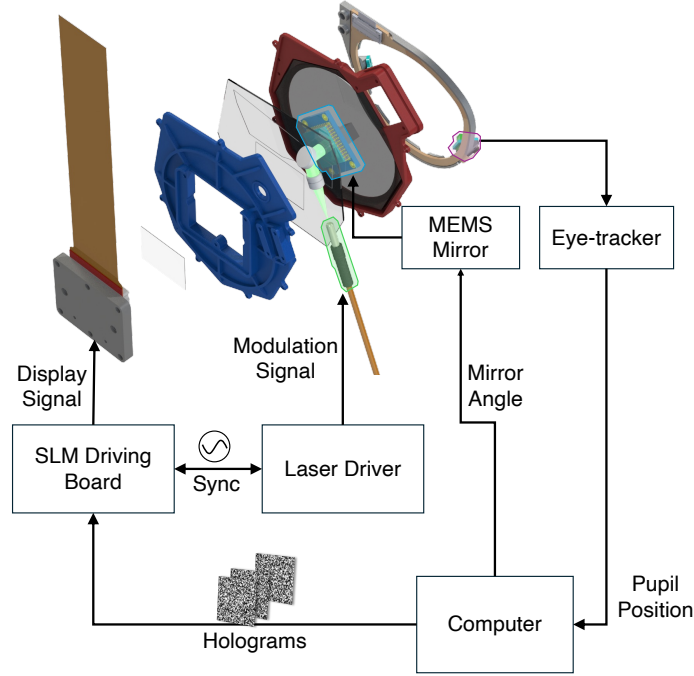


Supplementary Fig. S5 A. Configuration of our apochromatic holographic eyepiece lens. Right circular polarization (RCP), left circular polarization (LCP), geometric phase hologram lens (GPL). **B.** Experimental results demonstrating performance of our apochromatic holographic lens, compared with a conventional diffractive PBP lens. For a conventional PBP lens, the point spread functions of red, green, and blue colors are captured together at each focus plane demonstrating its severe chromatic dispersion, while our lens is captured at the single focus plane.

pupils can be scanned across the entire eyebox or locally around the user's current pupil size and position for the efficient use of optical power and temporal bandwidth with the eye-tracking. In our prototype, the frame rates of each component are as follows: eye-tracker (120Hz), MEMS mirror (400Hz), laser modulation (20KHz), and SLM (60Hz). Due to the limited refresh rate of the SLM used in the prototype, experimental images are captured using a long camera exposure without being synchronized with the eye-tracking camera. However, it is possible to achieve real-time operation with the use of high-frame-rate SLMs that are commercially available.¹⁰

Power and brightness. In our system, the estimated total optical efficiency is calculated as follows. The fiber coupling efficiency is about 50% due to coupling loss, beam collimation efficiency is 29% in average. The MEMS and waveguide combined efficiency is about 5%. Polarization filters and holographic lens result 40% efficiency. In total, the optical efficiency of the system is 0.3%, from the laser light source delivered to the instantaneous pupil. Targeting 1000 nit brightness and D65¹¹ white balance, the R,G,B laser sources should emit 137, 509, 43 lumens per watt, and the required average optical output power for each color is 2.28, 1.27,

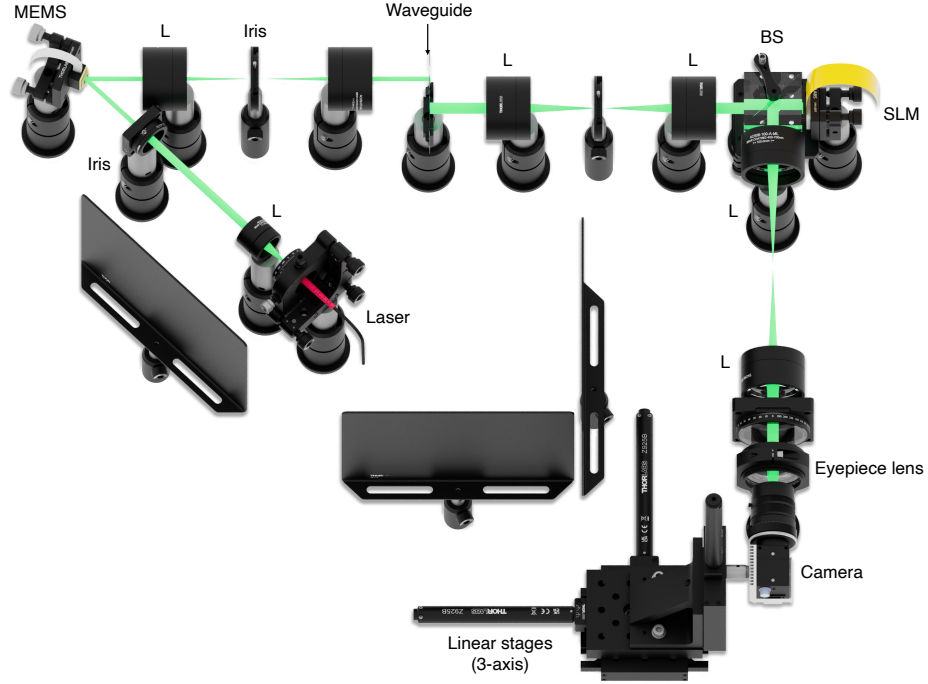
0.74 mW, respectively. Assuming a wall-plug efficiency of 8, 4.8, 5%, each requires 28 mW, 27 mW, 15 mW for R,G,B, resulting in a total battery power of 70 mW. This is an estimated value of our current prototype, and each efficiency has a room for improvement. For example, the waveguide illumination efficiency could be an order of magnitude higher in theory, and wall plug efficiencies can be also higher than our current off-the-shelf devices.



Supplementary Fig. S6 The schematic diagram of our prototype. The SLM patterns (holograms) generated for specific steering angles are sent to the SLM in sync with the the MEMS mirror and the laser in color sequential mode. The pupils can be scanned across the entire eye-box or locally around the user’s current pupil size and position for the efficient use of optical power and temporal bandwidth with the eye-tracking.

Details of experimental setup. We test our algorithmic framework on the thin prototype setup shown in Fig. 1 of the manuscript and the benchtop setup shown in Supplementary Fig. S7. For the benchtop setup, we use a refractive eyepiece lens and a surface relief grating waveguide, and expand the beam path using multiple 4- f systems to ablate the DC term from the surface relief grating waveguide, while maintaining the same geometric parameters as those of the thin prototype setup that uses a holographic waveguide and eyepiece. For both thin and benchtop prototypes, we employ the same off-the-shelf devices: Fisba Readybeam for the laser source, Mirrorcle devkit for the MEMS mirror, and Holoeye PLUTO-2.1-VIS-016 for the SLM. Our results demonstrate that our waveguide and wave propagation model generalize well regardless

of the type of waveguide. We use a FLIR Grasshopper3 GS3-U3-123S6C camera for capturing images.



Supplementary Fig. S7 Benchtop setup configuration. Three $4\text{-}f$ systems are used to relay the MEMS mirror to the in-coupler of the pupil-replicating waveguide, waveguide out-coupler to the SLM, and SLM to the refractive lens plane, respectively. The relayed planes are set to have identical geometric configurations to the thin prototype. Note that the high-order noises of the SLM are not filtered out intentionally, to match the optical characteristics of the thin prototype. L: lens, BS: beam splitter.

Supplementary Note 2: The theory of partial coherence

Partial coherence is the key to achieving high-quality results in large-étendue holographic displays, as we demonstrate in the manuscript, because it unlocks light fields with high-rank mutual intensity, and serves as the foundation for our wave propagation models and CGH rendering. In the following notes, we elaborate on the theory of partial coherence, and how it inspires our waveguide and wave propagation models.

In this note, we first motivate the use of partial coherence for the general image formation model in holographic displays and present it as the generalized version of both incoherent and coherent image formation models. We then discuss the discrete form of the image formation model in more detail, because it is more relevant to our practical implementation and provides more insights into the degrees of freedom. Specifically, we follow the celebrated results of the linear algebraic theory of partial coherence¹² and discuss the theory further in the context of holographic displays, clarifying how the rank of mutual intensity is affected by optical elements designed for étendue expansion. Finally, we empirically validate the relationship between the rank of mutual intensity and image fidelity for light fields using a toy example simulation.

The conventional image formation model for holographic displays is based on the scattering equation for a coherent wave, where each point on a wavefront acts as a source of spherical wavelets, following the Huygens–Fresnel principle.¹³ While this convolutional model generally applies to most imaging and display scenarios, our focus here is on the simplest form of light transport: *diffraction of quasi-monochromatic waves*. Let’s assume a planar wavefront $u_0 = a_0 e^{j\phi_0}$. As it propagates by a distance z , the wavefront evolves as

$$u_z(\xi, \eta, \bar{\lambda}) = h(x, y, \bar{\lambda}) * u_0(x, y, \bar{\lambda}), \quad u_{0,z} : \mathbb{R}^2 \mapsto \mathbb{C}, \quad (\text{S6})$$

where $\bar{\lambda}$ is the center or mean wavelength which we will omit if not necessary, $*$ is the convolution, and h is the Rayleigh–Sommerfeld diffraction kernel.¹⁴ Note that this is equivalent to the angular spectrum method formulation assuming perfect coherence.¹³ In this model, the intensity measured by a sensor or the human retina is

$$|I_z|^2 = |u_z(\xi, \eta)|^2 = |h(x, y) * u_0(x, y)|^2. \quad (\text{S7})$$

Unfortunately, this simple image formation model fails to accurately represent the physical measurements obtained from most setups, and this model mismatch sets a lower bound on the error observed in experimental results. The recent *Neural Holography* framework attempts to address this gap, by parameterizing the potential cause of the mismatch and learning it using data-driven approaches.^{10,15,16} In these approaches, understanding how to effectively model and parameterize the propagation pipeline from the Spatial Light Modulator (SLM) to viewers’ eyes plays a crucial role in achieving optimal performance. This mismatch is addressed using *coherent* parameterized wave propagation models that rely on the ability of convolutional neural networks to learn the arbitrary characteristics of the setup.^{10,15,16}

While these models have shown promising results for many holographic display systems,^{1, 10, 15–18} they become prohibitive when applied to large-étendue holographic display systems. This is because they require thousands of phase–intensity pairs and multiple refinement stages, where new phases are iteratively generated from the trained model and this process can take several days. Even in small-étendue settings without any expansion, this limitation remains significant, and it is exacerbated in a larger étendue setting. The substantial data and time requirements partly stem from the explicit representation and convolutional neural network themselves, which also struggles to generalize to unseen phase–intensity pairs.

In this work, we develop an intuitive yet simple approach for modeling the wave propagation in holographic displays, inspired by the theory of partial coherence.¹⁹ Specifically, we propose a CNN-free solution for scalable modeling of holographic display systems, developing a new wave propagation model that propagates an approximated mutual intensity instead of a single wavefront. Through low-rank factorization of the mutual intensity, the problem reduces to learning just a handful of wavefronts and propagating each of them, a method that offers several advantages, including better generalization to unseen phases, angles, and depths, and a reduction in the number of required training phases and time (See Supplemental Notes 3 and 4).

To begin with, we first note that Equation S7 may not accurately capture the time-varying nature of the wave, which can be represented as $u(x, y; t)$. This wave oscillates at a much higher frequency than the detector integration time ($\bar{\nu} = c/\bar{\lambda} \gg 60\text{Hz}$). Consequently, it becomes necessary to integrate the instantaneous intensity, denoted as $|u_z(\xi, \eta)|^2$, over a specific period of time. As a result, the image intensity can be described as follows

$$|I_z(\xi, \eta)|^2 = \langle |u_z(\xi, \eta)|^2 \rangle_t = \langle |h(x, y) * u_0(x, y)|^2 \rangle_t \quad (\text{S8})$$

$$= \iiint h(\xi_1 - x_1, \eta_1 - y_1) h^*(\xi_2 - x_2, \eta_2 - y_2) \langle u_0(x_1, y_1; t) u_0^*(x_2, y_2; t) \rangle_t dx_1 dy_1 dx_2 dy_2. \quad (\text{S9})$$

$$= \iiint h(\xi_1 - x_1, \eta_1 - y_1) h^*(\xi_2 - x_2, \eta_2 - y_2) J_0(x_1, y_1, x_2, y_2) dx_1 dy_1 dx_2 dy_2. \quad (\text{S10})$$

We assume that the system is stationary, implying that the impulse response does not depend on time.

Mutual Intensity. From Eq. S9, we naturally face the concept of the *mutual intensity*, denoted as J , which serves as a measure of the spatial coherence of light between two object points. (From this point onward, we will omit the subscript denoting zero.)

$$J(x_1, y_1, x_2, y_2) = \langle u(x_1, y_1; t) u^*(x_2, y_2; t) \rangle_t. \quad (\text{S11})$$

Describing a wavefront with this mutual intensity means we are viewing light as a *partially coherent field*, as it gives us the intuitive description of the coherence states, such as

1) Full coherence. If any arbitrary sampling points of the light source is perfectly coherent to each other, integrating the cross-correlation of them is merely the multiplication of time-independent terms;

$$J(x_1, y_1, x_2, y_2) = u(x_1, y_1) u^*(x_2, y_2) \quad (\text{S12})$$

2) Full incoherence. If any distinct sampling point of the light source is perfectly incoherent each other, the cross-correlation of them would vanish as it integrates over time, so

$$J(x_1, y_1, x_2, y_2) = \langle u(x_1, y_1; t) u^*(x_2, y_2; t) \rangle_t = I(x_1, y_1) \delta(x_1 - x_2, y_1 - y_2) \quad (\text{S13})$$

3) Partial coherence. To describe the intermediate state (partial coherence) between these two extremes, the theory of partial coherence^{12,20,21} expresses a partially coherent wavefront u in terms of an orthonormal set of coherent modes, $\{q_1, q_2, \dots, q_m\}$, referred as the coherent mode decomposition^{22,23}

$$J(x_1, x_2, x_2, y_2) = \sum_{m=1}^{\infty} \lambda_m q_m(x_1, y_1) q_m^*(x_2, y_2) = \sum_{m=1}^{\infty} \lambda_m J_{q_m}(x_1, x_2, x_2, y_2). \quad (\text{S14})$$

Here, λ_m are the eigenvalues and q_m are the eigenfunctions of the integral equation

$$\iint J(x_1, y_1, x_2, y_2) q_m(x_2, y_2) dx_2 dy_2 = \lambda_m q_m(x_1, y_1). \quad (\text{S15})$$

Partially coherent wave propagation models. Note that the mutual intensity of a quasi-monochromatic wave, regardless of its partial coherence state, can be expressed as a combination of mutual intensity functions associated with a set of fully coherent wavefronts. With this decomposition, we can derive a general image formation model using the coherent mode decomposition, following Goodman¹⁴ (Eqs. S16 – S19). Both full incoherence and full coherence can be encompassed within this model, representing the two extreme cases.

Substituting Eq. S14 into Eq. S10, we obtain

$$|I_z(\xi, \eta)|^2 = \iiint h(\xi_1 - x_1, \eta_1 - y_1) h^*(\xi_2 - x_2, \eta_2 - y_2) \sum_{m=1}^{\infty} \lambda_m q_m(x_1, y_1) q_m^*(x_2, y_2) dx_1 dy_1 dx_2 dy_2, \quad (\text{S16})$$

$$= \sum_{m=1}^{\infty} \lambda_m \iiint h(\xi_1 - x_1, \eta_1 - y_1) h^*(\xi_2 - x_2, \eta_2 - y_2) q_m(x_1, y_1) q_m^*(x_2, y_2) dx_1 dy_1 dx_2 dy_2, \quad (\text{S17})$$

$$= \sum_{m=1}^{\infty} \lambda_m \iint h(\xi_1 - x_1, \eta_1 - y_1) q_m(x_1, y_1) dx_1 dy_1 \left(\iint h(\xi_2 - x_2, \eta_2 - y_2) q_m(x_2, y_2) dx_2 dy_2 \right)^*, \quad (\text{S18})$$

$$= \sum_{m=1}^{\infty} \lambda_m |h(x, y) * q_m(x, y)|^2, \quad (\text{S19})$$

Without the loss of generality, we can assume $\lambda_0 > \dots > \lambda_n > \lambda_{n+1} > \dots > \lambda_{\text{inf}}$. Again, this equation can be applied to two extreme cases which agree Eq. S12 or Eq. S13.

1) Full coherence. Only single mode exists, which means λ_0 is only nonzero:

$$|I_z(\xi, \eta)|^2 = \lambda_1 |h(x, y) * q_1(x, y)|^2 = |h(x, y) * u(x, y)|^2, \quad (\text{S20})$$

which simplifies to the conventional image formation model in Eq. S7.

2) Full incoherence. In this scenario, as the q 's form an orthonormal basis, each q_m represents a single pixel with a delta function at that location. To indicate the activation of a mode, we will use a two-dimensional index, m and n . Consequently, the integral equation Eq. S15 can be expressed as follows:

$$q_{m,n}(x, y) = \delta(x - m, y - n), \quad (\text{S21})$$

$$\iint J(x_1, y_1, x_2, y_2) q_{m,n}(x_2, y_2) dx_2 dy_2 = \lambda_{m,n} q_{m,n}(x_1, y_1), \quad (\text{S22})$$

$$\iint J(x_1, y_1, x_2, y_2) \delta(x_2 - m, y_2 - n) dx_2 dy_2 = \lambda_{m,n} \delta(x_1 - m, y_1 - n), \quad (\text{S23})$$

$$J(m, n, m, n) = \lambda_{m,n} \leftrightarrow \lambda_{m,n} = I(m, n) = \langle |u(m, n)|^2 \rangle \quad (\text{S24})$$

Again,

$$|I_z(\xi, \eta)|^2 = \sum_{m=1}^M \sum_{n=1}^N \lambda_{m,n} |h(x, y) * q_{m,n}(x, y)|^2 = \sum_{m=1}^M \sum_{n=1}^N \lambda_{m,n} |h(x, y) * \delta(x - m, y - n)|^2 \quad (\text{S25})$$

$$= \sum_{m=1}^M \sum_{n=1}^N I(m, n) |h(x - m, y - n)|^2, \quad (\text{S26})$$

$$= |h(x, y)|^2 * \langle |u(x, y)|^2 \rangle \quad (\text{S27})$$

Thus, at this point, we arrive at the fundamental results of coherent and incoherent light transport: in an incoherent display system, the linearity holds in terms of intensity; the impulse response of such a system corresponds to the squared magnitude of the amplitude impulse response.¹⁴

Discrete fields and effective degrees of freedom

The above continuous equations can also be established within the simple framework of linear algebra. In this section, we will focus on the case of discrete wave fields, as it is the typical way we handle them. Additionally, we will explore the core insight that captures the essential property of coherence. It is also relevant to the discussion of degrees of freedom, which are often constrained by a discrete number of pixels, whether they are in displays or sensors. Hence, working within the framework of discrete representations offers an elegant and less redundant approach to dealing with partial coherence.

We can span any 2D band-limited wavefront into a 1D vector - $[f(1, 1), f(1, 2), \dots, f(m, n)]$ as $[f(1), f(2), \dots, f(mn)]$, so effectively we restrict our discussion to the 1D case, while the extension to higher dimensions is straightforward. In this 1D case, the mutual intensity can be represented as a two-dimensional matrix as $\mathbf{J} \in \mathbb{C}^{N \times N}$, where $N = mn$ is the number of controllable SLM pixels that we will assume in this subsection (i.e. number of degrees of freedom).

Given this matrix representation, the coherence mode decomposition of an arbitrary mutual intensity matrix J can be described by a singular value decomposition:

$$\mathbf{J} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H, \quad (\text{S28})$$

where $\mathbf{\Lambda}$ is a diagonal matrix consisting of the singular values λ_i and $\mathbf{Q} \in \mathbb{C}^{N \times N}$ is an orthonormal matrix with columns q_i :

$$\mathbf{Q} = (q_1, q_2, \dots, q_N), \quad (\text{S29})$$

which is equivalent to Eq. S14.

Note that Eq. S28 represents rank-N construction of the matrix J . The matrix representation of the mutual intensity provides a convenient way to assess the level of coherence in a partially coherent field. A fully coherent field, characterized by a single mode, can be represented as a rank-one matrix. Conversely, a fully incoherent field has no nonzero values off the diagonal, having rank N .

Linear operators on mutual intensity. The discrete representation offers several useful insights. For example, we can obtain the intensity as the trace of the mutual intensity matrix, $I = \text{tr}(\mathbf{J})$, or as the sum of the eigenvalues. In this subsection, we will examine various optical operations from the perspective of linear algebra in detail. Specifically, we will explore the role of several operations commonly used in holographic display systems, including wave propagation, phase mask application, time-multiplexing, beam steering, pupil replication, and lenslet arrays. We analyze how these operations can impact the rank of the wavefront, which essentially determines the independent degrees of freedom required to represent the spatial-angular space. *The key insight we gain here is that certain operations do not increase the rank, thus preserving the effective degrees of freedom needed for faithfully reconstructing scenes.*

1) Linear Propagation of Mutual Intensity. The linear propagation model of the coherent wavefront can be described as a matrix-vector multiplication form;

$$\mathbf{f}_o = \mathbf{H}\mathbf{f}_i, \quad (\text{S30})$$

where each row of $\mathbf{H}$ represents the point spread function (PSF) or propagation kernel and $\mathbf{f}_i$ is the input field and $\mathbf{f}_o$ is the output field. For the input and output mutual intensity, $\mathbf{J}_i$ and $\mathbf{J}_o$ described in Eqs. S28 and S29, the propagation is

$$\mathbf{J}_o = \mathbf{H}\mathbf{J}_i\mathbf{H}^H \quad (\text{S31})$$

$$= \mathbf{H}\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H\mathbf{H}^H \quad (\text{S32})$$

$$= \mathbf{H}\mathbf{Q}\mathbf{\Lambda}(\mathbf{H}\mathbf{Q})^H \quad (\text{S33})$$

$$= \sum_i^N \lambda_i (\mathbf{H}\mathbf{Q})_i (\mathbf{H}\mathbf{Q})_i^H, \quad (\text{S34})$$

As in the continuous description, the output mutual intensity is the linear sum of the mutual intensities of all propagated modes, weighted by their eigenvalues. Note that the propagation does not change the rank since the number of modes is conserved.

