Holographic display étendue expansion with a binary π-metasurface: supplemental document

This document provides additional detail for the primary manuscript: the component list, the measured phase look-up tables of the phase light modulator, the metasurface specifications, the simulation model and its parameters, and extended simulation and experimental results.

1. OPTICAL SYSTEM AND COMPONENT LIST

The optical setup implements a single passive expansion stage; the beam path is that of Fig. 1(a) of the primary manuscript, with the phase light modulator (PLM) imaged onto the metasurface at unit magnification so that each PLM mirror registers one-to-one with a metasurface supercell. Table S1 lists every component with vendor, part number and role.

Table S1. Component list.

Element	Vendor	Part number	Key specification and role
Illumination	FISBA	READYBeam Ind1	Fiber-coupled RGB laser; 450, 520 and 660 nm; collimated at the PLM
Modulator	Texas Instruments	XDLPM670E1FYZ	0.67-in. 4-bit piston-mode MEMS phase light modulator (PLM; DLP6750Q1EVM evaluation module); 10.8 μm phase-cell pitch
Controller	Texas Instruments	DLPC900 EVM	Drives the PLM; Video Pattern Mode locked to the HDMI source (firmware 6.10.178)
Expander	SNOChip Inc.	–	Silicon-nitride binary π-metasurface on fused silica
Beamsplitter	Thorlabs	BP250	Couples illumination in and the modulated return out
Relay L_1, L_2	Thorlabs	AC508-75-A-ML ($\times 2$)	Achromatic doublets, $f = 75$ mm; 1:1 relay at unit magnification
Fourier DC block	Front Range Photomask	350 μm dot	DC stop (350 μm-diameter opaque dot) and iris at the relay Fourier plane; blocks the PLM zero order
Fourier lens	Thorlabs	AC508-150-A-ML	$f = 150$ mm; forms the far field on the sensor
Camera	FLIR	BFS-U3-200S6M	Monochrome; 2.4 μm pixel pitch

A. Camera-plane sampling and calibration

For a hologram of N_j samples of pitch p_j along axis j, a Fourier lens of focal length f gives the ideal scalar sampling

$$\delta u_j = \frac{\lambda f}{N_j p_j}. \tag{S1}$$

*Author-prepared version of the supplemental document for: C. Wang, Z. Zhou, E. Tseng, V. Chu, W. T. Chen, J. E. Fröch, A. Majumdar, and F. Heide, "Holographic display étendue expansion with a binary π-metasurface," Opt. Lett. **51**(19) (2026), https://doi.org/10.1364/OL.613712.*

A camera-plane calibration measures a scale of 0.66 times the value predicted by Eq. (S1). This factor is an empirical calibration result, which we attribute to relay clipping and the finite pupil of the relay optics, and it is not an analytic magnification. We apply it whenever simulation and capture are compared.

B. Phase light modulator

The modulator is a Texas Instruments 0.67-in. piston-mode MEMS phase light modulator, device XDLPM670E1FYZ, operated as part of the DLP6750Q1EVM evaluation module, in which each mirror translates axially to impart phase, driven by a DLPC900 EVM controller. In phase operation the array presents 1358×800 cells on a $10.8\,\mu$m pitch, of which 1280×800 are addressed in the HDMI drive mode used here. The controller runs in Video Pattern Mode locked to the external HDMI source (DLPC900 firmware 6.10.178): each 60 Hz video frame is decomposed into 24 one-bit planes, eight per color channel, displayed for $694\,\mu$s each, which is the hardware origin of the eight subframes per color available for temporal multiplexing when all three colors are driven; a sequential single-laser capture with color synchronization off can assign all 24 planes to one laser. Each frame's 24 planes amount to $24 \times 60 = 1440$ phase patterns per second, which is what makes full-color operation at $K = 8$ per color a video-rate mode on this hardware. Raising every color to the $K = 24$ of the Fig. 3 captures would need $24 \times 3 \times 60 = 4320$ patterns per second, three times the streaming interface rate; conversely, at 1440 patterns per second those 72 subframes per frame would reduce the video rate to 20 Hz. The measured speckle contrast (Sec. 5) saturates by $K = 8$, so the display does not need to go beyond it. The illuminated aperture is therefore $1280 \times 10.8\,\mu\text{m} = 13.824$ mm by $800 \times 10.8\,\mu\text{m} = 8.64$ mm. Phase is quantized to 4 bits, that is 16 displacement levels.

All three primaries are driven from a single fixed mirror-bias voltage. The entire chromatic calibration therefore lives in three measured 16-entry phase look-up tables (LUTs), plotted in Fig. S1. Because a fixed piston displacement produces a wavelength-dependent phase, the achievable modulation depth falls with increasing wavelength: 7.54, 6.22 and 4.37 rad at 450, 520 and 660 nm. The red channel therefore does not reach a full 2π, which is an acknowledged limitation of operating from one fixed bias.