2) Time multiplexing on mutual intensity. Based on the definition in Eq. S11, it is evident that the mutual intensity represents the average of individual mutual intensities at each time slot. This can be represented by stacking each wavefield in columns, as

$$\mathbf{F} = (\mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \quad \mathbf{f}_n), \quad (\text{S35})$$

And multiply it with its transpose. The rank increases with the number of distinct columns, so with n frames, the upper bound of the rank is set to n , and the partially coherent wavefronts can approximate the light field using a rank- n approximation. However, this does not hold true if the same fields are used repeatedly over time.

$$\mathbf{J}_F = \mathbf{F}\mathbf{F}^H \quad (\text{S36})$$

3) Oversampling in mutual intensity. We note that oversampling (repetition) of any optical field does not increase the rank of its mutual intensity.

$$\mathbf{f}_O = \mathbf{O}\mathbf{f}. \quad (\text{S37})$$

Here $\mathbf{O}$ is the oversampling matrix, which duplicates each entry k times. For example, when we sample it twice we use

$$\mathbf{O}_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (\text{S38})$$

As oversampling does not increase the effective degree of freedom, so does not increase the rank of mutual intensity. It can be formally derived from the rank inequalities:

$$\mathbf{J}_{f_O} = \mathbf{O}\mathbf{f}\mathbf{f}^H\mathbf{O}^H = \mathbf{O}\mathbf{J}_f\mathbf{O}^H. \quad (\text{S39})$$

$$\text{rank}(\mathbf{J}_{f_O}) = \min(\text{rank}(\mathbf{O}), \text{rank}(\mathbf{J}_f)) (\because \text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))) \quad (\text{S40})$$

$$= \text{rank}(\mathbf{J}_f), \quad (\text{S41})$$

since $\text{rank}(\mathbf{O}) = N$.

4) The role of a mask (element-wise multiplication) in mutual intensity. Some previous works²⁴ apply a phase mask, often with smaller feature sizes, to the system to expand the étendue. The coherent field is multiplied element wise after oversampling. This element-wise multiplication is equivalent to multiplying by a diagonal matrix, with the mask values on its diagonal. *Note that the following inequality holds even when the pitch of the mask is smaller than the pixel pitch of the wavefront. Applying a mask does not increase the effective degrees of freedom of the system.*

$$\mathbf{J}_{f_O \cdot \mathbf{m}} = (\text{diag}(\mathbf{m})\mathbf{f}_O)(\text{diag}(\mathbf{m})\mathbf{f}_O)^H = \quad (\text{S42})$$

$$= \text{diag}(\mathbf{m})\mathbf{J}_{f_O}\text{diag}(\mathbf{m})^H = (\mathbf{m}\mathbf{m}^H) \odot \mathbf{J}_{f_O}. \quad (\text{S43})$$

$$\text{rank}(\mathbf{J}_{f \cdot \mathbf{m}}) \leq \text{rank}(\mathbf{m}\mathbf{m}^H)\text{rank}(\mathbf{J}_{f_O}) \quad (\because \text{rank}(A \odot B) \leq \text{rank}(A)\text{rank}(B)) \quad (\text{S44})$$

$$= \text{rank}(\mathbf{J}_{f_O}) \quad (\text{S45})$$

$$= \text{rank}(\mathbf{J}_f) \quad (\text{S46})$$

5) The role of pupil replication in mutual intensity. The role of pupil replication can be analyzed in a similar manner to the previous discussion on oversampling. In this case, pupil replication involves duplicating the entry with a certain shift (perhaps with subsequent changes following propagation). We can formulate the matrix that carries out this operation, denoted as R ; $\mathbf{f}_r = R\mathbf{f}_i$. The same rank analysis is applicable here; however, since the rank of R is bounded by N , it also maintains the existing rank without increasing it.

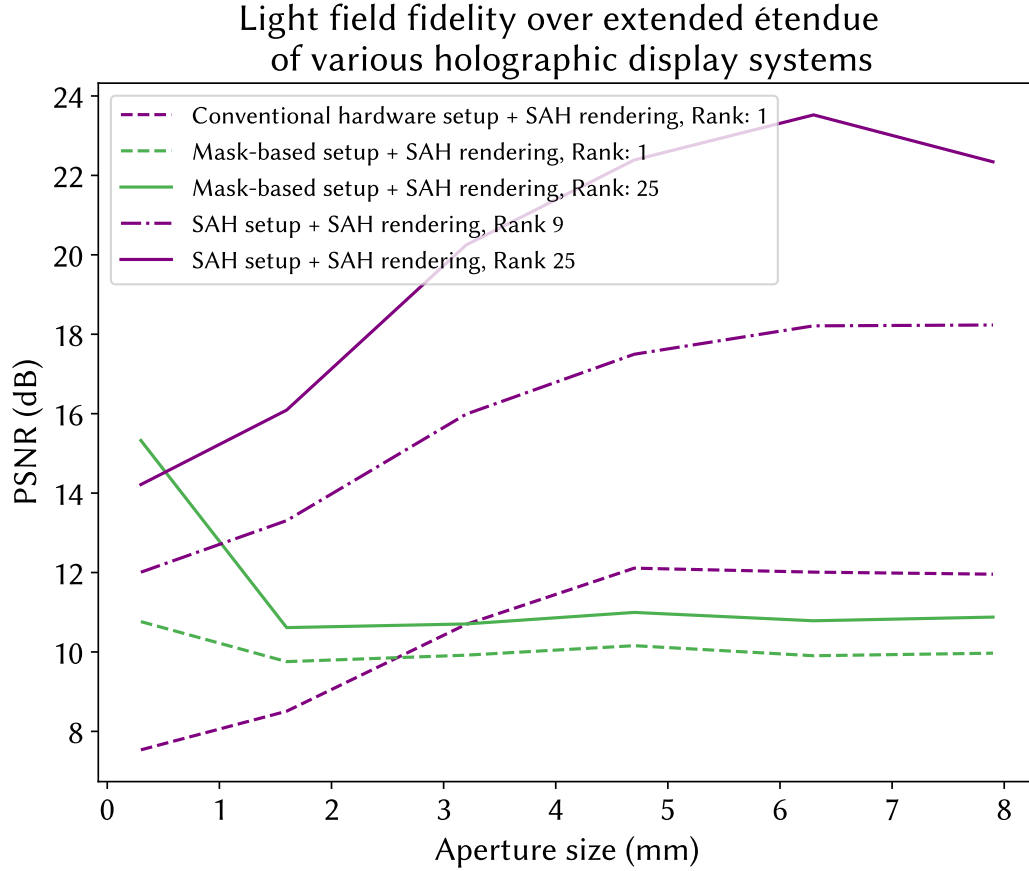
6) The role of backlight steering in mutual intensity. Backlight steering could be modeled as multiplying a phase ramp $e^{j\mathbf{k}\cdot\mathbf{r}}$. In discrete cases, if the angle exceeds the bandwidth with N pixels support, it cannot fully express the steering due to aliasing. One can dramatically oversample each pixel to support that steep angle. However, using the Fourier property, we can merely shift (migrate) the whole spectrum in Fourier domain, with the shift amount of k , and express this operation in a system matrix S . Note that, solely with the steering, the number of non-zero entries of $\mathbf{f}_s$ is still N , while it could access the wider range of the Fourier spectrum in k space.

$$\mathbf{f}_s = \mathbf{F}^{-1}\mathbf{S}\mathbf{F}\mathbf{f} \in \mathbb{R}^{(n+k)\times n} \quad (\text{S47})$$

However, with time multiplexing, where we stack the field profile in column dimension (See Eq. S35), we can determine each shift k_i to not overlap in rows, where we get the number of rows as $\max(k_i)$ for $t = 1, \dots, T$ where T is the number of frames. In this case, while the number of rows of $\mathbf{F}_s = (\mathbf{f}_1 \ \mathbf{f}_2 \ \dots \ \mathbf{f}_T)$ could be arbitrary within the limit of the steering system, the rank, or the nonzero rows, is limited up to nT .

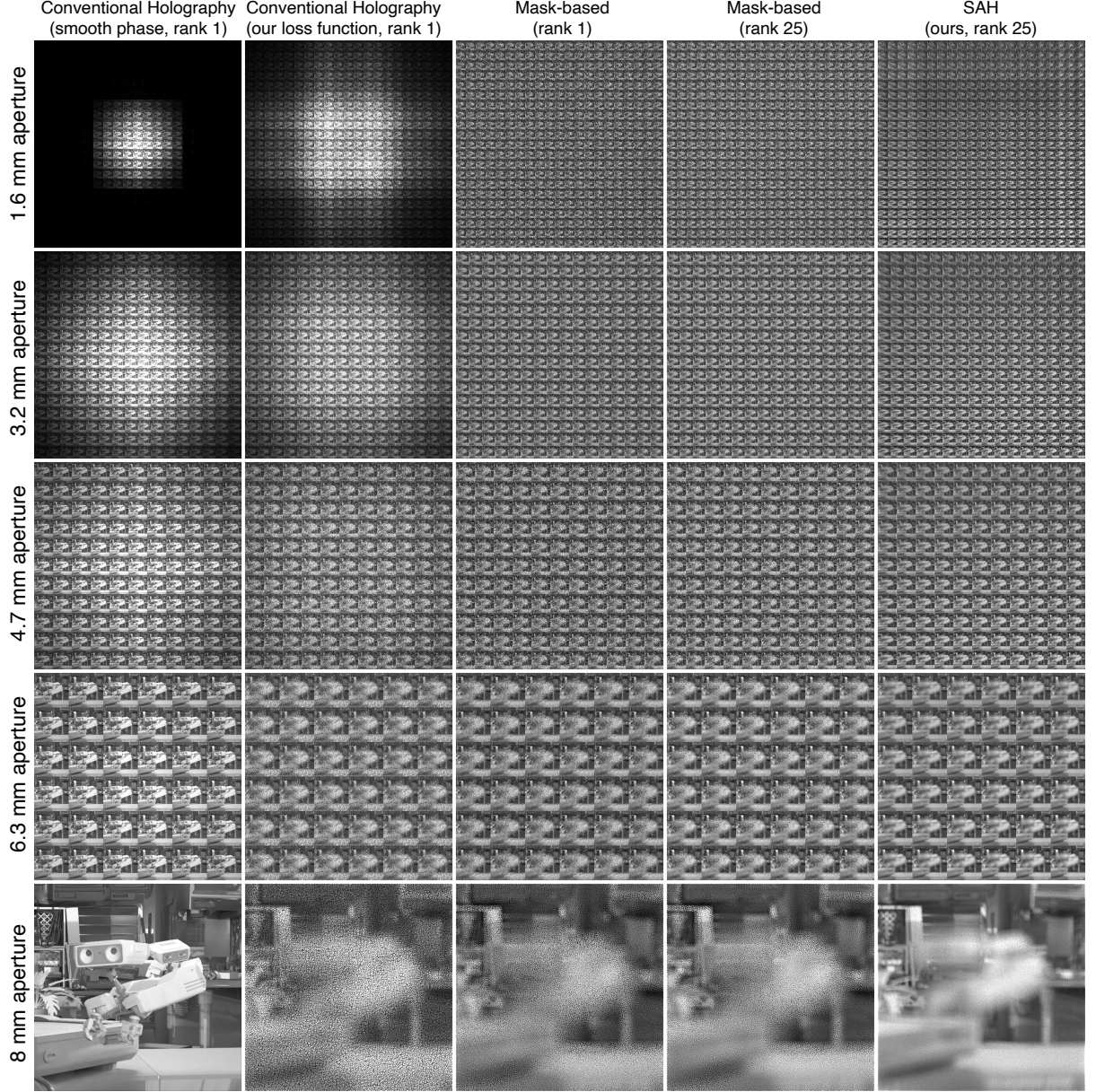
Validation

In this section, we empirically test the relationship between the rank and the effective degrees of freedom discussed in the previous section by supervising the wavefront through example operations, such as mask-based étendue expansion and synthetic aperture generation (steering + time-multiplexing), for producing a target light field of 25×25 views, which covers an 8×8 mm² eyebox. In this setting, it corresponds to an eyebox that is 5×5 times larger than what an SLM supports, and we simulate the mask right in front of the SLM with a pixel pitch that is five times smaller for mask-based étendue expansion. In practice, we encounter higher-order copies, and we must account for these higher orders.²⁵ For all methods, we simulate 5×5 higher orders and jointly optimize the target light fields.²⁵ We use the same pixel pitch and green channel from our hardware setup but with a smaller resolution of 600×600 to fit the memory requirements for the mask-based étendue expansion (Note that our method is scalable and not constrained by memory). To this end, we optimize complex-valued wavefronts using the same loss function (Eq. 5 in the main manuscript, see Supplementary Note 5), and examine the light field fidelity for various aperture sizes (See Supplementary Note 8).



Supplementary Fig. S8 Comparison of the effective degrees of freedoms of various holographic display setups using the average PSNR (in dB) of the reconstructed light fields in intensity. We simulate and compare the light field to the target, across the extended étendue of $8\text{mm} \times 8\text{mm}$, for all aperture sizes.

Supplementary Figure S8 shows the average PSNR (in dB) of the reconstructed light fields in intensity of various holographic display setups across the extended étendue of $8\text{mm} \times 8\text{mm}$, using all aperture sizes. We observe an interesting trend here: systems with rank-1 operations, even with the étendue expansion, do not significantly improve the performance in generating light fields, regardless of aperture size. Specifically, the mask-based expansion allows access to a larger étendue, thereby improving performance for smaller aperture sizes, but it significantly reduces contrast and lowers performance for larger apertures. As expected, only high-rank operations enable a partially coherent wavefront to generate high-fidelity light-field reconstructions. We also simulate the same high rank for mask-based expansion (*green solid line*), which only marginally improves performance, indicating that the increased bandwidth does not fully translate to *effective degrees of freedom* (See Supplementary Note 8) in this system. Notably,



Supplementary Fig. S9 Qualitative comparison of the effective degrees of freedoms of various holographic display setups using the reconstructed light fields. We simulate and compare the light field to the target, across the extended étendue of $8\text{mm} \times 8\text{mm}$, for all aperture sizes. The interval between each viewpoint is fixed to that of target light field views, i.e. $\frac{\lambda}{5p}$.

this trend has been observed in multiple studies,^{23,26} but we are the first to validate it in a large-étendue, high-resolution setting.

Supplementary Figure S9 presents the reconstructed light fields for various holographic

display setups. Each row corresponds to a light field generated with a different aperture size, ranging from 1.6 mm to 8 mm. The first column represents the smooth phase method using a central view image, which concentrates all energy at the center of the eyebox. Consequently, the light field visibility is limited by the aperture size, like a pinhole camera scenario. In this setup, the light field can be seen from all viewpoints if an aperture size larger than approximately 4 mm is used. The second column shows the same hardware setup but with our rendering approach. Since the system generates higher orders, it typically provides optimization access to a larger étendue than what the SLM bandwidth allows. By simulating higher orders, the optimization redistributes the energy to fit the light fields for all aperture sizes (driven by our loss function, see Supplementary Note 5), thus enabling a larger effective étendue, even with the same hardware. The mask-based expansion inherently accesses a larger étendue, but it suffers from low contrast in all cases. The last column shows our system, which cover the entire effective étendue with the highest image fidelity, demonstrating the larger effective degrees of freedom. We discuss this further in Supplementary Note 8. Note that our setup is straightforward to extend this to other color channels, while the mask-based method should jointly account for this, which would result in a worse performance.

Supplementary Note 3: Waveguide models

In this note, we present several novel waveguide models designed to mitigate aberrations and interference artifacts in our system. The waveguide generates the incident beam that is modulated by the SLM, but is prone to significant aberrations and interference artifacts due to the complex wave propagation inside the waveguide. As a result, accurately modeling waveguide propagation has been a long-standing challenge in this field. Recent studies have demonstrated modeling the complex waveguide propagation of coherent light¹ using a interferometric calibration system. However, this approach does not account for partial coherence, a critical aspect of our system when using partially coherent light that arises from optical path length differences between output beams. To our knowledge, there is currently no established method for modeling partially coherent wave propagation within waveguides. A straightforward approach might be to learn each separated beam based on its angle. However, this method is not scalable for large étendue and partially coherent settings, and it does not generalize well to unseen angles.

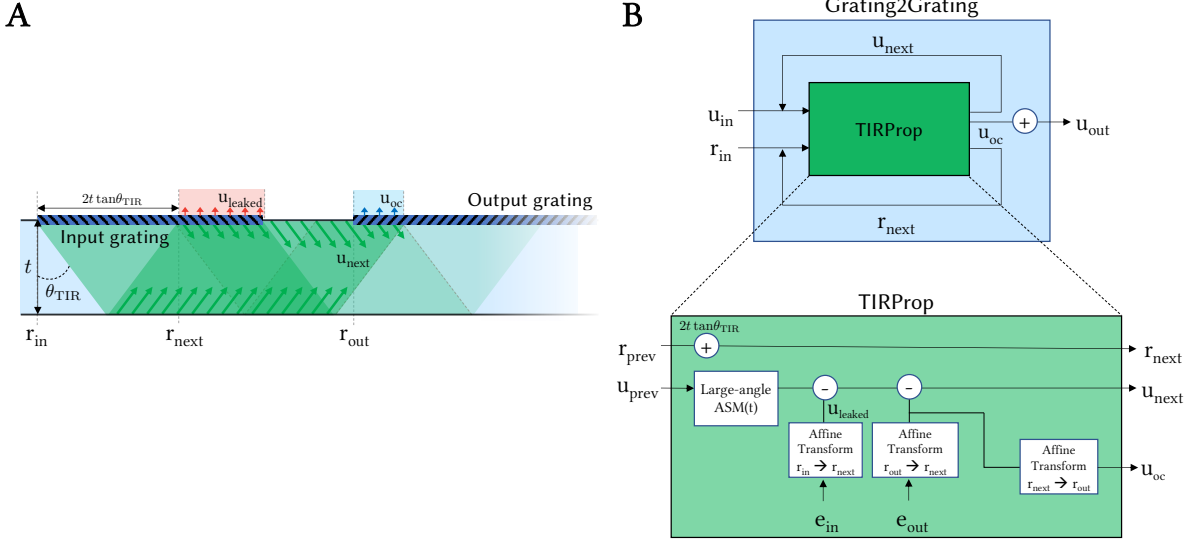
To address this, we first develop a fully physics-based waveguide model which inspires our neural implicit waveguide model. In this note, we demonstrate advantages of this implicit neural representation of wavefronts, in terms of fast convergence, while revealing the full potential in the next Supplementary Note: when combined with a mutual-intensity-based wave propagation model, it generalizes well to unseen depths, angles, and phases, while requiring an order of magnitude less data.

Explicit representations. The most straightforward approach to representing waveguides is to explicitly store the output wavefront at all incident angles using a tensor as below. Each element in the tensor corresponds to the output wavefront for a particular incident angle. While this representation is straightforward and can work well in certain cases, it may not be practical if the number of incident angles is large, as it can lead to a significant increase in memory requirements. One way to get around it is to allocate a small number of angles for the explicit wavefront tensor and use interpolation to estimate the values for intermediate angles within the memory limits.

$$u_{\text{wg}} \in \mathbb{C}^{H \times W \times N \times K}, \quad (\text{S48})$$

where H and W is the resolution of the wavefront, N is the number of calibrated angles, and K is the number of channels to approximate the mutual intensity.

Physics-based representations. Note that we usually need to adopt a wide range of angles, typically around 20° . In explicit models, this requires storing wavefronts at over 100 discrete angles, resulting in billions of parameters and making the memory requirements prohibitive. To address this, we develop a physics-based waveguide model that parameterizes key properties such as diffraction efficiency of gratings and geometrical parameters, instead of learning



Supplementary Fig. S10 Illustration of physics-based modeling of a waveguide. **A.** The side view visualization of the waveguide slab shows the propagation from the input grating to the output grating. The out-coupled beam (in blue shades) and the leaked beam at the in-coupler (in red shades) are visualized. The coordinates of the input coupler, out-coupler, and the next patch are also displayed. **B.** Visualization of the modules for the physics-based waveguide model. The propagation between the gratings is modularized as the *Grating2Grating* module, which consists of a series of *TIRProp* modules.

the whole output wavefronts. By learning these parameters, we can generate the wavefront at arbitrary angles at runtime. In this model, the wavefront out of the waveguide is described as the output of a function, which is conditioned on the incident angle θ_{in} and is based on physical principles.

$$u_{wg, \theta_{in}} = \text{WG}_{\theta_{in}}(u_{in}), \quad \text{WG} : \mathbb{C}^{M_i \times N_i} \mapsto \mathbb{C}^{H \times W}, \quad (\text{S49})$$

where u_{in} is the input field of the waveguide, with a resolution of $M_i \times N_i$.

Our model aims to accurately simulate the wave propagation inside the waveguide and diffraction at grating couplers, which has traditionally been approached using either ray approximation methods or computational electromagnetics tools such as the Rigorous coupled-wave analysis (RCWA). These approaches, however, can be either inefficient or inaccurate. Here, we simulate the internal waveguide propagation and grating effects using wave optics.¹⁴ One of the challenges we face is simulating internal wave propagation at large angles, as this exceeds the bandwidth of the simulation. However, we observe that when a wave is propagated at a large angle, the bandwidth is simply shifted rather than extended. This insight enables us to accurately simulate wave propagation by combining the shifted angular spectrum method with

the shifted sampling window and affine transform, which takes into account geometric parameters, similar to Eq. S58. In this way, we are able to accurately model wave propagation at large angles. Grating effects, on the other hand, can be influenced by both spatial variations in the grating due to fabrication imperfections and the angular response of the grating itself. The former can be modeled using a complex-valued mask, while the latter can be simulated using the RCWA. It is worth noting that the RCWA coefficients often have a smooth envelope, which can be modeled through simple convolution with a kernel of small support.

If there are no interactions with gratings at the surfaces, the series of propagations with n total internal reflections can be represented as a single propagation with a distance that is n times longer. However, when the field comes into contact with the output grating or the input grating again, a portion of the light is out-coupled or leaked by the gratings (Supp. Fig. S10A). We propose modeling this series of actions as a recurrent module to represent the propagation through the waveguide. This physics-based model consists of a series of *Grating2Grating* modules, which represent the propagation from one grating to another. Each of these *Grating2Grating* modules performs the propagation between every other total internal reflection recurrently, which we have abstracted as *TIRProp*. A single *TIRProp* module takes in a 2D wavefield of the in-coupler and the coordinate of the center of the input field, and outputs information about the wavefield at the next interaction with the same surface: the out-coupled field, the remaining field to continue, and the coordinate of the continued field.

Our physics-based model offers several key advantages. First, it is fully analytical and therefore differentiable, which enables us to parameterize various geometric and physical parameters such as grating pitch and diffraction efficiency. These parameters can be learned using camera-in-the-loop training but also can be optimized for various purposes in waveguide design. Additionally, the model can accommodate a wide range of input angles beyond the sampling limit using the shifted angular spectrum method and geometric transformations and is suitable for a pupil-steered holography system. Lastly, it requires less memory than explicit representation while still offering robust generalization to untrained angles.

Algorithms 1–3 describe our physics-based waveguide model. While this model provides useful insights into wave propagation in waveguides, such as the lateral shift of the output due to phase shifts from the steered input and phase discontinuities, we note that it lacks the degrees of freedom necessary to model all the imperfections in the setup, such as distortion and warping. This motivates the need for a more flexible model that is agnostic to the geometric parameters of the waveguide and offers additional benefits.

Implicit neural waveguide representation. An affine transformation or warping of each reflected patch, which corresponds to the patches through different numbers of total internal reflections, occurs as a function of the steered angle. Importantly, the transformation of each copy differs because the propagation distance will be proportional to the tangent of the angle. The multiple copies are either coherently added up or incoherently added up based on the optical

ALGORITHM 1: Physics-based waveguide model

N : Number of gratings
 u_{in} : input field
 $\mathbf{e}_i$: diffraction efficiency of i th grating, $i = 1, \dots, N$.
 $\mathbf{p}_i$: grating pitch of i th grating, $i = 1, \dots, N$.
 $\mathbf{c}_i$: coordinate of the center of i th grating, $i = 1, \dots, N$.
 θ_{in} : incident angle on the incoupler
 n : the refractive index of the waveguide

$u_1 \leftarrow u_{in}$
 $\theta_1 \leftarrow \sin^{-1}(\sin(\theta_{in})/n)$ // Snell's law
foreach i **in** $1 \dots N$ **do**
 // Propagation from a grating to the next grating
 $u_{i+1}, \theta_{i+1} \leftarrow \text{Grating2Grating}(u_i, \theta_i, \mathbf{e}_i, \mathbf{e}_{i+1}, \mathbf{c}_i, \mathbf{c}_{i+1}, \mathbf{p}_i)$
end
return u_{N+1}

ALGORITHM 2: Grating2Grating

u_{prev} : input field
 θ_{prev} : incident angle on the input grating
 $\mathbf{e}_{prev}$: diffraction efficiency of the input grating
 $\mathbf{e}_{next}$: diffraction efficiency of the output grating
 $\mathbf{c}_{prev}$: coordinate of the center of the input grating
 $\mathbf{c}_{next}$: coordinate of the center of the output grating
 $\mathbf{p}$: grating pitch of the input grating
 n : the refractive index of the waveguide

$\theta_{next} \leftarrow \sin^{-1}(\sin(\theta_{prev}) + \frac{\lambda}{n\mathbf{p}})$ // Grating equation
 $u \leftarrow u_{prev}$
 $\mathbf{r} \leftarrow \mathbf{c}_{prev}$
 $u_{out} \leftarrow 0$
while *field is between gratings* **do**
 // propagation between contact at one surface
 $u, \mathbf{r}, u_{oc} \leftarrow \text{TIRProp}(u, \mathbf{r}, \theta_{next}, \mathbf{e}_{prev}, \mathbf{e}_{next}, \mathbf{c}_{prev}, \mathbf{c}_{next})$
 $u_{out} \leftarrow u_{out} + u_{oc}$
end
return u_{out}, θ_{next}

path difference (OPD). One can manually consider the fact and implement so that only adjacent copies interfere, but this would be computationally intensive. Moreover, one needs to fine-tune every model based on waveguide specifications which could be cumbersome.

ALGORITHM 3: TIRProp

u_{prev} : input field
 $\mathbf{r}_{prev}$: coordinate of the input field
 θ : propagation angle
 $\mathbf{e}_{prev}$: diffraction efficiency of the input grating
 $\mathbf{e}_{next}$: diffraction efficiency of the output grating
 $\mathbf{c}_{prev}$: coordinate of the center of the input grating
 $\mathbf{c}_{next}$: coordinate of the center of the output grating
 t : the thickness of the waveguide

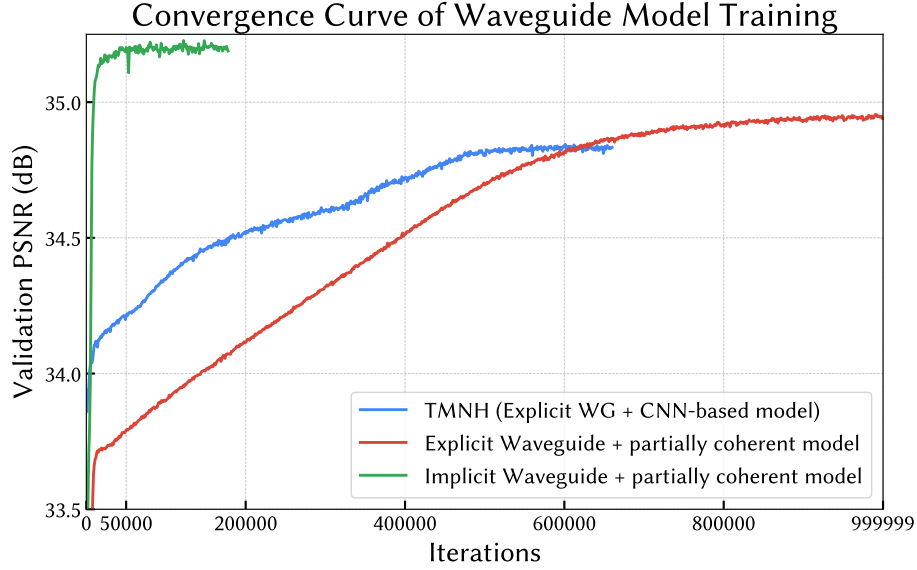
$\mathbf{r}_{next} \leftarrow \mathbf{r}_{prev} + 2t \tan \theta$
 $u_{next} \leftarrow \text{Shifted_ASM}(u_{prev}, 2t, \theta)$
 $u_{leaked} \leftarrow u_{next} \cdot \text{Affine_Transform}_{\mathbf{c}_{prev} \rightarrow \mathbf{r}_{next}}(\mathbf{e}_{prev})$
 $u_{oc} \leftarrow u_{next} \cdot \text{Affine_Transform}_{\mathbf{c}_{next} \rightarrow \mathbf{r}_{next}}(\mathbf{e}_{prev})$
 $u_{next} \leftarrow u_{next} - u_{leaked} - u_{oc}$
 $u_{oc} \leftarrow \text{Affine_Transform}_{\mathbf{r}_{next} \rightarrow \mathbf{c}_{next}}(u_{oc})$
return $u_{next}, \mathbf{r}_{next}, u_{oc}$

Here, based on physical insights, we propose a novel representation of the waveguide leveraging emerging neural representations.²⁷ In our waveguide model, we query by inputting the coordinates of the output pixels and the steered angle, forming multiple (K) channels of incident partially coherent wavefronts

$$u_{wg, \mathbf{u}_0}(\mathbf{r}, \mathbf{u}_0) = \text{WG}(\mathbf{r}, \mathbf{u}_0), \quad \text{WG} : \mathbb{R}^4 \mapsto \mathbb{C}^{K \times 1}. \quad (\text{S50})$$

One significant advantage of using implicit neural wavefronts is that warping can be efficiently achieved by querying different coordinates. Specifically, we first pass the angles and spatial coordinates to a small coordinate transform multilayer perceptron (MLP), which outputs the warped coordinates for each channel. These coordinates are then input into a multiresolution hash-based MLP,²⁷ producing multiple features that are decoded by a decoder MLP. Another benefit is that it is agnostic to any specific configuration. When outputting multiple channels, we can increase the output of the decoder MLP without significantly increasing the number of model parameters (see Table 1 in the main manuscript). However, this approach may limit the model’s ability to output arbitrary modes, so we employed multiple implicit neural waveguide representations for larger modes (e.g., $K = 16$ for the prototype).

The output of this model is a multi-channel wavefront, representing incident wavefronts at the Spatial Light Modulator (SLM), residual to the steered illumination angle. We manage the angle separately and input it into the off-axis wave propagation operator, while the residual phase is learned through training, as shown in Figure 2 in the main manuscript and Supplementary Figure S14. It is worth noting that the model has effectively learned the physics; for example, the abrupt change at the boundary of patches.



Supplementary Fig. S11 Convergence Curve of PSNR During Waveguide Model Training.

We show the convergence curve for training three models using 1000 phase–intensity pairs, including the previous state-of-the-art CNN-based model (TMNH),¹⁰ and the partially coherent wave propagation model (See Supp. Note 4) with explicit and implicit waveguide model. The models using explicit representations require hundreds of thousands of iterations to converge, while ours requires fewer than an order of magnitude of iterations (50k vs 1M) to converge, with better performance, uniquely enabling scalable training for a large étendue.

Comparison of waveguide models. Our waveguide model offers several benefits over previous state-of-the-art parameterized wave propagation models^{10,15,28} and uniquely enables scalable training for a large étendue. These benefits include fast convergence, better generalization capability requiring fewer training images for comparable performance, and generalization for unseen angles. While we will demonstrate most of these advantages in the next note, here we provide a direct comparison of our models against the previous state-of-the-art explicit CNN-based models.¹⁰

In Supplementary Fig. S11, we highlight one of the key benefits of implicit waveguide models. Specifically, we trained waveguide models using a dataset from the prototype (Figure 2 in the main manuscript) to evaluate their efficacy and performance, and the dataset consists of 1000 phase–intensity pairs. We evaluate them on a validation set of 300 pairs. Notably, our model converges much faster, requiring fewer than an order of magnitude of training iterations, which enables training in the large-étendue setting. Note that, while the difference between performances of these models in validation PSNR is only slightly different, the CNN-based model (TMNH) does a poor job of predicting intensity for unseen phases (See Figure 2 in the

manuscript). We attribute this to the CNN easily learning the spatial variance across the incident wavefront, which may not be physically accurate and fails to generalize to unseen phases. We also note that the explicit representation tends to fall into local minima more easily leading to slightly worse performance as shown in Supp. Fig. S11 (red line vs green line).

Supplementary Note 4: Partially coherent wave propagation models

In this note, we describe in detail the full partially coherent wave propagation model developed for our system. First, we use the Wigner distribution function (WDF),²⁹ as it provides a more intuitive description in the spatial frequency domain using convolutions, though the counterpart for the mutual intensity can also be derived. Next, we propose computationally efficient methods for handling the mutual intensity, following the same approach as in Eq. S14.

The full image formation model from the laser to the retina is as follows. The input beam for the waveguide, with an incident angle of $\mathbf{u}_0$, propagates through the waveguide, which generates a partially coherent wave denoted as its WDF $W_{\text{WG}}(\mathbf{r}, \mathbf{u} - \mathbf{u}^m)$, which we approximate with our waveguide model. The wave modulated by a phase-only SLM can be computed as the convolution of W_{WG} and the WDF of the SLM modulation W_{SLM} in the frequency domain.

$$W_{\text{SLM}}(\mathbf{r}, \mathbf{u}) = \iint_{-\infty}^{\infty} \left\langle e^{i\phi(\mathbf{r} + \frac{\mathbf{r}'}{2})} e^{-i\phi(\mathbf{r} - \frac{\mathbf{r}'}{2})} \right\rangle e^{-i2\pi\mathbf{r}'\mathbf{u}} d\mathbf{r}', \quad (\text{S51})$$

$$W_i(\mathbf{r}, \mathbf{u}) = W_{\text{SLM}}(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{u}} W_{\text{WG}}(\mathbf{r}, \mathbf{u} - \mathbf{u}^m), \quad (\text{S52})$$

where $\mathbf{r}, \mathbf{u}$ and $\otimes$ denote spatial and frequency coordinates and the convolution operator, respectively.

Next, this partially coherent wave is propagated in the free space, by the distance z to form the desired mutual intensity around the focal plane of the eyepiece. This can be simulated as the spatial convolution with the WDF of propagation kernel, h_z . Finally, the light field intensity observed through an aperture p can be computed as the four-dimensional convolution with the aperture WDF, W_p . These operations are all linear, thus we can simply average over time when we scan over the synthetic aperture with multiple phase patterns.

$$l(\mathbf{r}, \mathbf{u}) = \langle W_i(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{r}} W_{h_z}(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{r}, \mathbf{u}} W_p(\mathbf{r}, \mathbf{u}) \rangle, \quad (\text{S53})$$

$$= \left\langle \underbrace{\underbrace{W_{\text{SLM}}(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{u}} W_{\text{WG}}(\mathbf{r}, \mathbf{u} - \mathbf{u}^m)}_{\text{SLM modulation}} \otimes_{\mathbf{r}} W_{h_z}(\mathbf{r}, \mathbf{u})}_{\text{wave propagation}} \otimes_{\mathbf{r}, \mathbf{u}} W_p(\mathbf{r}, \mathbf{u}) \right\rangle. \quad (\text{S54})$$

intensity through aperture

With these equations, we can map phase patterns shown on the SLM $\{\phi\}$ with instantaneous aperture positions $\{\mathbf{u}^m\}$ to the image that a user would observe with a pupil p positioned at $\mathbf{u}$ while focusing at a particular depth, z . Note that, with the assumption of an ideal waveguide with uniform output, paraxial free-space propagation, and a large pupil, the model becomes

the Radon–Wigner transform²⁹ $l(\mathbf{r}) = \langle \int W_{\text{SLM}}(\mathbf{r} - \lambda z(\mathbf{u} - \mathbf{u}^m) d\mathbf{u}) \rangle$. Our wave propagation model uniquely characterizes and learns the partially coherent nature of the system, in the form of W_{WG} and W_{hz} , instead of a single coherent wavefront.

While this representation is computationally prohibitive to deal with, it provides rich insights into the behaviour of the signal in the spatial frequency domain, such as the migration of the energy, which motivates the use of bandwidth migration for coherent wave propagation models described in the next section.

Efficient wave propagation for synthetic apertures. The incident beam for creating synthetic apertures can be modeled with a wave vector $\mathbf{k}$ as plane waves $e^{j\mathbf{k}\cdot\mathbf{r}}$, but when the angle is steep enough, this approach exceeds the bandwidth the simulation supports and this causes aliasing. To accurately model wave propagation, one must oversample the pixels to accommodate the high-frequency plane wave, which, unfortunately, demands excessive computation and is often prohibitive. To address this challenge, we employ bandwidth migration leveraging Fourier properties. In this subsection, we describe an efficient diffraction calculation method that carries the same cost as the angular spectrum method.¹³ Note that we first deal with a single coherent, quasi-monochromatic wavefront, but we discuss its expansion to the partially coherent extension in the next section.

First, multiplying a plane wave to a wavefront f is equivalent to a coordinate shift in the Fourier domain:

$$\mathcal{S}(f, \mathbf{u}_0) = f(\mathbf{r}) e^{j2\pi(\mathbf{u}_0 \cdot \mathbf{r})}, \quad (\text{S55})$$

$$\mathcal{F}(f(\mathbf{r}) e^{j2\pi(\mathbf{u}_0 \cdot \mathbf{r})}) \leftrightarrow F(\mathbf{u} - \mathbf{u}_0), \quad (\text{S56})$$

where $\mathcal{S}$, $\mathcal{F}$ and F denote the steering, Fourier transform operator, and the Fourier transform of the wavefront, respectively. $\mathbf{r}$ and $\mathbf{u}$ represent coordinates in the spatial and Fourier domains, respectively. As mentioned, one can oversample the wavefront in the spatial domain, to avoid aliasing. However, the energy of the oversampled wavefront in the Fourier domain is concentrated in the central part, and the other parts are zero, which simply indicates this strategy is inefficient. The steering operation just steers the energy in the spatial angular domain, where we are interested in. Thus, we opt to migrate the energy in the Fourier domain, and sample at off-axis coordinates, as proposed by Matsushima et al.³⁰ Subsequently, by shifting the wavefront geometrically, we can seamlessly add up each mode of partially coherent wavefronts sequentially in the spatial domain (or across the field-of-view). This approach allows us to compute it at nearly the same computational cost as the original diffraction calculation:

$$f(\mathbf{r} - \mathbf{r}_0) \leftrightarrow \mathcal{F}^{-1}(F(\mathbf{u}) e^{j2\pi(\mathbf{u} \cdot \mathbf{r}_0)}), \quad (\text{S57})$$

Please note that Eq. S57 is essential not only for sampling the useful wavefront but also for preventing aliasing in the Fourier domain through the application of a linear phase ramp. By sampling at shifted coordinates, we can efficiently introduce spatial shifts without encountering aliasing. When considering the propagation kernel, the final wave propagation at a steered angle can be described as follows:

$$\mathcal{P}_z(\mathcal{S}(f, \mathbf{u}_0))(\mathbf{r} - \mathbf{r}_0) = \mathcal{F}^{-1}\left(F(\mathbf{u} - \mathbf{u}_0) \mathcal{H}_z(\mathbf{u} - \mathbf{u}_0) e^{j2\pi(\mathbf{u} \cdot \mathbf{r}_0)}\right), \quad (\text{S58})$$

Here, $\mathcal{P}_z$ denotes a wave propagation operator, such as the Fresnel diffraction or the angular spectrum method,¹³ and $\mathcal{H}_z$ represents the propagation kernel for propagation distance z in the Fourier domain. Equation S58 represents the wavefront propagated and sampled at the off-axis coordinates.³⁰ Again, to seamlessly merge these wavefronts, we geometrically shift the intensities and linearly add them up in a larger canvas in a reference coordinate system. In Supplementary Figure S12, we present an example of seamless rendering. As the illumination is directed at an angle θ , the center of the wavefront shifts in the spatial domain to $\mathbf{r}_0 = z \tan \theta$. We illustrate three examples in the second row. To achieve a seamless merge of content across the field of view (FOV), it is essential to render the correct content as shown in the third column of the first row. Otherwise, inconsistent merging occurs, leading to incorrect rendering, as shown in the second column. Supplementary Figure S12B demonstrates the same principle in 2D and 4D (light field content). As expected, when the light field is merged seamlessly, it exhibits the right photometric effect,³¹ such as the natural defocus behavior as shown in the top right.

Partially coherent wave propagation model. In this section, we describe our partially coherent wave propagation model. We also give the description in the phase-space optics, because it naturally connects to the theory of Synthetic Aperture Holography that we will present in the next Supplementary Note. For a more comprehensive background of the phase-space optics, we refer to Testorf et al.²⁹

In our model, we first approximate the partially coherent wavefront (mutual intensity) out of the waveguide, as a set of multiple coherent modes $f_{wg} = \{f_1, f_2, \dots, f_M\}$, known as coherent mode decomposition.^{22,23} The aforementioned off-axis wave propagation can be used for each mode, and we can simply sum them up in intensity to calculate the observed image at any given plane. From the phase-space optics perspective, the propagation of mutual intensity is a linear operation where its Wigner distribution function (Eq. S84) is convolved the Wigner counterpart of the diffraction kernel. The steering operation $\mathcal{S}$ shifts the entire energy in the Wigner domain at the amount of steered angle $\mathbf{u}_0$:

$$W_{\mathcal{P}(\mathcal{S}(f, \mathbf{u}_0))}(\mathbf{r}, \mathbf{u}) = W_f(\mathbf{r}, \mathbf{u} - \mathbf{u}_0) \otimes_{\mathbf{r}} W_h(\mathbf{r}, \mathbf{u} - \mathbf{u}_0), \quad (\text{S59})$$

where $\mathcal{P}, \mathcal{S}, \otimes_{\mathbf{r}}$ denotes the propagation, steering, and convolution over spatial dimension, respectively and h is the propagation kernel, such as the Rayleigh–Sommerfeld kernel.¹³ Eq. S59 shows that the light distribution is uniquely defined in the spatial angular domain by the propagation and steered angle. The observable light field can be obtained by convolving the WDF with the WDF of the synthetic aperture function, which shares the same bounding box.³² As the wavefronts propagate, the lateral shift of the wavefront across the FOV is far beyond the range supported by the simulation. Consequently, we migrate the bandwidth and geometrically shift the entire light field to seamlessly merge multiple modes, which uniquely allows us to access to simulate large Etendue at the same memory footprint (See Methods and the Full image formation section below.).

Modelling aberrations at large étendue using implicit neural representation. Pupil aberrations are usually modeled using Zernike polynomials.¹⁵ While a natural extension of this to a large étendue setting might be to use an MLP to predict the Zernike coefficients, this method struggles to support high-frequency aberrations (see Supplemental video). We use our implicit neural wavefront model to model the pupil aberrations at the Fourier plane. This approach can serve as a drop-in replacement and generalizes well to unseen angles due to its inherently continuous nature (see Supplementary Fig. S13). Similar to Equation S50, we model the aberration $\mathcal{A}$ with our implicit model WF :

$$\mathcal{A}(\mathbf{u}, \mathbf{u}_0) = \text{WF}(\mathbf{u}, \mathbf{u}_0), \quad \text{WF} : \mathbb{R}^4 \mapsto \mathbb{R}, \quad (\text{S60})$$

In the Wigner domain, this is equivalent to

$$W_{\mathcal{P}(f)}(\mathbf{r}, \mathbf{u}) = W_f(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{r}} W_h(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{r}} W_{e^{j\mathcal{A}}}(\mathbf{u}, -\mathbf{r}), \quad (\text{S61})$$

Monotonic, angle-dependent, nonlinear response of the SLM. Note that often the SLM’s response is nonlinear and spatially varying, and significantly affect the image quality of holographic displays.^{15,28} To address this challenge, several technique and modeling have been proposed. In our work, we extend Peng et al’s modeling to even adopt the angle dependent response, by using the spatially varying code per pixel, that is generated by our implicit neural representation INR that takes steered angle as input, outputting a spatially varying code. We also additionally constrain the MLP to be monotonic, with Lipschitz constraint.³³

$$\mathbf{c}(\mathbf{r}, \mathbf{u}) = \text{INR}(\mathbf{r}, \mathbf{u}), \quad \text{INR} : \mathbb{R}^4 \mapsto \mathbb{R}, \quad (\text{S62})$$

$$\phi_{\text{SLM}}(\mathbf{r}) = \text{MLP}_{\text{LUT}}(V, \mathbf{c}) \quad (\text{S63})$$

Full image formation model and its training. Now we have a full model for the wave propagation in our system, and we aim to retrieve the mutual intensity function of the partially coherent wavefront that is incident on the SLM, while we jointly optimize other parts, such as the Aberration and LUT to better characterize the system. The mutual intensity function is four dimensional which is computationally prohibitive to be stored using an explicit discrete representation, and we propose to learn the lower-rank approximation of the mutual intensity, parameterizing the multiple wavefront modes for rank- M : $\{f_1, f_2, \dots, f_M\}$.