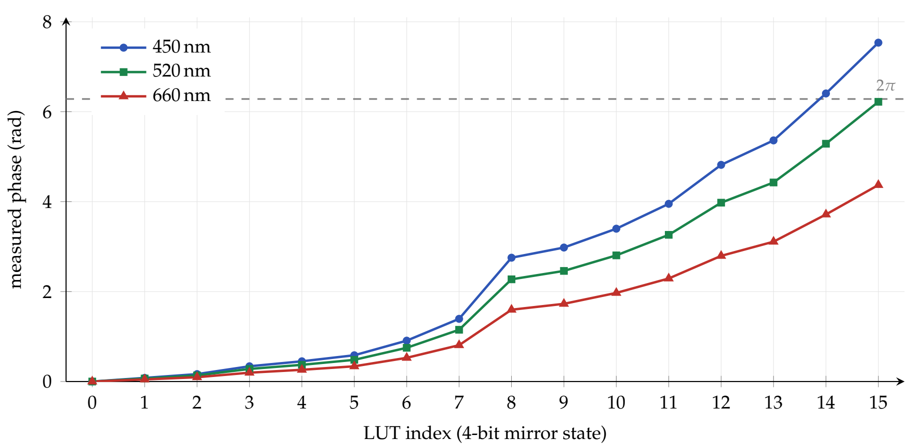

Fig. S1. Measured 16-entry PLM phase look-up tables at the three primaries, from a single fixed mirror-bias voltage.

2. BINARY π-METASURFACE EXPANDER

The expander is a passive silicon-nitride metasurface patterned on a 0.5 mm fused-silica substrate, with a 13.8 mm $\times$ 8.6 mm active area and a $10.8\,\mu$m supercell pitch that registers one-to-one with the PLM mirrors. The pattern is a balanced binary mask, with a measured fill fraction of approximately 0.496, imposing a phase of 0 or π.

Each supercell consists of SiN nanopillars of either 191 or 336 nm diameter, 640 nm tall, on a 386 nm center-to-center lattice (28 pillars per $10.8\,\mu$m supercell); scanning electron micrographs of a representative region are given in the primary manuscript [its Fig. 1(b)]. The nanopillars were fabricated by depositing a 640-nm-thick SiN film onto the fused-silica substrate and spin-coating a 200-nm-thick electron-beam resist; the resist was patterned by electron-beam lithography (EBPG 5150+, Raith Inc.) and developed, an aluminum layer was deposited and lifted off to form a hard mask, the SiN was etched in an inductively coupled plasma etcher, and the hard mask was removed in Al Etchant Type A (Transene Inc.) for 10 minutes. The two pillar populations are dispersion-engineered so that their complex transmissions $t_A(\lambda)$ (191 nm) and $t_B(\lambda)$ (336 nm) maintain a π phase difference across the visible [Fig. 1(c) of the primary manuscript]. In full-wave design simulation, the phase difference stays within 13° of π on average over 400–700 nm (7°, 17° and 19° at 450, 520 and 660 nm), departing near resonances of the 336 nm pillar at 560–595 nm, where its transmission amplitude dips to 0.33. The circularly symmetric pillars make the response polarization-insensitive.

In the same full-wave simulation the grating zero order stays below 3% at all three primaries (Table S2), the ±1 orders carry approximately 42% each (an ideal binary-π grating diffracts $4/\pi^2 \approx 40.5\%$ into each first order), and the total transmission is 82, 79 and 93%. The order efficiencies are diffraction efficiencies of the transmitted light: each order's power divided by the total transmitted power of the same simulation. This is the contrast-relevant convention, since the un-expanded ghost and the expanded image share the same throughput; multiplying by the transmission column gives the incident-power fractions (zero order 0.22%/1.04%/2.56% at the three primaries).

Table S2. Unscattered fraction, simulated and measured, and simulated reconstruction quality. The $s = 4$ contrast rows restrict the background to the un-expanded ghost zone (see text).

Quantity	450 nm	520 nm	660 nm
Resin index $n(\lambda)$, Cauchy fit	1.5223	1.5157	1.5081
DOE $\|c_0\|^2$, π at 520 nm	7.0%	0.0%	11.9%
DOE $\|c_0\|^2$, π at 660 nm	51.2%	19.1%	0.0%
Metasurface $\|c_0\|^2$ (full-wave sim.)	0.27%	1.31%	2.75%
Metasurface $\|c_0\|^2$ (measured, via sim. transmission)	9.1%	7.5%	3.5%
Image contrast at $s = 2$, metasurface (fabricated, sim.)	11.8:1	18.4:1	20.2:1
Image contrast at $s = 2$, DOE (π at 520 nm)	9.5:1	16.0:1	11.9:1
Image contrast at $s = 4$, metasurface (fabricated, sim.)	2.5:1	3.5:1	3.0:1
Image contrast at $s = 4$, DOE (π at 520 nm)	1.7:1	3.3:1	1.3:1
Image contrast at $s = 4$, DOE (π at 660 nm)	0.7:1	1.2:1	2.6:1