To retrieve the partially coherent wavefront, we capture the intensity (or amplitude) of the resulting mutual intensity at the target plane a_{captured} with a sensor, at an arbitrary steered MEMS angle $\mathbf{u}_0$. It is important to note that we only capture at a single plane to retrieve the wavefront, and we do not require any interferometric setup. Starting with the multiple wavefronts, these wavefronts are multiplied by the SLM field $e^{j\phi_{\text{SLM}}}$ as described in Eq. S63. Then, we add angle-dependent DC terms for each mode, which is also outputted from the waveguide models as separate channels:

$$f_{\text{SLM}}^k = f_{\text{wg}}^k \cdot e^{j\text{MLP}_{\text{LUT}}(V, \mathbf{c})} + u_{\text{DC}}^k, \text{ for } k = 1, \dots, M. \quad (\text{S64})$$

Next, these wavefronts are propagated using the off-axis propagation described in Eq. S58, while we also sample at off-axis coordinates. Thus, this propagation is approximately identical to typical diffraction at the same computational cost, as long as it is within the bandwidth limit. Note that we simulate higher orders following Gopakumar et al.²⁵ Simultaneously, we geometrically shift the entire intensity to account for the propagation. Consequently, we use a larger canvas that exceeds the size of the original wavefront, specifically $l_{\text{SLM}} + z \tan(\theta_{\text{max}})$ for the maximum steering angle θ_{max} . The shifted intensities are added up in this large canvas after shift, and the loss function with respect to the captured intensity is calculated. Please note that a_{captured} is transformed to this reference frame using the homography at the nominal angle:

$$f_{\text{target}}^k = \mathcal{P}_z \left(\mathcal{S} \left((f_{\text{SLM}}^k, z), \mathbf{u}_0 \right) \right) (\mathbf{r} - \mathbf{r}_0), \quad k = 1, \dots, M., \quad (\text{S65})$$

$$a_{\text{out}} = \sqrt{\sum_k^M |f_{\text{target}}^k|^2}. \quad (\text{S66})$$

In conclusion, in our training procedure, we optimize the model parameters using the partial Wigner projection at each full synthetic aperture generated by a steered illumination, for a loss function

$$\min_{\{f_{\text{WG}}, f_{\text{DC}}\}, \text{MLP}_{\text{LUT}}, \text{WF}} \sum_{j=1}^J \sum_{\mathbf{u} \in \mathbf{U}} \mathcal{L}(a_{\text{out}}, a_{\text{captured}}), \quad (\text{S67})$$

where J is the number of (SLM phase, captured image) pairs and $\mathbf{U}$ is the set of steered angles in the dataset.

Our partially coherent model offers numerous advantages over conventional single-mode wave propagation models and even surpasses the performance of state-of-the-art CNN-based models. One of the most primary advantages is the generalization capability, which stems from its inherent linearity in the mutual intensity domain. As a result, the model shows comparable performance even with a magnitude less amount of data and training time. To demonstrate this, we compare our model with the recently proposed hybrid–physical wave propagation model by Choi et al.¹⁰ on datasets of different sizes. We trained models on datasets with different numbers of pairs, ranging from 10 to 10000, and evaluated the model’s performance on a test set of a fixed size (100 pairs). As shown in Fig. 2D in the manuscript.

Calibration and dataset acquisition. While most of the aberrations are learned in our model, we first calibrate the angles to capture the dataset and put it in the rectified coordinate. We calculate the homography between the sensor plane and the image plane as in previous works.^{10,15} From this homography relationship or the mean shift of dot patterns with respect to the central angle, we can calculate the angle of illumination. Note that there are severe aberrations and one can adjust the effective propagation distance depending on angles. The residual aberrations are learned in the model as the form of the phase profile in the Fourier plane. (See Supplementary Figure S13). Captured images are transformed into a coordinate that corresponds to the central angle, and appear to be similar to what we see in Supplementary Figure S12B. This uniquely ensures that our method merges the intensity over multiple apertures seamlessly.

Ablation study. Table S1 shows an ablation study to demonstrate the efficacy of each component in our model. Specifically, we trained each model using 300 phase-intensity pairs captured at 72 angles and evaluated them on the training and test sets from the same angles, as well as on 9 held-out angles that were not included in the training set. While each component adds performance, it clearly shows that modeling partial coherence is the key to learning wave propagation in the setup.

Visualization of the partially coherent wave propagation model. In Supplementary Fig. S13A, we show experimentally captured resolution chart images for different pupil positions. Naïve steering approaches (center) have severe aberrations at the edge of the extended eyebox due to angle-dependent aberrations and incident beam profiles. These effects are not easily modeled by existing, approximately shift-invariant CNN-based models. Using our implicit model, we can efficiently calibrate the system, achieving high image quality uniformly across the extended exit pupil, as shown in the inset. In Supplementary Fig. S13B, we show the visualization of learned aberrations ($\mathcal{A}$) for the extended exit pupil. Note that this learned representation is inherently continuous and generalizes well to unseen pupil positions, as demonstrated by the extensively sampled pupil positions shown here. Each aberration map includes higher-order aberrations; the central region is cropped and shown in Fig. 3B of the main manuscript.

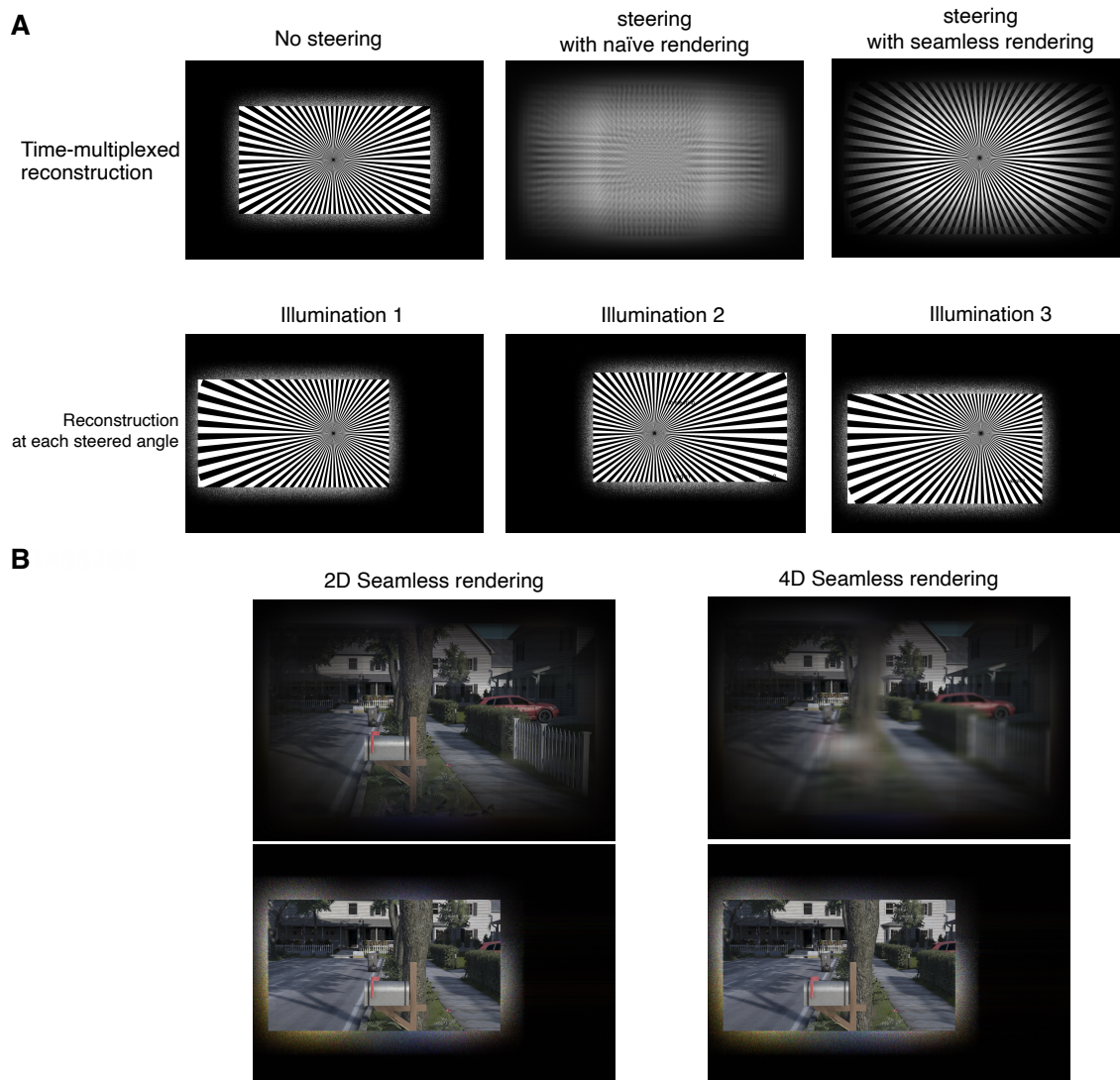
In Supplementary Fig. S14, we show the visualization of learned waveguide intensity for the extended exit pupil. Note that this learned representation is inherently continuous and generalizes well to unseen pupil positions, as demonstrated by the extensively sampled pupil positions shown here and the supplementary video.

Generalization capabilities of the implicit neural waveguide model. In Supplementary Figs. S15 and S16, we provide additional experimental results using 2D target images captured at multiple distances. Here, our model was trained on a single depth and a single aperture state (with the full aperture open) but achieves high-quality experimental results across varying conditions, demonstrating the generalization capabilities of our neural implicit model. In Supplementary Table S2, we also provide quantitative results measuring the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), averaged over five test images, showing that our method generalizes well across all tested conditions. We attribute this strong generalization ability to the physics-inspired nature of our model, which does not rely on any convolutional neural networks.

In Supplementary Fig. S17 and Table S3, we provide experimentally captured results obtained with multiple pupil sizes. Our model is trained on the full aperture at the center while achieving robust results for all pupil sizes.

Supplementary Table S1 Ablation study on our partially coherent wave propagation models on multiple angles. All models are trained on 72 angles. Peak signal-to-noise ratio (PSNR) is evaluated for train and test sets and 9 heldout angles.

Models	Train	Test	Heldout
ASM	26.6 dB	26.6 dB	26.8 dB
+ High-order	27.3 dB	27.6 dB	27.4 dB
+ SLM wavefront (implicit)	29.8 dB	29.9 dB	30.0 dB
+ Fourier phase	30.7 dB	30.8 dB	30.9 dB
+ LUT	31.6 dB	31.7 dB	31.4 dB
+ DC	31.7 dB	31.8 dB	31.7 dB
+ partial coherence (2 modes)	32.8 dB	32.9 dB	32.7 dB
+ partial coherence (3 modes)	33.5 dB	33.6 dB	33.3 dB
+ partial coherence (4 modes)	33.6 dB	33.7 dB	33.4 dB
+ partial coherence (6 modes)	34.1 dB	34.2 dB	33.8 dB



Supplementary Fig. S12 Illustration of geometric shifts across FOV and its seamless merging. Note that when correctly rendered with the light field, it shows the natural defocus blur as shown in **B**.

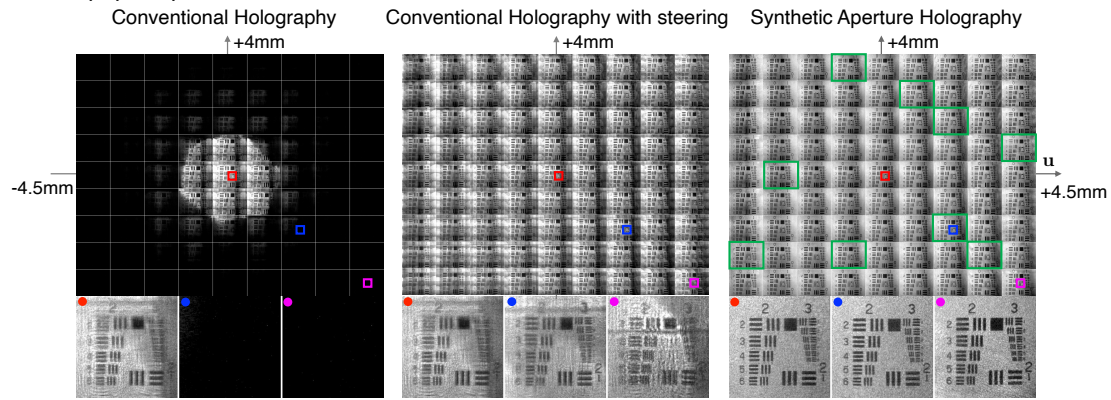
Supplementary Table S2 Quantitative evaluation of captured experimental results across varying depths reported in diopters (D). We report (PSNR (in dB) / SSIM) values averaged over five test images. Our model consistently achieves an average PSNR of over 20 dB across all depths, outperforming the baseline angular spectrum method (ASM) by 5 dB. Notably, despite being trained only on images captured at 0D, it robustly maintains high image quality across all untrained depths. All results are captured with a pupil size of 5mm diameter.

Model \ Distance	0D (inf)	0.6D	1.2D	1.8D	2.4D	3.0D
ASM	16.2 / 0.41	16.2 / 0.41	16.2 / 0.41	16.2 / 0.42	16.1 / 0.41	16.1 / 0.41
Ours	21.3 / 0.58	21.3 / 0.59	21.2 / 0.59	21.0 / 0.58	20.8 / 0.57	20.8 / 0.59

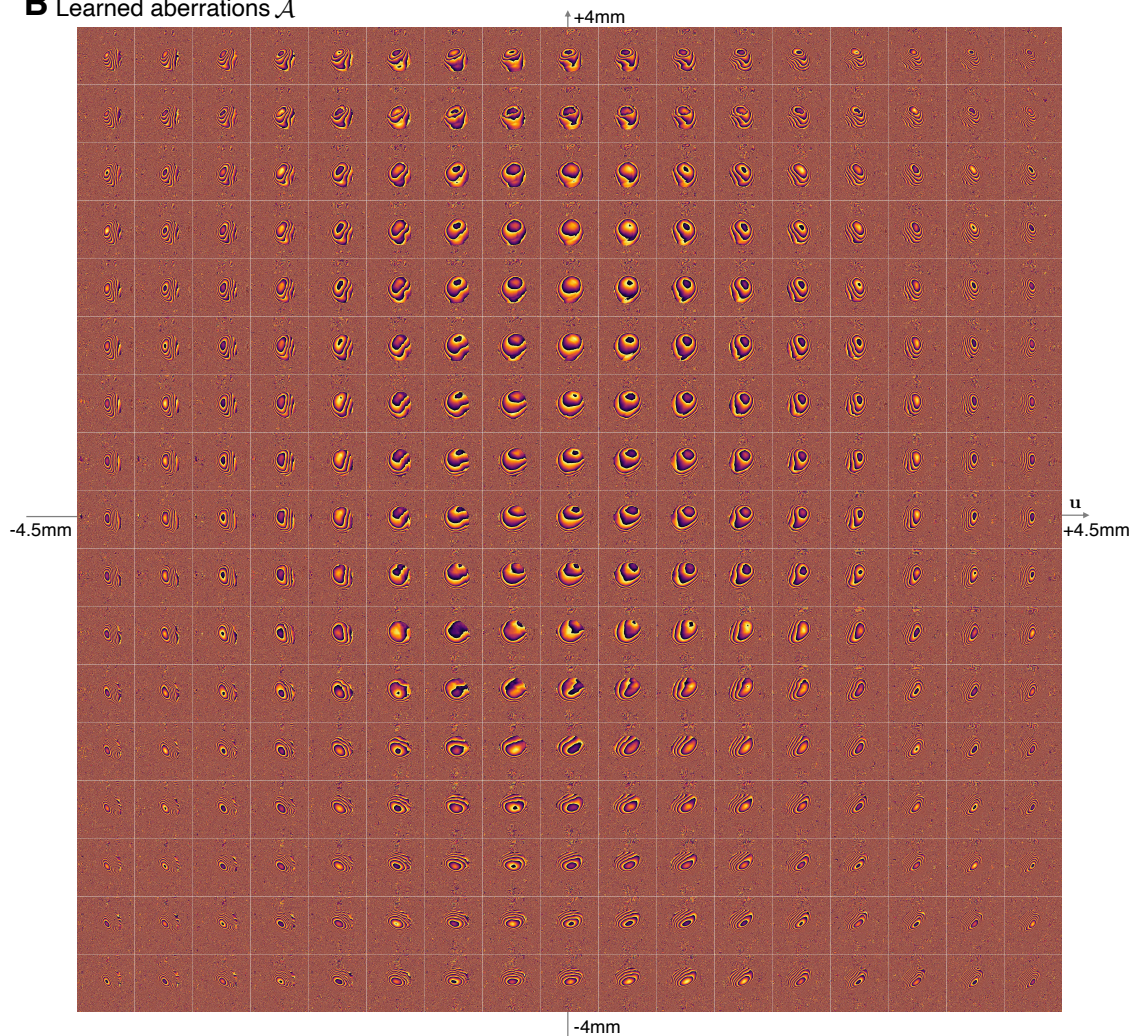
Supplementary Table S3 Quantitative evaluation of captured experimental results across varying pupil sizes. We report PSNR (in dB) and SSIM values averaged over five test images. Our model consistently achieves an average PSNR of over 20 dB across all pupil states, outperforming the baseline angular spectrum method (ASM) by 5 dB. Notably, for all states, we do not adaptively change phase patterns; instead, our synthetic aperture holography generates a true light field, ensuring high quality across all aperture states. Moreover, our model is not specifically trained on these pupil states—it is trained on a full aperture state of 30mm diameter, including all higher orders, allowing it to flexibly account for all pupil states when used for CGH.

Model \ Aperture size (diameter)	8mm	5mm	4mm	2.5mm
ASM	15.5 / 0.39	16.2 / 0.41	15.9 / 0.41	15.7 / 0.40
Ours	21.0 / 0.60	21.3 / 0.58	21.3 / 0.59	21.6 / 0.58

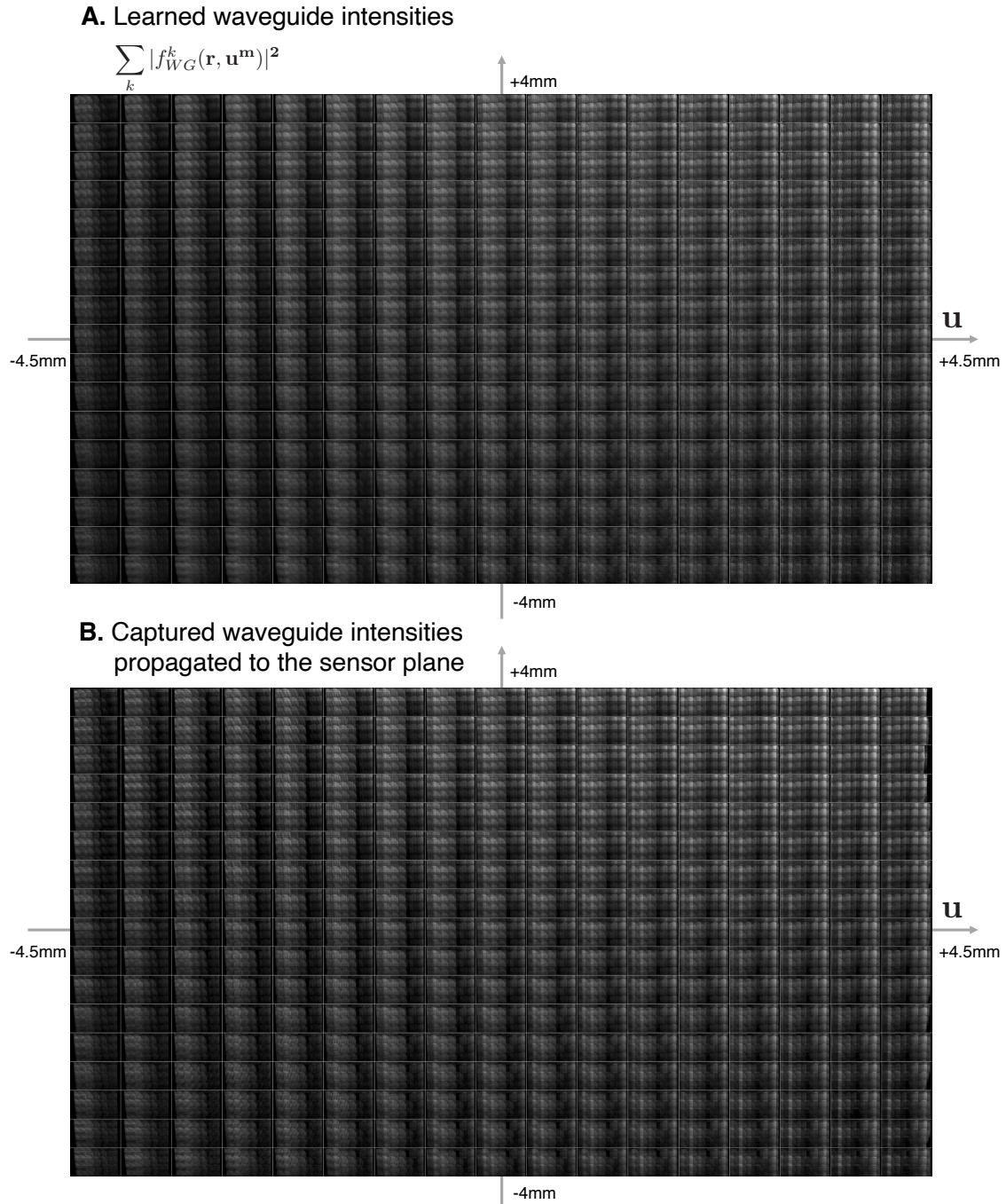
A Exit-pupil expansion



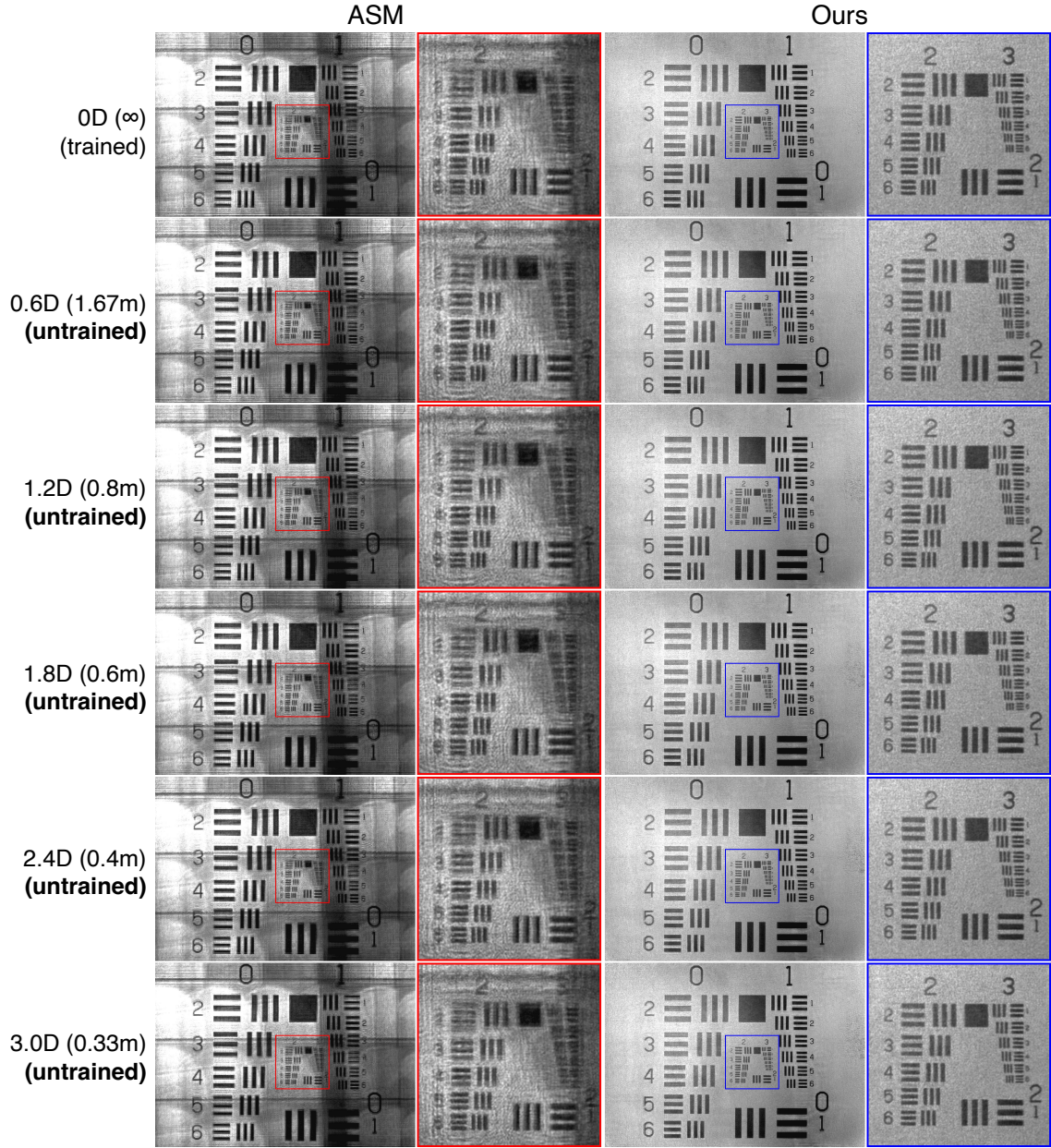
B Learned aberrations $\mathcal{A}$



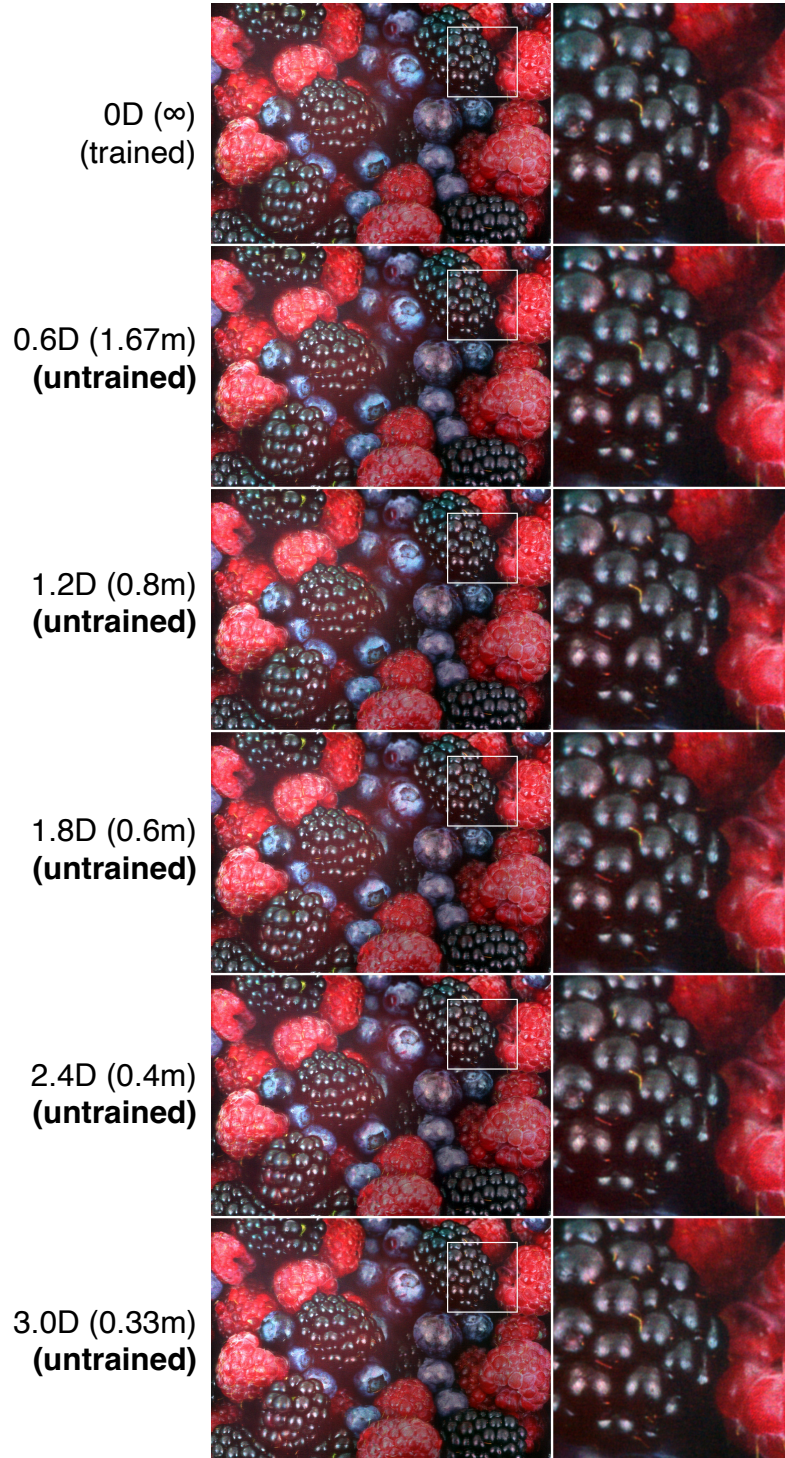
Supplementary Fig. S13 A. Experimentally captured resolution chart images for different pupil positions. Naïve steering approaches (*center*) have severe aberrations at the edge of the extended eyebox due to angle-dependent aberrations and incident beam profiles. These effects are not easily modeled by existing, approximately shift-invariant CNN-based models. Using our implicit model, we can efficiently calibrate the system, achieving high image quality uniformly across the extended exit pupil, as shown in the inset. Green boxes denote angles that are not trained, demonstrating its generalization capability. **B.** Learned aberrations ($\mathcal{A}$) for the extended exit pupil. Note that this learned representation is inherently continuous and generalizes well to unseen pupil positions, as demonstrated by the extensively sampled pupil positions shown here. Each aberration map includes higher-order aberrations; the central region is cropped and shown in Fig. 3B of the main manuscript.



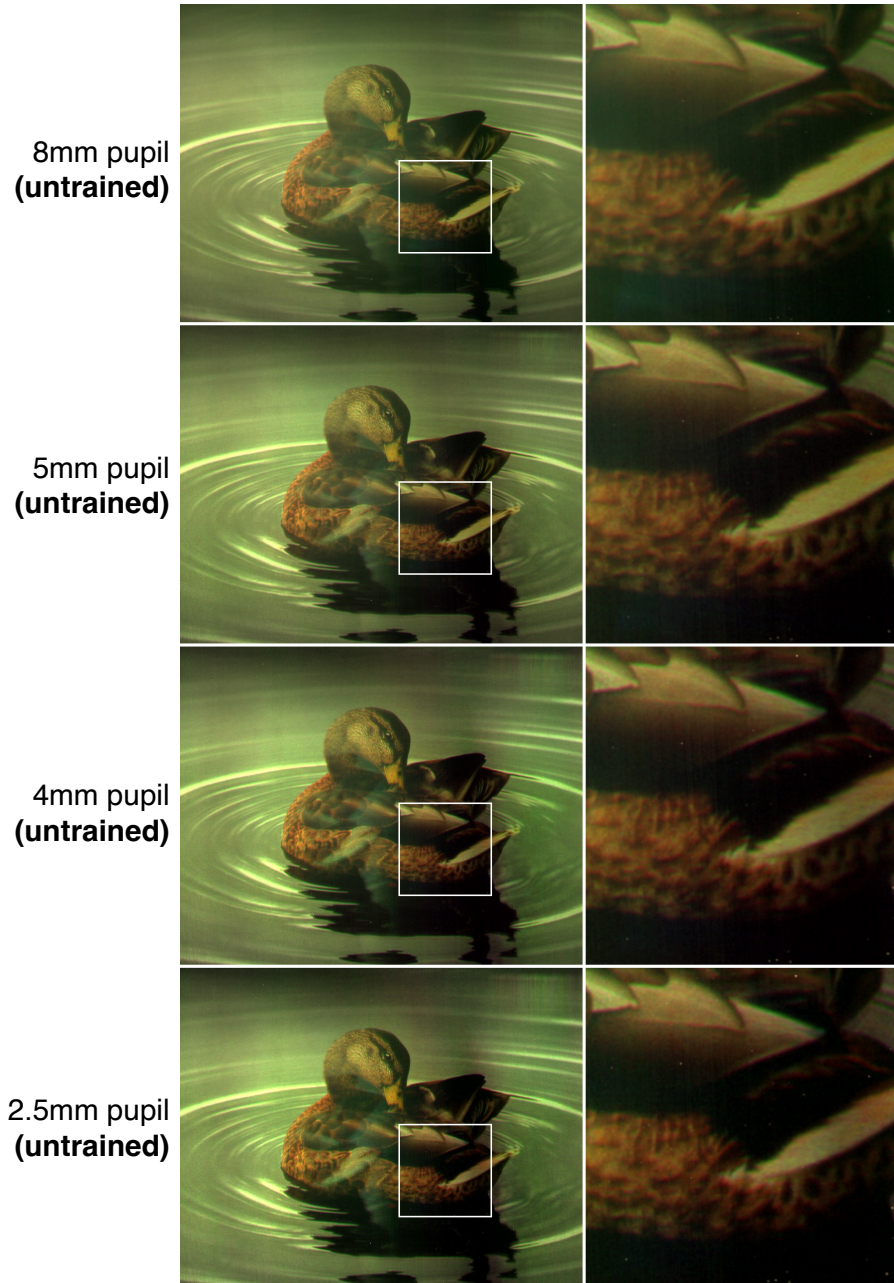
Supplementary Fig. S14 A. Visualization of the learned waveguide intensity for the extended exit pupil. Note that this learned representation is inherently continuous and generalizes well to unseen pupil positions, as demonstrated by the extensively sampled pupil positions shown here. Each waveguide intensity is calculated as the sum of the coherent-mode intensities. **B.** For reference, we additionally show the captured waveguide intensities (at the image plane) with a blank phase pattern on the SLM, whose trend matches the learned waveguide intensities.



Supplementary Fig. S15 Experimental results of a resolution chart captured at multiple distances. We capture experimental results of CGH methods using the angular spectrum method model (left) and our trained model (right). Note that our model is trained only on images captured at 0D (the top row), yet it achieves high image quality across all depths, demonstrating the generalization capabilities of our model.



Supplementary Fig. S16 Experimental results of a natural image captured at multiple distances. Phase patterns are optimized using our implicit neural waveguide model, which is trained only on images captured at 0D (top row). Our mutual-intensity-based implicit neural waveguide model maintains high image quality across all depths, demonstrating strong generalization capabilities.



Supplementary Fig. S17 Experimental results of a natural image captured with varying pupil sizes. The phase patterns are optimized using our implicit neural waveguide model, which is trained exclusively on images captured with the full-aperture state (30 mm) and not on any of the other pupil sizes shown here. Our mutual-intensity-based implicit neural waveguide model preserves high image quality across all pupil sizes, demonstrating strong generalization capabilities. Note that, our CGH algorithm optimizes the higher diffraction orders, which contributes to lower contrast for a larger pupil.

Supplementary Note 5: Synthetic Aperture Holography

In this note, we elaborate on the theory of the rendering part of *Synthetic Aperture Holography* (SAH) systems and its extension to large-étendue settings; our SAH variant uses a waveguide with scanning elements, which we refer to as *Synthetic Aperture Waveguide Holography*. First, we develop the theory of SAH through the lens of phase-space optics, which provides rich insights into the behavior of partially coherent wavefronts. Specifically, we formulate the computer-generated holography (CGH) as a coherence retrieval problem, as an extension of the classical phase retrieval.³⁴ From this perspective, we discuss various types of CGH supervisions and revisit its analogies in phase-space optics. The direct outcome of this is a complete loss function for SAH, which we introduced in Eq. 3 in manuscript. Consequently, we discuss the conventional 2D, 3D, and 4D CGH within the phase-space optics framework, clarifying how they constrain the mutual intensity, or the WDF.

Conventional CGH algorithms are often treated as a phase retrieval problem, where they retrieve a single coherent wavefront (or amplitude, or phase, based on their spatial light modulators) for the target image. While this approach usually assumes perfect coherence, in most practical cases, the complete image formation model from the source to the retina would be partially coherent, as we discussed in Supplementary Notes 2–4. In SAH, we extend the concept of phase retrieval for CGH to coherence retrieval, where we retrieve the mutual intensity of the light field instead of a single wavefront. Conventional CGH is a special case of this coherence retrieval, in which we restrict the rank of mutual intensity to be 1. This framework generalizes the previous image formation models^{10,17} and offers rich insights into existing supervisions for CGH from phase-space optics perspectives.

The goal for SAH is to physically implement the target light distribution, which is parameterized by a synthetic image plane (the $\mathbf{r}(xy)$ plane) and the pupil plane (the $\mathbf{u}(uv)$ plane), using the partially coherent holographic engine (See Supplementary Figure S20). Under the paraxial approximation, the value of the irradiance image that appears on the retina is given by³¹

$$I_{\text{target}} = \int l_{\text{GT}}(\mathbf{r}, \mathbf{u}) A(\mathbf{u}) d\mathbf{u}, \quad (\text{S68})$$

where A represents an arbitrary synthetic aperture function at the pupil plane. When a partially coherent engine generates a Wigner distribution function W_f (or a mutual intensity J) associated with the M coherent modes $f_{1,\dots,M}$, the retinal image resulting from the synthetic aperture A is the projection of the Wigner distribution of wavefronts that pass through the aperture, along the angular axis u . Since applying the aperture function at the pupil plane is equivalent to convolution with the corresponding filter function p in the spatial domain (for each single mode), we obtain

$$I_{\text{generated}} = \int W_{(\{f\}_m \otimes_{\mathbf{r}} p)}(\mathbf{r}, \mathbf{u}) d\mathbf{u} \quad (\text{S69})$$

$$= \int (W_{\{f\}_m}(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{r}} W_p(\mathbf{r}, \mathbf{u})) d\mathbf{u}. \quad (\text{S70})$$

In other words, we aim to generate a light field from the partially coherent holographic engine, namely Eq. S70, to match the target light field images in Eq. S68.

$$I_{\text{generated}} \simeq I_{\text{target}} \leftrightarrow \int l_{\text{GT}}(\mathbf{r}, \mathbf{u}) A(\mathbf{u}) d\mathbf{u} \simeq \int W_{\{f\}_m}(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{r}} W_p(\mathbf{r}, \mathbf{u}) d\mathbf{u}. \quad (\text{S71})$$

Ideally, this could be achieved by matching the integrands:

$$l_{\text{GT}}(\mathbf{r}, \mathbf{u}) A(\mathbf{u}) \simeq W_{\{f\}_m}(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{r}} W_p(\mathbf{r}, \mathbf{u}). \quad (\text{S72})$$

Unfortunately, matching the two in the above equation is challenging for several reasons. Firstly, calculating the Wigner distribution is computationally prohibitive, which makes the direct optimization of Eq. S72 not feasible. Secondly, due to the quadratic nature of the Wigner function, optimizing for individual aperture states does not necessarily ensure a match for other apertures. For example, consider that we have optimized the engine for two small apertures. Then, as soon as we consider the light passing through the union of two apertures, the generated light field would not match the incoherent sum of target light fields, with the crosstalk as they coherently adds up. Consequently, it is crucial to ensure that the approximation of Eq. S72 is valid across all potential apertures.

To address these issues, we first note that a holographic display system that would support M coherent modes would only be able to reproduce the desired light field with a rank- M representation of the mutual intensity. Thus, we consider the modal expansion of the mutual intensity, and truncate it to the first M modes, which effectively represents the mutual intensity in a computationally feasible way. Moreover, we propose to indirectly match finite projections of the Wigner distribution, blurred by the aperture Wigner function, with the corresponding irradiance image values that is the finite integral of target light fields, **for all possible aperture states** $\mathbf{U} \subset \mathcal{U} \times \mathcal{V}$, where $\mathcal{U} \times \mathcal{V}$ is the set of aperture functions for given light fields. In other words, we optimize for the loss function described in Eq. S73, (which is the same with Eq. 3 of the main paper)

$$\min_{\{f_m\}} \sum_{\mathbf{U} \subset \mathcal{U} \times \mathcal{V}} \mathcal{L} \left(\left\langle \left| f(\mathbf{r}; z, m) \otimes_{\mathbf{r}} \int_{\mathbf{u} \in \mathbf{U}} dp(\mathbf{r}, \mathbf{u}) \right|_m^2 \right\rangle, \int_{\mathbf{u} \in \mathbf{U}} l_{\text{GT}}(\mathbf{r}, \mathbf{u}) d\mathbf{u} \right). \quad (\text{S73})$$

Here the synthetic aperture function A is replaced as the sum of finite sub-apertures $\int dp$. As we are optimizing the loss function for a set of wavefronts $\{f_m\}$ (or phase patterns $\{\phi_m\}$),

this formulation is similar to the coherence retrieval where the mutual intensity is discretized as a Hermitian positive semidefinite matrix.³⁴ It is important to note that this approximation not only offers a practical method for CGH to approximate the correct light field but also provides valuable insights into multiple aspects of partially coherent CGH. These include understanding the role of the synthetic aperture function and offering an interpretable explanation of CGH loss functions for target distributions through the perspective of phase–space optics.

The role of synthetic aperture function. Equation S70 reveals important insights regarding the role of the synthetic aperture function. It is important to note that the convolution operation between the aperture Wigner Distribution Function (WDF) and the main WDF is always evaluated, even without the integral, when we assess the observable light field. This is the consequence of the uncertainty principle in optics – we cannot observe precisely both the location and direction of a photon. Indeed, for any Wigner distribution function, the product of the spatial spread and the angular spread – measured as the variance along each axis – has a lower bound:

$$\Delta \mathbf{r} \Delta \mathbf{u} \geq \frac{\lambda}{4\pi}, \quad (\text{S74})$$

where $\Delta \mathbf{r}$, $\Delta \mathbf{u}$ are the variance of the WDF along $\mathbf{r}$ and $\mathbf{u}$ axes, respectively. It is also noteworthy that after applying this WDF blur, the observable light field becomes non-negative.³² This implies that this could be understood as a minimal unit of *the effective étendue* and we will revisit this interesting and significant relationship between the observable light field and the effective étendue in Supplementary Note S8.

The blur kernel, namely the aperture WDF, can have a small extent either on the spatial axis (indicating a large aperture) or on the angular axis (indicating a small aperture), but not on both simultaneously. This represents each limitation of the 4D supervision (fine angular control but low resolution) and the 2D/3D supervision (sparse angular control but high resolution), as they utilize the corresponding evaluation of the observable light field for the loss function (See Supplementary Figs. S18 and S19). SAH achieves the best of both worlds by jointly considering these aspects, which we will elaborate on in the following sections.

CGH supervisions through the lens of phase–space optics

In this section, we make it clear that the conventional 2D, 3D, and 4D CGH are trying to optimize the subsets of SAH loss function, clarifying that how they address ill-posed problems in CGH that may not fully reproduce the plenoptic function without sufficient prior information.

In 2D and 3D supervision cases, we use full apertures $A = 1$, or the aperture WDF being the delta function so we can omit along with the convolution in Eq. S72. This yields

$$\int l_{\text{GT}}(\mathbf{r}, \mathbf{u}) d\mathbf{u} \simeq \int W_{\{f\}_m}(\mathbf{r}, \mathbf{u}) d\mathbf{u}. \quad (\text{S75})$$

2D supervision. Note that the above equation exactly corresponds to the conventional 2D image supervision, where one directly supervises the time-averaged intensity with a 2D image. For right-hand side, this corresponds to the Wigner projection along the entire angular axis, where you get the time-averaged intensity. See Supplementary Figures S20 and S21.

3D supervision. From the literature on synthetic light field photography,³¹ we can obtain the 3D focal stack using the digital refocusing technique simply by integrating light fields along sheared axis from the LHS. This provides the incoherent 3D focal stack at depth z

$$I = \int l_{\text{GT}}(\mathbf{r} - \lambda z \mathbf{u}, \mathbf{u}) d\mathbf{u}. \quad (\text{S76})$$

In phase-space optics, integrating it along sheared axis corresponds to the Radon–Wigner transform where the Wigner distribution is projected along a sheared axis.²⁹ The shearing operation could be replaced by the convolution with $\delta(y - ax)$, which is the form of the Wigner distribution functions of Fresnel kernel. Thus, from RHS we obtain the intensity image propagated after z using the Fresnel propagation $\{f\}_m \otimes_{\mathbf{r}} h$.

$$I = \int W_{\{f\}_m}(\mathbf{r} - \lambda z \mathbf{u}, \mathbf{u}) d\mathbf{u} \quad (\text{S77})$$

$$= \int W_{\{f\}_m}(\mathbf{r}, \mathbf{u}) \delta(\mathbf{r} - \lambda z \mathbf{u}) d\mathbf{u} \quad (\text{S78})$$

$$= \int W_{\{f\}_m}(\mathbf{r}, \mathbf{u}) \otimes_{\mathbf{r}} W_h(\mathbf{r}, \mathbf{u}) d\mathbf{u} \quad (\text{S79})$$

$$= \int W_{\{f\}_m \otimes_{\mathbf{r}} h}(\mathbf{r}, \mathbf{u}) d\mathbf{u}. \quad (\text{S80})$$

This supervision constrains the WDF along the projection with the full aperture, but with several sheared axes based on the depths. This indirectly provides some degree of spatio-angular control but lacks accurate view-dependent effects. This is also a common approach for retrieving the complex wave field using phase-space tomography.³⁵ See Supplementary Fig. S20.

4D supervision. In light-field supervision,¹⁰ we assume a finite, fixed pupil size, which can be efficiently calculated using the Short Time Fourier Transform (STFT).^{10,36} This corresponds to a 4D convolution of the WDF with the aperture WDF.³² While this blurring operation is intuitive and exactly corresponds to Equation S70, it can also be viewed as a finite projection along the angular axis, with the width corresponding to the pupil size (see Supplementary Figure S20), resulting in the observable light field, which is always positive.

This light-field supervision guides our optimization using the individual light field for mutual intensity, sampled through the corresponding pupil position and size; however, it does not account for interference between pupils. This can be understood as the crosstalk term in the WDF domain,³² mainly due to the coherent image formation model of the hologram. This

crosstalk significantly lowers the contrast of the images when sampled, as shown in Fig. 4 of the main manuscript.

ALGORITHM 4: Synthetic Aperture Holography (SAH)

```

 $\mathcal{V}$  : a set of viewing directions  $\{v_i = \sin\theta_i\}$ 
 $\mathcal{W}$  : a set of window functions  $\{w_i\}$ 
 $K$  : number of iterations,  $\approx 500$ 
 $\text{loss}(\cdot)$  : loss function (e.g. L2 loss)
 $\text{loss\_bp}(\cdot)$  : backpropagation through loss function
 $\text{model\_p}(\cdot; \theta)$  : propagation through parameterized model
 $\text{model\_bp}(\cdot; \theta)$  : backpropagation through parameterized model
 $s$  : scalar
 $x_{\text{SLM}}$  : SLM pixels
 $a_{\text{target}}(\cdot, v)$  : target image at a viewpoint  $v$ 
 $\theta_{\text{bl}}$  : steering angle using backlight
foreach  $k$  in  $1 \dots K$  do
     $\mathcal{V}_k, \mathcal{W}_k \leftarrow \text{sample}(\mathcal{V}, \mathcal{W})$ 
     $\{q_m\} \leftarrow \text{model\_p}(e^{j\phi_{\text{SLM}}}; \theta_{\text{bl}})$   $\triangleright$  partially coherent modes
     $\mathcal{L} \leftarrow 0$ 
    foreach  $v, w$  in  $\mathcal{V}_k, \mathcal{W}_k$  do
        %% This loop can be run in parallel %%
         $l(x, v) \leftarrow \sqrt{\sum_m^M |q_m(x) * (w \cdot e^{j\frac{2\pi}{\lambda} \sin(\theta - \theta_{\text{bl}}) \cdot x})|^2}$   $\triangleright$  Generalized radiance functions
         $\mathcal{L} \leftarrow \mathcal{L} + \text{loss}(s \cdot l(x, v), a_{\text{target}}(x, v))$ 
    end
     $\{\phi_{\text{SLM}}, s\} \leftarrow \text{model\_bp}(\text{loss\_bp}(\mathcal{L}))$ 
end
return  $\phi_{\text{SLM}}$ 

```

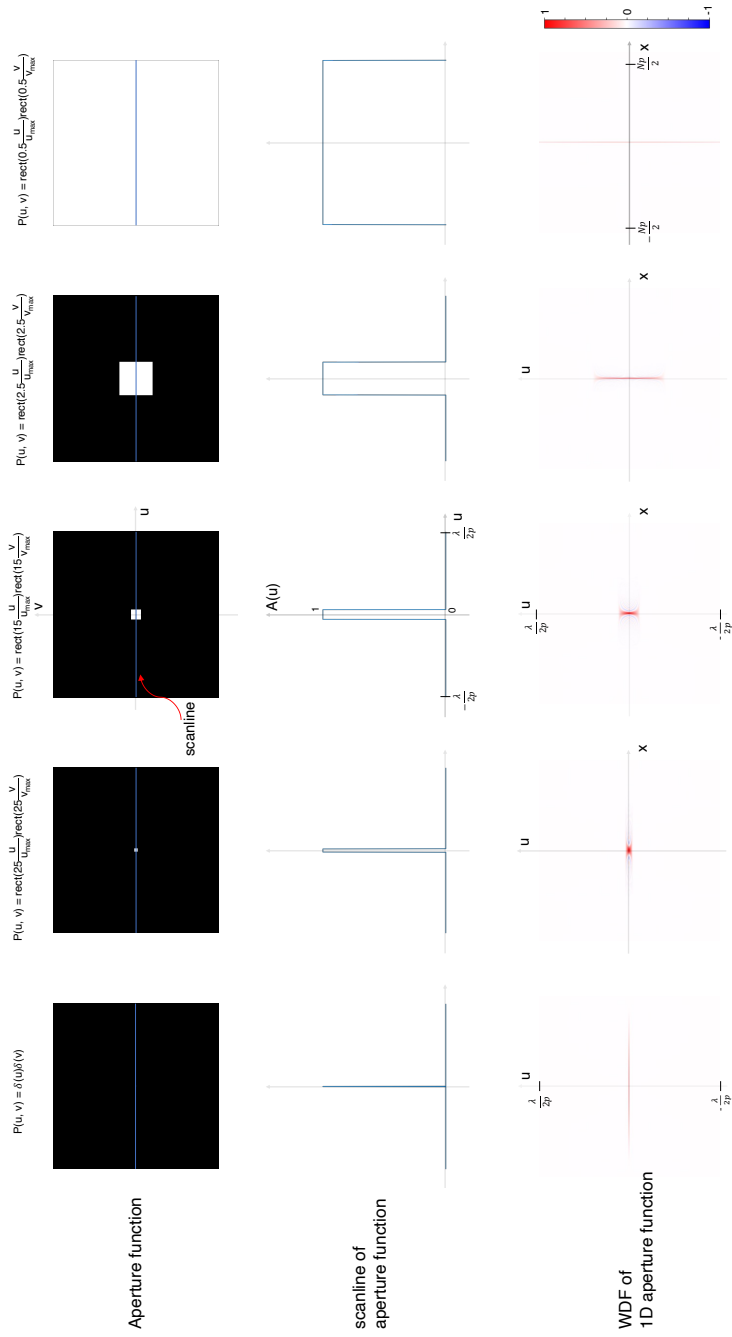
Synthetic Aperture Holography. We present the pseudocode for our Synthetic Aperture Holography rendering in Algorithm 4. Note that 3D and 4D supervisions are special cases of this algorithm, depending on the sampling strategy for the sets of viewing directions and window functions, $\mathcal{V}$ and $\mathcal{W}$.

In Supplementary Figure S20, we illustrate the comparison between Synthetic Aperture Holography (SAH) and conventional 3D holography. In conventional 3D holography (top left), the full wavefront at multiple depths (represented by red, orange, and purple planes) is supervised by corresponding target images. This method samples across the entire angular axis following a shearing operation in the spatio-angular domain. When viewed through the Wigner distribution function in phase optics space, the spatio-angular information of the wavefronts (illustrated by gray shades) is supervised after being projected along the sheared axes (bottom left).

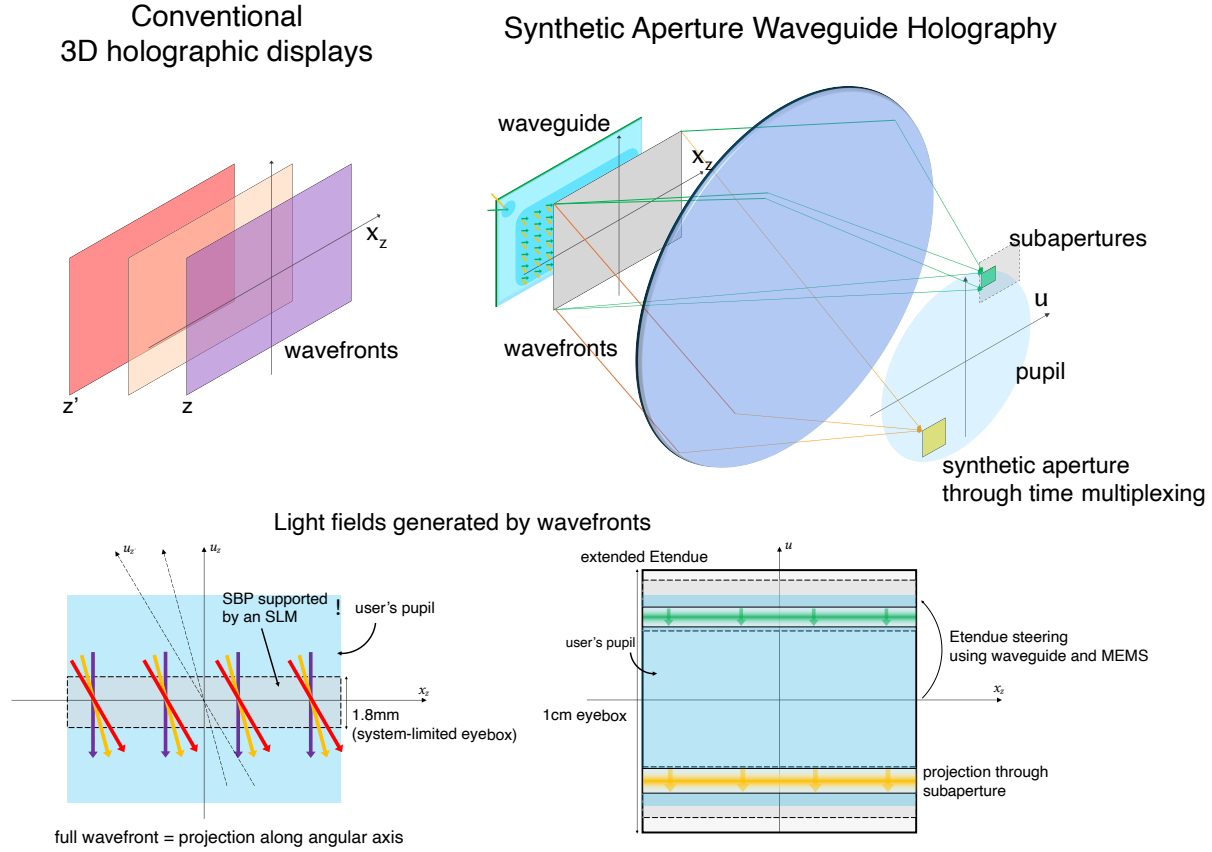
In Synthetic Aperture Holography, we supervise the projections along these axes with finite width rather than across the full axis. This approach samples specific sub-apertures at certain positions and sizes (represented by green and yellow), allowing fine control over spatio-angular information. We consider all possible combinations of these sub-apertures to form a synthetic aperture, which inspired its own name.

These properties are well demonstrated in Supplementary Figure S21. In Supplementary Figure S21A, we show the amplitude of the partially coherent wavefronts optimized using different CGH methods, including 2D supervision,¹⁵ 4D supervision,¹⁰ and our method. In conventional 2D holography, the full wavefront is supervised by target images, and the behavior of the resulting observable light field from the hologram depends on the initialization of the phase or the pre-assumed target phase profile (*bottom epipolar plane images*). The light-field hologram supervises the optimization for light-field view images through the sub-apertures, ensuring full light-field reconstruction over the full NA while suffering from low resolution when viewed through the whole aperture (*center*). Our synthetic aperture supervision achieves the best of both worlds by considering the synthetic aperture that consists of sub-apertures. In Supplementary Figure S21B, we show a static eyebox control uniquely enabled by our SAH supervision. Since our method provides a principled way to generate the light field using partially coherent engines, we can control the eyebox distribution by arbitrarily setting the Fourier profile of the target light fields. Here, we show that we can control the size of the eyebox using our optimization. Note that this cannot be done using the light-field supervision, due to the crosstalk between the sub-apertures that they do not consider. On the right, we show the results optimized using the stochastic variants of our method,^{37,38} following the approach described in their manuscript, which provides an eyebox shape overfitted to certain pupil shapes and positions. We attribute this to the gradient descent optimization for the hologram getting stuck in local minima, while our method is robust to this issue due to its physically correct principle for reconstructing incoherent light fields using partially coherent engines.

Our waveguide system can shift the entire bandwidth beyond what a typical SLM supports, enabling supervision over large-baseline light fields when combined with time multiplexing. This is achieved with only a marginal increase in system thickness, as demonstrated in the manuscript.



Supplementary Fig. S18 The role of synthetic aperture function. Here, we demonstrate the 2D aperture function, a 1D scanline of it, and the corresponding WDF for the scanline across various sizes of rectangular-shaped pupils. It is important to note that these WDFs act as kernels for the Wigner distribution when evaluating irradiance values, highlighting a clear trade-off between angular and spatial blurs. For large pupils, we can achieve diffraction-limited resolution, albeit without control over angle selection. Conversely, using small pupils results in always blurred images, yet allows for control over angles by moving the pupils - this merely shifts the entire WDF along the u -axis.



Supplementary Fig. S20 Comparison between conventional 3D holography and Synthetic Aperture Holography (SAH). In conventional 3D holography (*top left*), the full wavefront at multiple depths (denoted as red, orange, purple) is supervised by target images, corresponding to constraining the projection of the wavefronts along the entire angular axis after a spatio-angular shearing operation (*bottom left*). This approach does not support fine angular variations of the wavefronts. Additionally, the eyebox supported by the system is typically much smaller than the user's pupil (*gray shades vs. blue shades*). In contrast, SAH supervises projections with finite-width sampling over specific sub-apertures (*top right*), enabling fine control of spatio-angular information. The synthetic aperture is formed by combining these sub-apertures and also encompasses the full wavefront case when all sub-apertures are considered. The waveguide system extends the effective étendue (*bottom right, bold outline*) through steering and time multiplexing, with minimal added system thickness.

Supplemental Note 6: Additional simulation results

Supplemental Figure S22 shows the Zemax simulation of our holographic lens demonstrating the apochromatic characteristics and the conceptual diagram of 3D eyebox expansion we present in Figure 4 in the main manuscript.

To train a wave propagation model at enlarged étendue, we capture a dataset consisting of 300 pairs of SLM phases and captured images at 9×9 angles. We then train the model at 72 angles, leaving 9 angles as held-out data. Unlike previous approaches where only the region of interest (ROI) is supervised for training, we penalize all the regions where energy can be present to ensure the model sees the data of the projection in the full spatio-angular domain (See Supplementary Note 4). The ROI is calculated as $l_{\text{SLM}} + 2z \tan(\arcsin(\lambda/2p))$, where l_{SLM} and p represent the size and pixel pitch of the SLM, respectively. We use the distributed data-parallel (DDP) of PyTorch for training, with a batch size of 8 on 8 NVIDIA RTX A6000s and a learning rate of $3e^{-4}$.

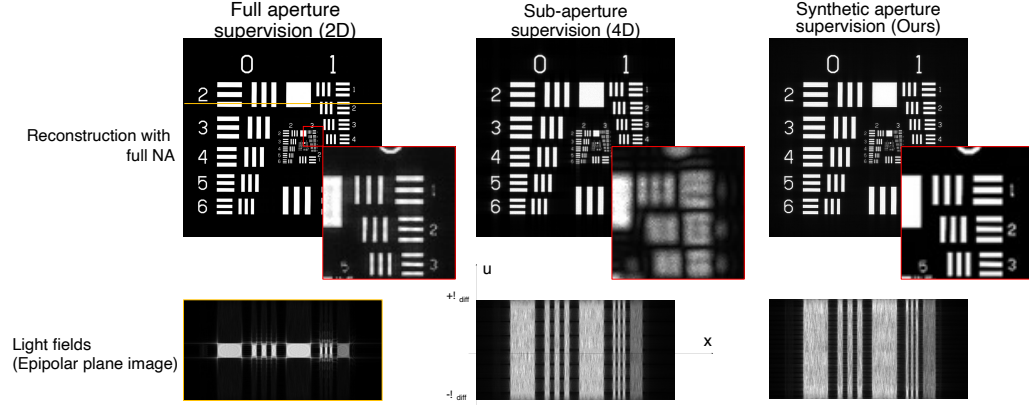
Supplemental Note 7: Additional captured results

In this note, we present additional captured results from our prototypes.

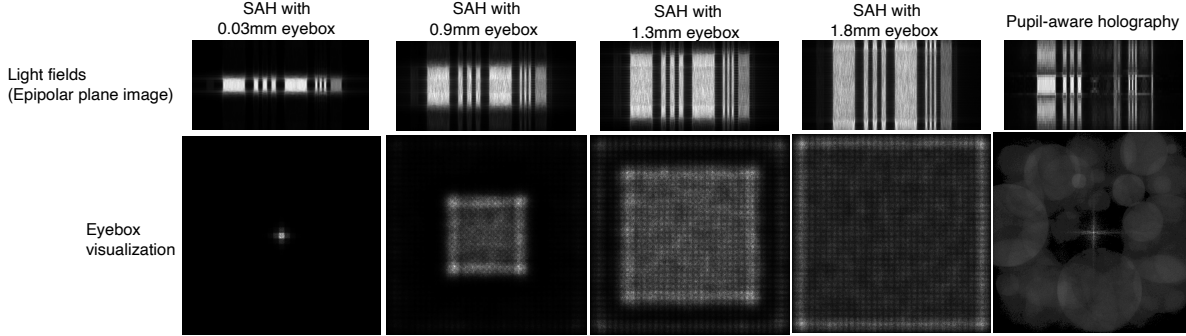
In Supplementary Figure S23, we show experimentally captured results of 3D scenes with varying pupil sizes. Each of the six rows shows one example scene and each of the columns one pupil diameter setting. The two closeups under each image show the changes in depth of field: while the sharpness of a focused object (left) remains constant, an out-of-focus object appears increasingly blurry as the pupil diameter grows (right).

In Supplementary Figures S24–S27, we present additional experimentally captured results of 3D scenes with various focal states and pupil positions. These results demonstrate high image quality with refocusing capabilities and parallax effects as further demonstrated in the supplemental video.

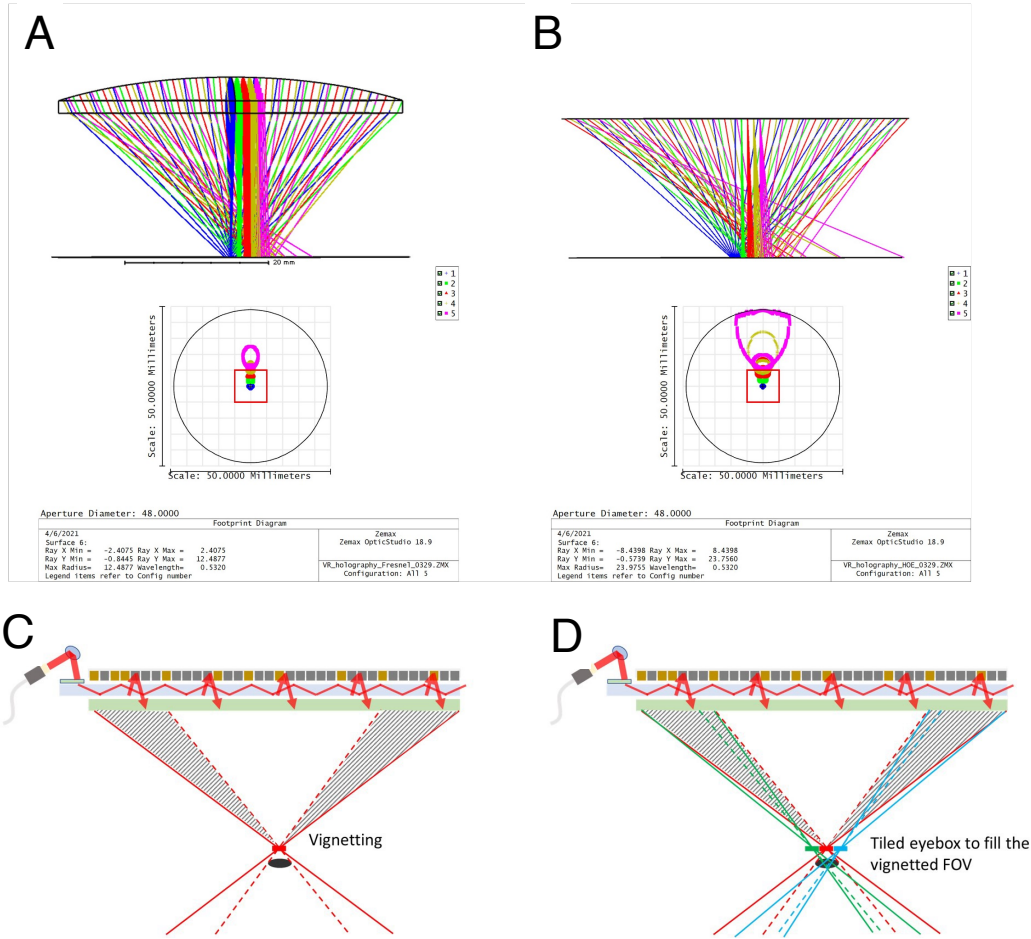
A Comparison of CGH supervisions



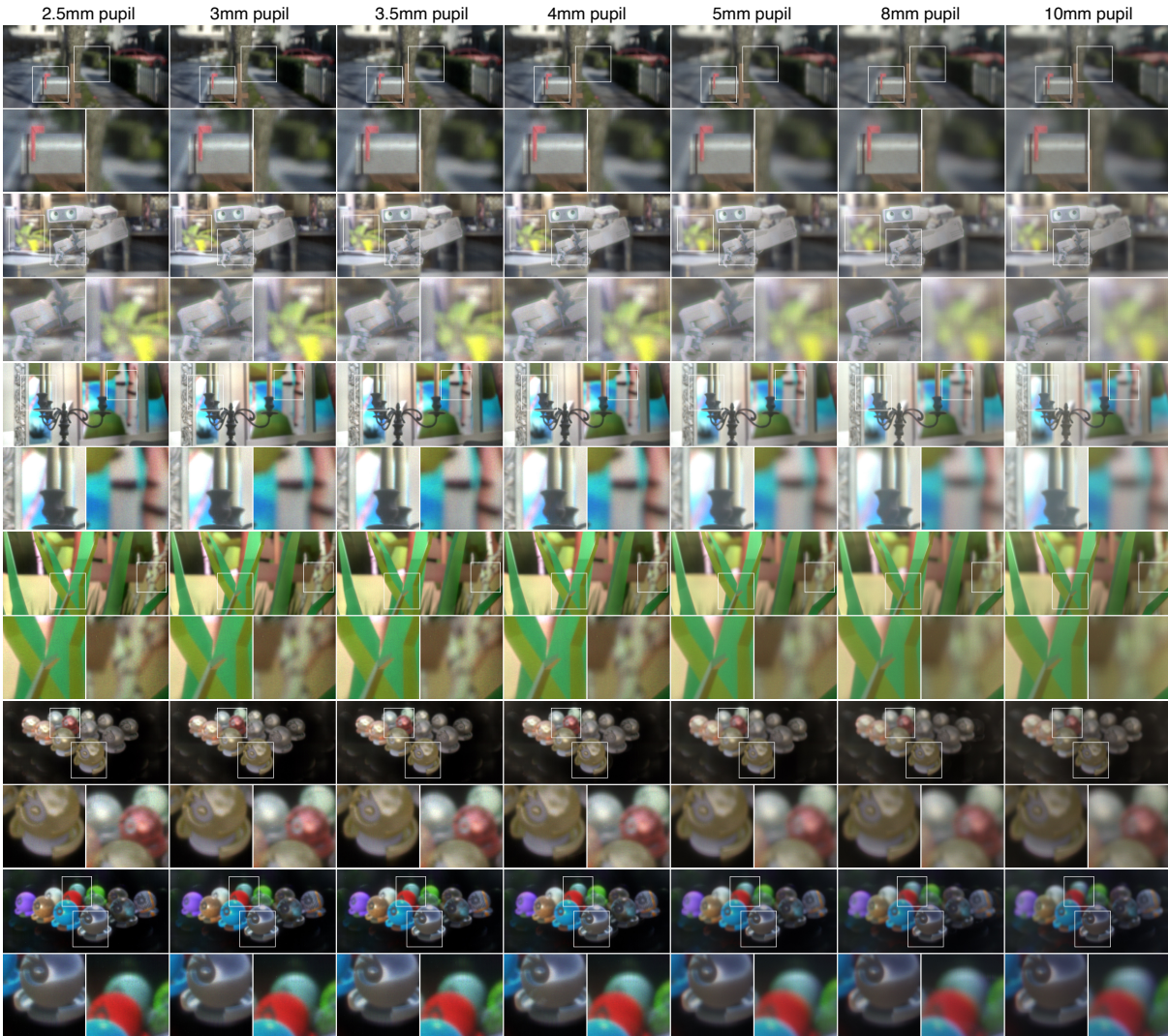
B Static eyebox control



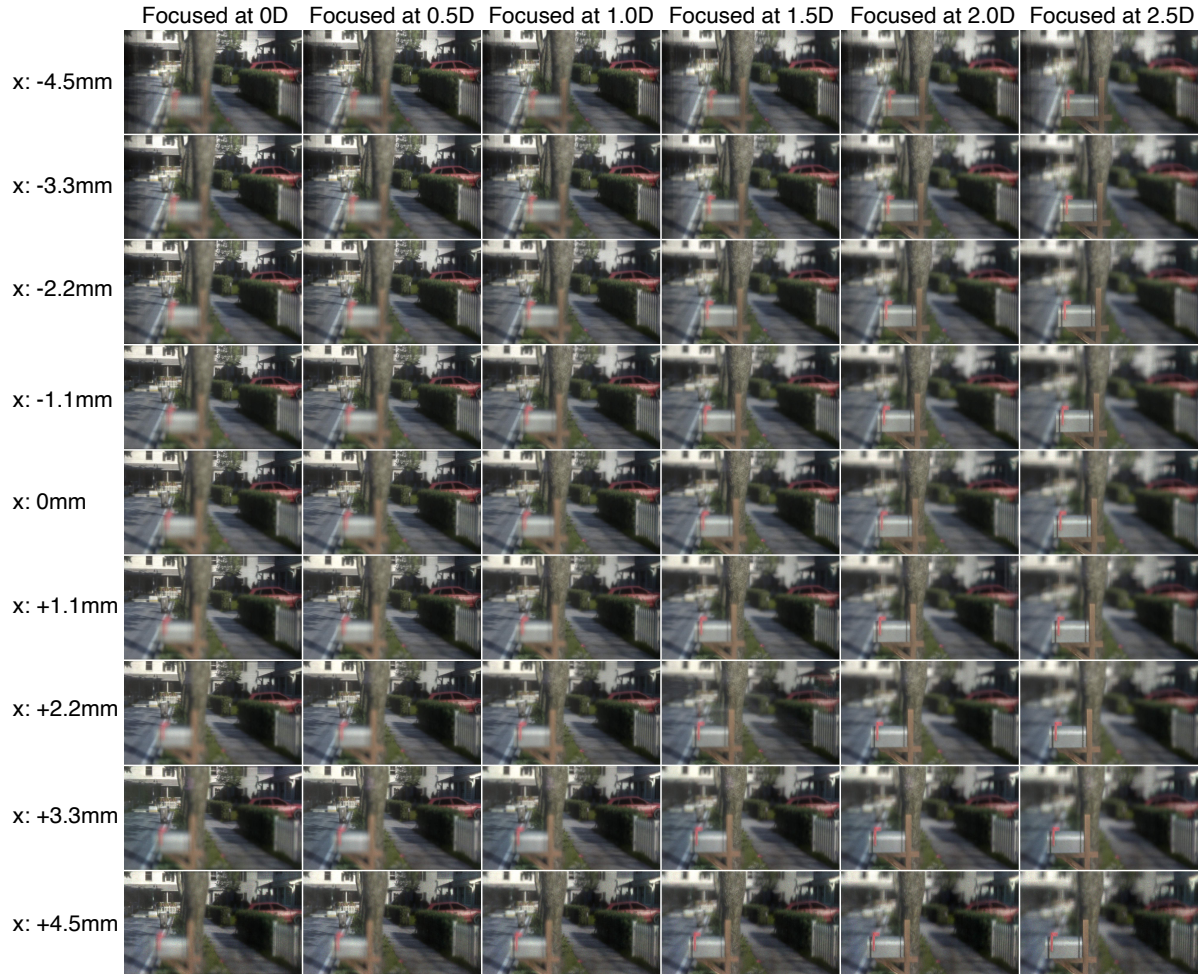
Supplementary Fig. S21 A. Comparison between conventional 2D holography, light-field (4D) holography, and Synthetic Aperture Holography (SAH). In conventional 2D holography, the full wavefront is supervised by target images, and the behavior of the resulting observable light field from the hologram depends on the initialization of the phase or the pre-assumed target phase profile. The light-field hologram supervises the optimization for light-field view images through the sub-apertures, ensuring full light-field reconstruction over the entire NA, but suffering from low resolution when viewed through the whole aperture (*center*). Our synthetic aperture supervision achieves the best of both worlds by considering the synthetic aperture that consists of sub-apertures. **B.** Static eyebox control. Since our method provides a principled way to generate the light field using partially coherent engines, we can control the eyebox distribution by arbitrarily setting the target light fields. Note that this cannot be done using light-field supervision due to the crosstalk between the sub-apertures they do not consider. On the right, we show the results optimized using the stochastic variants of our method, which produces an eyebox shape overfitted to specific pupil shapes and positions.



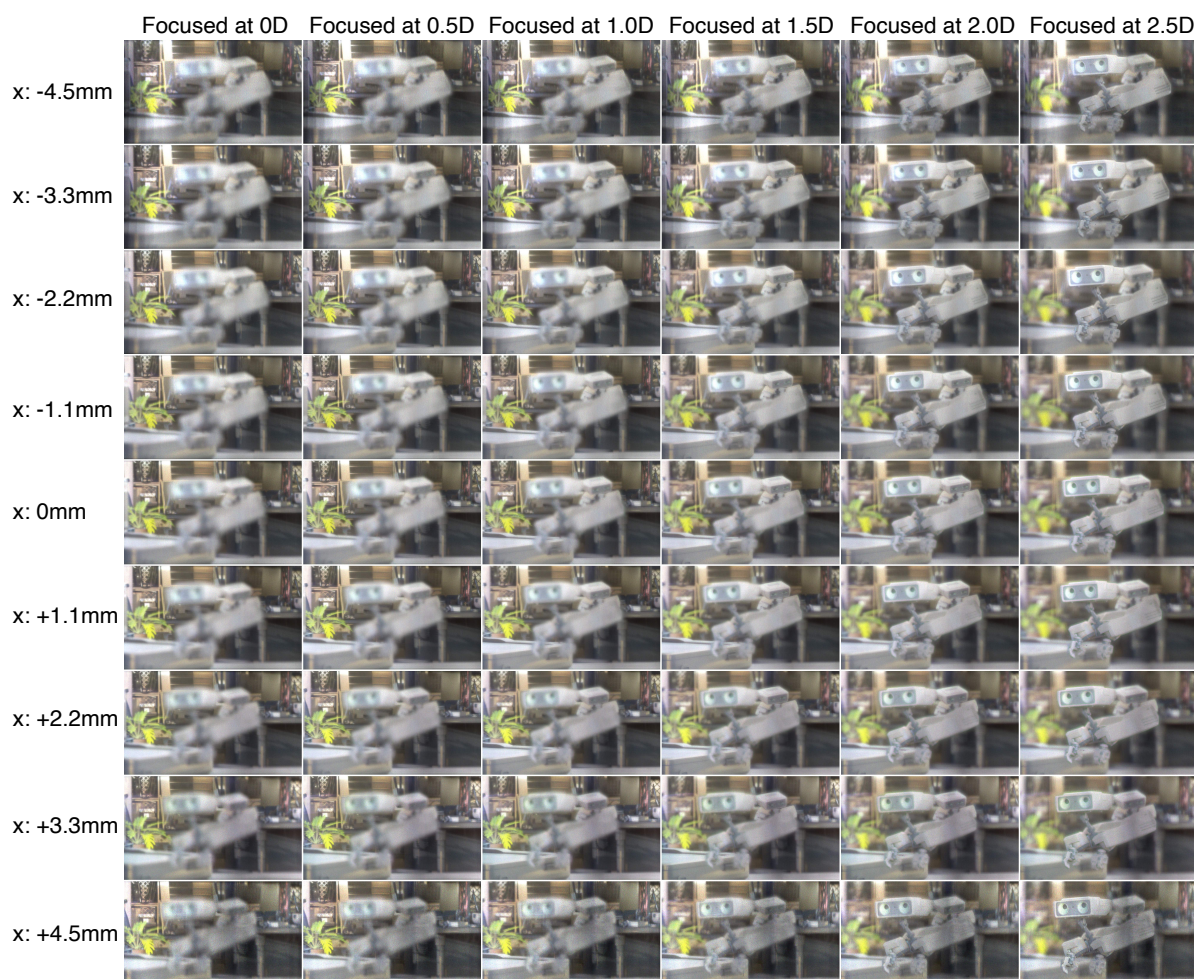
Supplementary Fig. S22 A. Zemax Simulation of conventional diffractive lens that suffers from aberration. **B.** Zemax Simulation of our holographic eyepiece that demonstrates the apochromatic imaging characteristics. **C.** Illustration of eyebox tiling using localized temporal scanning of exit pupils. Vignetting caused by eye-relief mismatch. Red square represents exit pupil and black ellipsoid indicates eye pupil. Striped area indicates vignetted FOV which does not enter the eye pupil. **D.** Locally tiled eyebox illustrated as green and blue color. Vignetted FOV enters the pupil with local pupil scanning.



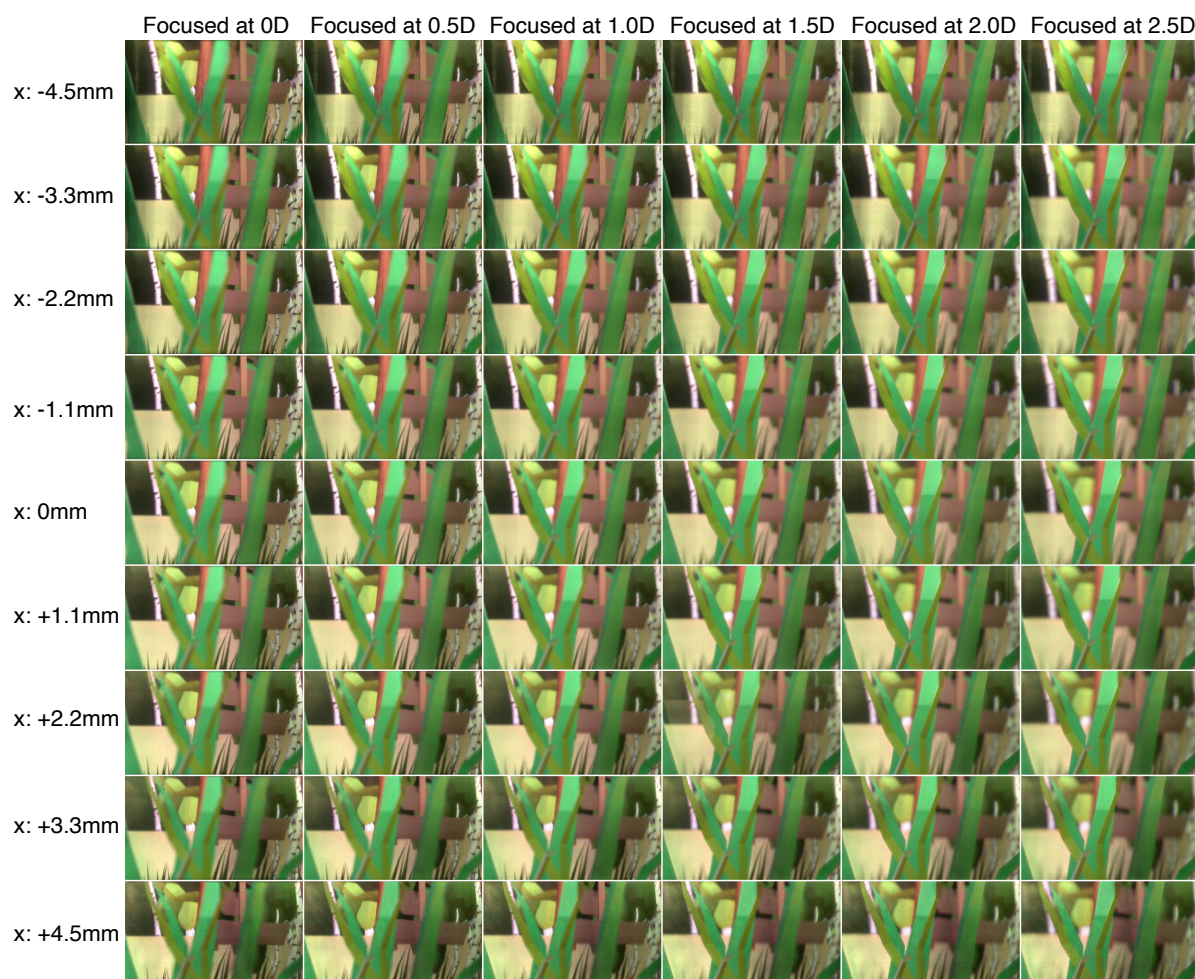
Supplementary Fig. S23 Experimentally captured results of 3D scenes with varying pupil sizes. Each of the six rows shows one example scene and each of the columns one pupil diameter setting. The two closeups under each image show the changes in depth of field: while the sharpness of a focused object (*left*) remains constant, an out-of-focus object appears increasingly blurry as the pupil diameter grows (*right*).



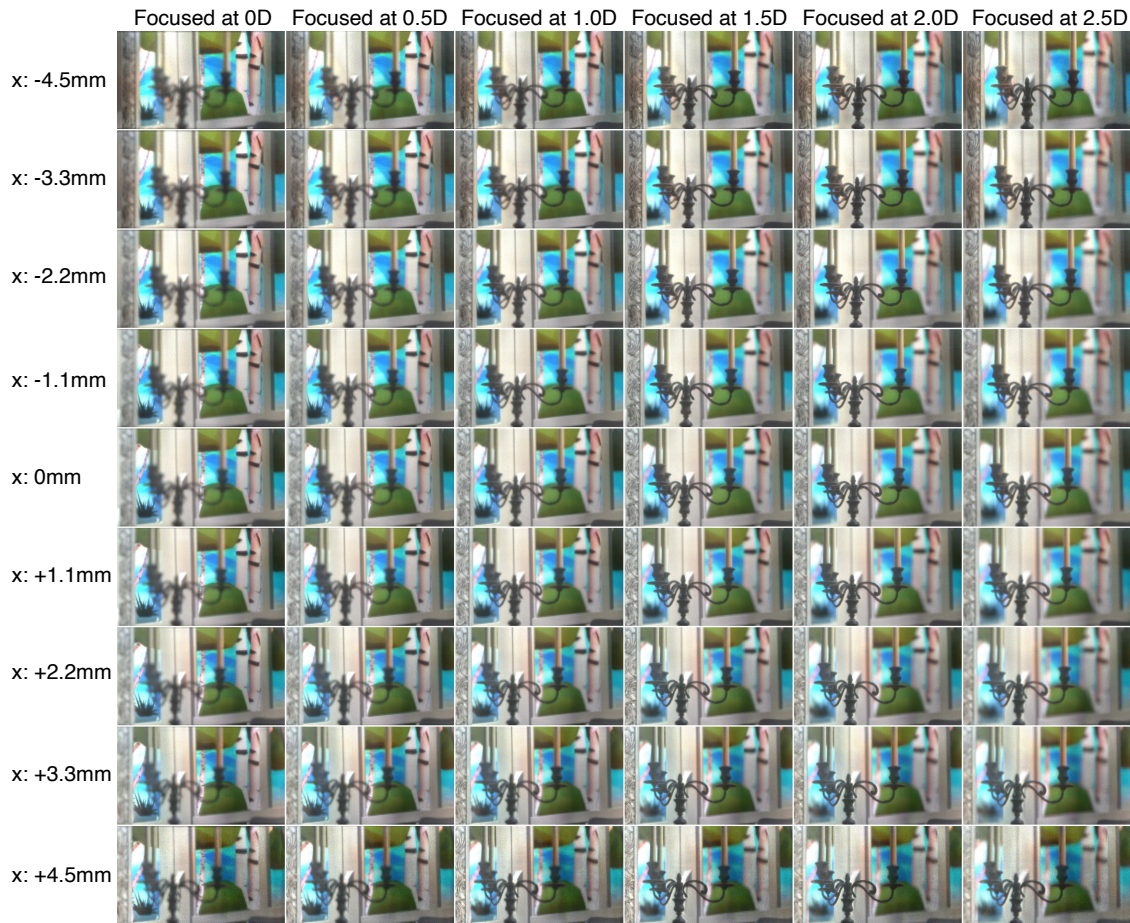
Supplementary Fig. S24 Experimentally captured results of 3D scenes with various focal states and pupil positions. These results demonstrate high image quality with refocusing capabilities and parallax effects as further demonstrated in the supplemental video.



Supplementary Fig. S25 Experimentally captured results of 3D scenes with various focal states and pupil positions. These results demonstrate high image quality with refocusing capabilities and parallax effects as further demonstrated in the supplemental video.



Supplementary Fig. S26 Experimentally captured results of 3D scenes with various focal states and pupil positions. These results demonstrate high image quality with refocusing capabilities and parallax effects as further demonstrated in the supplemental video.



Supplementary Fig. S27 Experimentally captured results of 3D scenes with various focal states and pupil positions. These results demonstrate high image quality with re-focusing capabilities and parallax effects as further demonstrated in the supplemental video.

Supplementary Note 8: Space–bandwidth product, bandwidth, and effective étendue

In this note, we elaborate on the concept of the *effective étendue* in holographic display systems. Since étendue is highly relevant to (or often referred to as) the space–bandwidth product, or bandwidth, we first clarify the definition of terms. Next, we define the *effective étendue* for holographic displays, emphasizing the difference between the étendue of a system and that of a signal. This distinction is established in the context of the spatio-angular domain or phase–space optics.²⁹ We then compare a few étendue expansion approaches within this framework, showing that our architecture allows for an increase in the *effective étendue*, with full representation capability for extended étendue (*effective degrees of freedom*), whereas other approaches, such as mask-based étendue expansion,²⁴ fail to achieve increase the effective degrees of freedom, while it might increase its étendue.

The key takeaway from this section is the distinction between *effective étendue* and *effective degrees of freedom* — the former refers to the amount of observable light field through the optical system, potentially within a certain time period, while the latter refers to the system’s capability to express arbitrary light fields with higher image fidelity, also within a certain time period, for example, one that a human can resolve.

Space–bandwidth product and bandwidth

The optical extent, or space–bandwidth product (SBP), is a measure of the amount of information contained in the scene generated by the display system. It is defined as the product of the area of the spatial region and the two-dimensional bandwidth supported by the optical system.^{14,39} Additionally, it is beneficial to consider the bandwidth (BW), or the SBP rate, as the eyes receive and average out optical signals over a finite period. The numerical SBP of the optical system and its BW are defined as:⁴⁰

$$|SBP_{\text{system}}| = A_{XY} B_{XY}, \quad (\text{S81})$$

$$BW = |SBP| \times f, \quad (\text{S82})$$

where A_{XY} is the area of the spatial region and B_{XY} is the two-dimensional bandwidth supported by the system, and f is the frame rate of the signals.

From Eq. S81, it is clear that the numerical SBP of the system corresponds to the 4D volume of the spatio-angular space (x, y, u, v) supported by the system. Due to the complexity involved in analyzing and visualizing 4D functions, we focus on the 2D space of the coordinates (x, u) for simplicity, since the extension to 4D is straightforward. The SBP in this 2D space can be understood as the area within that domain, as shown in Supplementary Fig. S28A. This boundary represents the maximum extent of the spatial or angular dimensions that can be modulated (or generated) by the device.³⁹ For instance, in the case of phase-only spatial light

modulators (SLM), a_x denotes the size of the SLM, while b_u relates to the maximum spatial frequency, which is determined by the pixel pitch p of the SLM: $b_u = \lambda/p$. It is well known that the numerical SBP of the system is preserved as the number of controllable pixels N_x .⁴¹ In this case, $|SBP_{\text{system}}| = a_x b_u = \frac{a_x \lambda}{p} = N_x \lambda$.

However, we argue that the SBP of the system is not sufficient to describe holographic display systems in general. To elaborate, we distinguish the SBP of the optical signals from the SBP of the system. The user observes a set of optical signals generated within a period of time, rather than the entire combination of signals that can be generated by the system. Especially, the coherent optical signals do not occupy the entire extent of the SBP at any instantaneous time, mainly due to the rank-1 constraint in its mutual intensity matrix¹² (See Supplementary Note 2). The bandwidth supported by the optical system sets the boundary for the bandwidth of an optical signal.

$$SBP_{\text{signal}} \subset SBP_{\text{system}} \quad (\text{S83})$$

In order to numerically measure the SBP of signals, we need a different metric than Eq. S81 to represent signal characteristics in the spatio-angular domain. While an optical signal can be represented with any space-frequency functions*, we use the Wigner distribution function,^{43,44} also known as *generalized radiance*⁴⁵ in the partially coherent case, which is defined as:

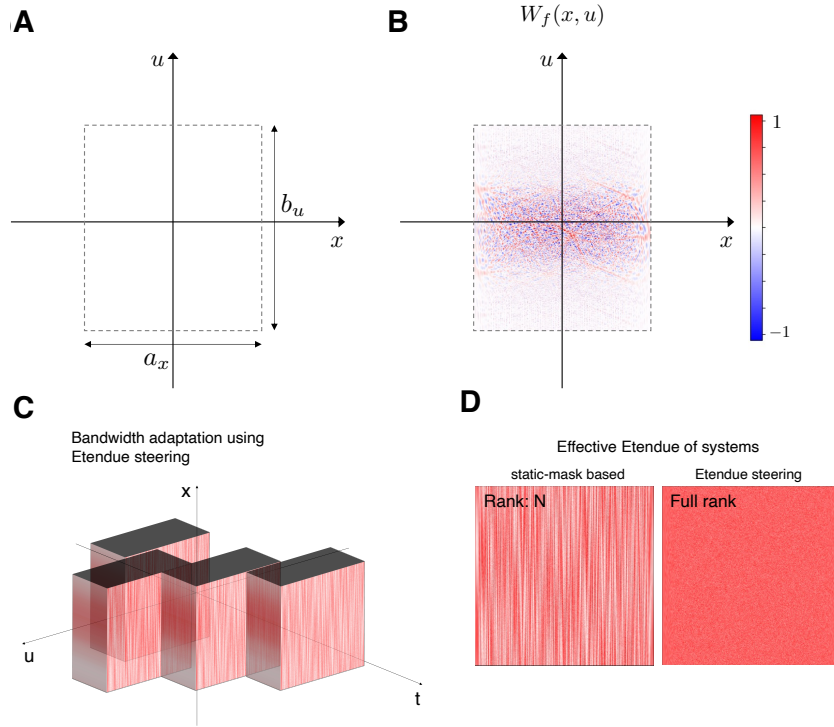
$$W_f(x, u) = \int_{-\infty}^{\infty} \left\langle f\left(x + \frac{x'}{2}\right) f^*\left(x - \frac{x'}{2}\right) \right\rangle e^{-i2\pi x' u} dx', \quad (\text{S84})$$

where x and u are the spatial and angular variables, respectively, and $f(x)$ is the optical signal. This function is interpreted as an energy density function in space and frequency.⁴⁶ Using the Wigner distribution function, we can describe the space-bandwidth product of the complex-valued signal more precisely:

Nonzero area. Following the definition of SBP of the system, we essentially want to measure the area that WDF of the signal takes in the same domain. This is not straightforward, as the WDF of a random signal typically appears as in Fig. S28B. Lohmann et al. attempted to define the SBP of a signal as the area where its WDF is non-zero, while acknowledging that this may not be precise for certain signals.³⁹ One could also use thresholding to measure it.

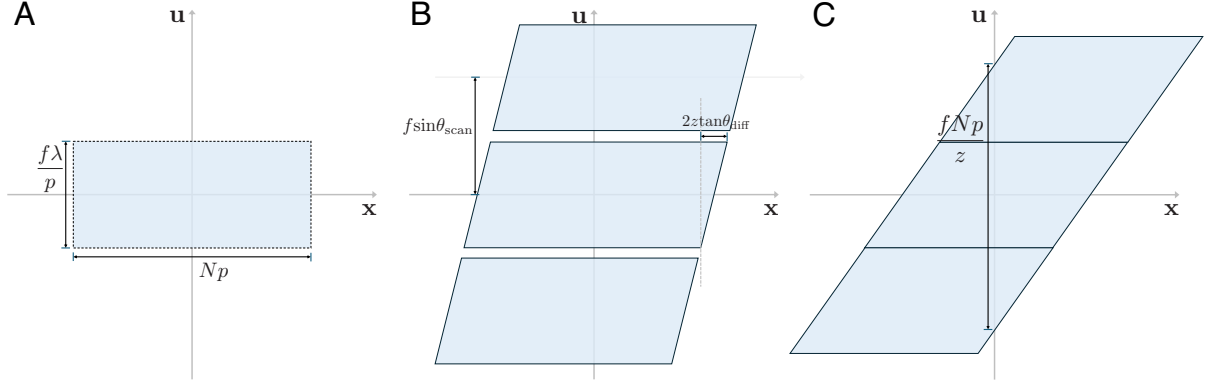
Effective area. Instead of measuring the non-zero area, we could measure the effective area of this representation. The value of the WDF can be interpreted as the amount of a certain space-frequency component contained in the signal,⁴⁷ and the square root of the diagonal terms of its second-order moments $m_{xx,uu}$ is viewed as the effective width $d_{xx,uu} = \sqrt{m_{xx,uu}}$ of the intensity in the x, u direction, respectively. Thus, we argue that we could use the effective area

*the space-frequency functions mostly belong to the Cohen class⁴²



Supplementary Fig. S28 A, B. Visualization of the space–bandwidth product in the spatio-angular domain: **A.** SBP of the system, which sets the boundary, **B.** The SBP of the coherent wavefront (complex-valued signal) can be understood from the signal’s behavior in the spatio-angular domain; here, the color values denote the value of the Wigner distribution function of the complex-valued signal. **C.** Illustration of dynamic allocation of signal in the spatio-angular domain. Backlight (or frontlight) steers the signal in the spatio-angular domain, and the effective étendue can be calculated as the area of the projection of these energies along the time axis over a given time interval. **D.** Time-multiplexed observable light field in the spatio-angular domain (horizontal axis is the spatial domain, the vertical axis is the angular domain) over sufficiently many random wavefronts in static-mask-based system and our system. Here, we simulate an étendue three times larger using a static mask (*left*, with a pixel pitch three times smaller) and a steering operation with time multiplexing (*right*, with a steering angle three times larger than the maximum diffraction angle). In mask-based systems, the signal is always convolved with the Wigner distribution of the mask, resulting in specific patterns that limit the ability to express arbitrary observable light fields, even with sufficient time multiplexing. In synthetic aperture holography, the system occupies the full spatio-angular domain and can express arbitrary light fields if sufficient time multiplexing is available.

of the optical signal in the Wigner domain as the product of the effective widths in spatial and angular domains for near-eye displays, while it has been discussed in time-frequency analysis of time signals.⁴⁸



Supplementary Fig. S29 Visualization of the space–bandwidth product (SBP) of synthetic aperture-based holographic display systems in the spatio-angular domain: **A.** SBP of the basic system, where the unit for the angular axis is converted in meters, based on the focal length of the eyepiece, f . **B.** SBP of the scanned-angle-limited SAH system, where the angular width of a single point in the field of view is limited by the scanned angle range. **C.** SBP of the propagation-limited SAH system, where the angular width is fundamentally limited by the shear operation in the spatio-angular domain.

$$|SBP_{\text{signal}}| = d_{xx}d_{uu} = \sqrt{m_{xx}m_{uu}}, \quad (\text{S85})$$

$$\mathbf{M} = \begin{bmatrix} m_{xx} & m_{xu} \\ m_{ux} & m_{uu} \end{bmatrix} = \frac{1}{E} \iint \begin{bmatrix} x \\ u \end{bmatrix} \begin{bmatrix} x^T & u^T \end{bmatrix} W(x, u) dx du, \quad (\text{S86})$$

$$E = \iint W(x, u) dx du. \quad (\text{S87})$$

Effective étendue

While the *effective area* might provide an intuitive measure, there are corner case scenarios where this approach fails (such as a few points at the edges). This necessitates a more robust measure for the effective étendue.

Given that these signals are ultimately projected onto the human retina and integrated over a certain period of time—during which the human eye may not resolve fine temporal details—in this paper, we define the term *effective étendue* as the area of the observable light field³² in the spatio-angular domain, integrated over a certain period of time.

One important factor to consider is that the relationship between the observable light field and the WDF, or the complex-valued signal, depends on the aperture state p , which determines the resolution across both the spatial and angular axes. Therefore, this *effective étendue* can be measured for a specific aperture size and may be better analyzed across multiple aperture sizes. More formally, the *effective étendue* e for aperture state p over the time period $[t_i, t_f]$ can be

expressed as:

$$e(l, p) = \int \int \int_{t=t_i}^{t=t_f} \mathbf{1}_{l(x,u,t)>c}(x, u, t) dt dx du \quad (\text{S88})$$

$$\mathbf{1}_{l(x,u,t)>c} = \begin{cases} 1 & \text{if } l(x, u, t) > c \\ 0 & \text{otherwise} \end{cases} \quad (\text{S89})$$

where the observable light field l can be obtained using Equation S70. This can be empirically calculated using light fields presented in Supplementary Figure S8. In other words, the *effective étendue* can be understood as the projection of the SBP of signal over time (See Supplementary Figure S28C). Note that the aperture state (size) p determines the resolution of x and u (See Supplementary Figures S18–S19).

From this perspective, steering operations do not increase the instantaneous étendue but instead enhance the *effective étendue*, while a fast spatial light modulator simultaneously increases the effective degrees of freedom, as detailed in Supplementary Note 2. We also highlight the potential of these étendue expansion approaches: with sufficient time multiplexing using an ultrafast spatial light modulator, the étendue steering technique can represent the full-rank mutual intensity. In contrast, static-mask-based approaches are limited by an upper bound in their ability to express arbitrary signals in the spatio-angular domain. Supplementary Figure S28D illustrates the time-multiplexed light field over the number of frames which is the full rank that the system supports. Mask-based expansion is restricted to a specific region of the spatio-angular domain, as it effectively convolves the WDF with a kernel defined by the mask. On the other hand, the étendue steering approach can fully explore and express the entire domain.

Eyebox size

As the steering operation steers the étendue (Eq. S59), each point in the entire field-of-view can only be represented with certain steering angles. We find that this fundamentally limits the eyebox, or equivalently the blur size, as stated in Equation 6 in the main manuscript. We visualize the two cases that the eyebox is limited in Supplementary Figure S29. In Supp. Fig. S29A, we visualize and label the size of the SBP in terms of eyebox size (with the focal length of the eyepiece f) and the panel size. The eyebox size is the length along the angular axis, i.e., $\frac{f\lambda}{p}$. As the system aims to generate images around the propagation distance z with steering and time multiplexing, it supports a larger effective étendue, as shown in Supp. Fig. S29B. This scenario assumes that the eyebox is limited by the scanned angle θ_{scan} , and the eyebox size is given by $\frac{f\lambda}{p} + 2f \sin \theta_{\text{scan}}$. If the system supports a wider steering angle or larger propagation distances, it would steer the field more in the spatial domain (see Supplementary Figure S12). Here, the eyebox size is limited by the wave propagation itself, and the size can be calculated as $\frac{L_{\text{SLM}} f \sin \theta_{\text{diff}}}{|z| \tan \theta_{\text{diff}}} \simeq \frac{L_{\text{SLM}} f}{|z|}$, where θ_{diff} is the maximum diffraction angle of the SLM. The approximation holds for most commercially available SLMs.

Supplementary References

- ¹ Jang, C., Bang, K., Chae, M., Lee, B. & Lanman, D. Waveguide holography for 3d augmented reality glasses. *Nature Communications* **15** (2024).
- ² Levola, T. Diffractive optics for virtual reality displays. *Journal of the Society for Information Display* **14**, 467–475 (2006).
- ³ Kogelnik, H. Coupled wave theory for thick hologram gratings. *Bell System Technical Journal* **48**, 2909–2947 (1969).
- ⁴ Yariv, A. & Yeh, P. Optical waves in crystal propagation and control of laser radiation (1983).
- ⁵ Kress, B. C. & Pace, M. Holographic optics in planar optical systems for next generation small form factor mixed reality headsets (2022).
- ⁶ Bruder, F.-K., Fäcke, T. & Rölle, T. The chemistry and physics of bayfol® hx film holographic photopolymer. *Polymers* **9** (2017). URL <https://www.mdpi.com/2073-4360/9/10/472>.
- ⁷ Escuti, M. J., Oh, C., Sánchez, C., Bastiaansen, C. & Broer, D. J. Simplified spectropolarimetry using reactive mesogen polarization gratings. In Shen, S. S. & Lewis, P. E. (eds.) *Imaging Spectrometry XI*, vol. 6302, 630207. International Society for Optics and Photonics (SPIE, 2006). URL <https://doi.org/10.1117/12.681447>.
- ⁸ Li, L., Kim, J., Shi, S. & Escuti, M. J. Color-selective geometric phase lens for apochromatic lens system. In Khoo, I. C. (ed.) *Liquid Crystals XXIV*, vol. 11472, 1147219. International Society for Optics and Photonics (SPIE, 2020). URL <https://doi.org/10.1117/12.2569165>.
- ⁹ Li, L., Shi, S., Kim, J. & Escuti, M. J. Color-selective geometric-phase lenses for focusing and imaging based on liquid crystal polymer films. *Opt. Express* **30**, 2487–2502 (2022). URL <https://opg.optica.org/oe/abstract.cfm?URI=oe-30-2-2487>.
- ¹⁰ Choi, S. *et al.* Time-multiplexed neural holography: A flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators. In *Proceedings of the ACM SIGGRAPH*, 1–8 (2022).
- ¹¹ Commission Internationale de l’Eclairage (CIE). CIE Standard Illuminant D65 (n.d.). URL <https://cie.co.at/datatable/cie-standard-illuminant-d65>. Accessed: [insert date of access].
- ¹² Ozaktas, H. M., Yüksel, S. & Kutay, M. A. Linear algebraic theory of partial coherence: discrete fields and measures of partial coherence. *JOSA A* **19**, 1563–1571 (2002).
- ¹³ Matsushima, K. & Shimobaba, T. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Optics express* **17**, 19662–19673 (2009).

- ¹⁴ Goodman, J. W. *Introduction to Fourier optics* (Roberts and Company, 2005).
- ¹⁵ Peng, Y., Choi, S., Padmanaban, N. & Wetzstein, G. Neural holography with camera-in-the-loop training. *ACM Transactions on Graphics (TOG)* **39**, 1–14 (2020).
- ¹⁶ Choi, S., Gopakumar, M., Peng, Y., Kim, J. & Wetzstein, G. Neural 3d holography: Learning accurate wave propagation models for 3d holographic virtual and augmented reality displays. *ACM Trans. Graph. (SIGGRAPH Asia)* (2021). Accepted for publication (*denotes equal contribution).
- ¹⁷ Peng, Y., Choi, S., Kim, J. & Wetzstein, G. Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration. *Science Advances* (2021).
- ¹⁸ Gopakumar, M. *et al.* Full-colour 3D holographic augmented-reality displays with metasurface waveguides. *Nature* (2024).
- ¹⁹ Wolf, E. New theory of partial coherence in the space–frequency domain. part i: spectra and cross spectra of steady-state sources. *JOSA* **72**, 343–351 (1982).
- ²⁰ Gamo, H. Iii matrix treatment of partial coherence. In *Progress in Optics*, vol. 3, 187–332 (Elsevier, 1964).
- ²¹ Goodman, J. W. *Statistical optics* (John Wiley & Sons, 2015).
- ²² Bastiaans, M. J. Application of the wigner distribution function to partially coherent light. *JOSA A* **3**, 1227–1238 (1986).
- ²³ Zhang, Z. *Analysis and synthesis of three-dimensional illumination using partial coherence* (Stanford University, 2011).
- ²⁴ Kuo, G., Waller, L., Ng, R. & Maimone, A. High resolution étendue expansion for holographic displays. *ACM Transactions on Graphics (TOG)* **39**, 66–1 (2020).
- ²⁵ Gopakumar, M., Kim, J., Choi, S., Peng, Y. & Wetzstein, G. Unfiltered holography: optimizing high diffraction orders without optical filtering for compact holographic displays. *Optics Letters* **46**, 5822–5825 (2021).
- ²⁶ Hamann, S., Shi, L., Solgaard, O. & Wetzstein, G. Time-multiplexed light field synthesis via factored wigner distribution function. *Opt. Lett.* **43**, 599–602 (2018).
- ²⁷ Müller, T., Evans, A., Schied, C. & Keller, A. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)* **41**, 1–15 (2022).
- ²⁸ Choi, S., Gopakumar, M., Peng, Y., Kim, J. & Wetzstein, G. Neural 3d holography: Learning accurate wave propagation models for 3d holographic virtual and augmented reality displays. *ACM Trans. Graph. (SIGGRAPH Asia)* (2021).

- ²⁹ Testorf, M., Hennelly, B. & Ojeda-Castañeda, J. *Phase-space optics: fundamentals and applications* (McGraw-Hill Education, 2010).
- ³⁰ Matsushima, K. Shifted angular spectrum method for off-axis numerical propagation. *Optics Express* **18**, 18453–18463 (2010).
- ³¹ Ng, R. *et al.* *Light field photography with a hand-held plenoptic camera*. Ph.D. thesis, Stanford university (2005).
- ³² Zhang, Z. & Levoy, M. Wigner distributions and how they relate to the light field. In *Proc. ICCP*, 1–10 (IEEE, 2009).
- ³³ Nolte, N., Kitouni, O. & Williams, M. Expressive monotonic neural networks. In *The Eleventh International Conference on Learning Representations* (2022).
- ³⁴ Bao, C., Barbastathis, G., Ji, H., Shen, Z. & Zhang, Z. Coherence retrieval using trace regularization. *SIAM Journal on Imaging Sciences* **11**, 679–706 (2018).
- ³⁵ McAlister, D., Beck, M., Clarke, L., Mayer, A. & Raymer, M. Optical phase retrieval by phase-space tomography and fractional-order fourier transforms. *Optics letters* **20**, 1181–1183 (1995).
- ³⁶ Padmanaban, N., Peng, Y. & Wetzstein, G. Holographic near-eye displays based on overlap-add stereograms. *ACM Transactions on Graphics (TOG)* **38**, 1–13 (2019).
- ³⁷ Chakravarthula, Praneeth and Baek, Seung-Hwan and Schiffers, Florian and Tseng, Ethan and Kuo, Grace and Maimone, Andrew and Matsuda, Nathan and Cossairt, Oliver and Lanman, Douglas and Heide, Felix. Pupil-Aware Holography. *ACM Transactions on Graphics (TOG)* **41**, 1–15 (2022).
- ³⁸ Schiffers, F. *et al.* Stochastic light field holography. In *2023 IEEE International Conference on Computational Photography (ICCP)*, 1–12 (IEEE, 2023).
- ³⁹ Lohmann, A. W., Dorsch, R. G., Mendlovic, D., Zalevsky, Z. & Ferreira, C. Space–bandwidth product of optical signals and systems. *JOSA A* **13**, 470–473 (1996).
- ⁴⁰ Li, J., Smithwick, Q. & Chu, D. Holobricks: modular coarse integral holographic displays. *Light: Science & Applications* **11**, 1–15 (2022).
- ⁴¹ Born, M. & Wolf, E. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).
- ⁴² Cohen, L. Time-frequency distributions-a review. *Proceedings of the IEEE* **77**, 941–981 (1989).

- ⁴³ Wigner, E. P. On the quantum correction for thermodynamic equilibrium. In *Part I: Physical Chemistry. Part II: Solid State Physics*, 110–120 (Springer, 1997).
- ⁴⁴ Bastiaans, M. J. The wigner distribution function applied to optical signals and systems. *Optics communications* **25**, 26–30 (1978).
- ⁴⁵ Walther, A. Radiometry and coherence. *JOSA* **58**, 1256–1259 (1968).
- ⁴⁶ De Bruijn, N. Uncertainty principles in fourier analysis. *Inequalities* **2**, 57–71 (1967).
- ⁴⁷ Xiao, J., Zhang, W. & Zhang, H. Inverse diffraction in phase space. *JOSA A* **40**, 175–184 (2023).
- ⁴⁸ Durak, L. & Arikan, O. Short-time fourier transform: two fundamental properties and an optimal implementation. *IEEE Transactions on Signal Processing* **51**, 1231–1242 (2003).