A. Measured unscattered fraction

We measured the unscattered fraction of the fabricated element on the setup used for the captures. The modulator is driven flat, the Fourier-plane DC stop is retracted, and an iris stops the beam so that it underfills the patterned area with margin; the sensor is defocused so the unscattered order arrives as a resolved image of the pupil rather than a saturating point, and we record the on-axis energy with the element in the beam and again with it removed. Because the reference frame carries all of the light, that ratio is the unscattered fraction referenced to the incident power, and no integration of the diffracted halo is required, which matters because the halo does not fit on the sensor. Attenuators sit in the common path ahead of the modulator and divide out. The measured incident-power zero order is 7.5, 5.9 and 3.3% at 450, 520 and 660 nm, against the 0.22, 1.04 and 2.56% the full-wave zero order of Sec. 2 implies at the same simulated transmission; dividing the measurement by that transmission gives the $|c_0|^2$ of Table S2. Three sessions in two optical configurations agree to within 15%, and stopping the beam to a third of its width, so that no light can pass beside the pattern, leaves the result unchanged. A few pixels at the core of the reference frames reach the sensor ceiling, five at 450 and twelve at 660 nm, which would bias the ratio upward; a separate acquisition of the same configuration, re-estimated with a matched-point-spread-function fit that masks the saturated core and extrapolates through it, reproduces the aperture photometry to within 5% at all three primaries.

Read through $|c_0|^2 = (1-2p)^2 + p(1-p)\delta^2$, the numbers imply a phase step departing from π by 35, 31 and 22° against the 7, 17 and 19° the designed meta-atom phases give directly, the residuals Eq. (S4) is evaluated with. A fabricated element can only depart further from π than its nominal geometry, so the direction is expected; the size is not, and we report it without attributing its magnitude. Two candidates are excluded. The comparison is against the full-wave grating simulation of Sec. 2, which resolves the supercell and the 386 nm pillar lattice, so the excess is not a shortcoming of the thin-element picture of Eq. (S4); and it is not a uniform pillar-size bias, since a rigid shift of the designed spectral response cannot reproduce the measurement at all: the designed zero order rises with wavelength where the measured one falls, no shift within ±150 nm reproduces that ordering, and the closest fit still misses by 5.7 percentage points. Two observations bound what remains. The designed phase difference itself departs from π by between 0.3 and 80.5° across 400–700 nm, so tens of degrees is the scale on which this dispersion-engineered pair responds to geometry, and the extra 28, 14 and 3° implied here is within ordinary diameter and sidewall tolerance. The iris also samples about 4 × 2.4 mm of the 13.8 × 8.6 mm patterned area, so an error that varies across the die is not averaged here as it is in the full-aperture captures. The comparison with the fixed-depth baseline narrows accordingly, and we state it as measured: the element no longer improves on a 520 nm-designed DOE at every primary, since at 450 nm the measured 9.1% exceeds that element's 7.0% and the DOE nulls exactly at its design wavelength. What survives is the achromatic bound that motivates the design, the measured worst case of 9.1% staying below the 11.9% of the 520 nm design and far below the 51.2% of a 660 nm one.

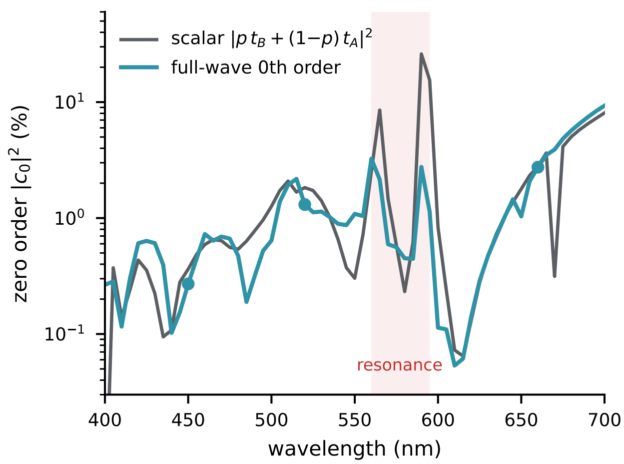

Fig. S2. Zero order of the meta-atom pair. The scalar thin-element estimate Eq. (S4) evaluated with the supplied t_A, t_B against the full-wave result; the two agree away from the 560–595 nm resonance band.

3. SIMULATION MODEL

A. Forward model

Let ϕ_P be the modulator phase on a coarse $(R/s) \times (C/s)$ superpixel grid, where the integer étendue expansion factor s groups $s \times s$ native cells into one superpixel, and let $\phi_E(\mathbf{x}) = \pi m(\mathbf{x})$, $m \in \{0, 1\}$, be the fixed expander phase on the native $R \times C$ grid. The operator U_s expands each modulator sample into a constant $s \times s$ block; the expander mask is never resized. Assuming coherent, unit-amplitude illumination and a single Fraunhofer propagation, the far-field intensity is

$$I(\mathbf{u}) = \left|\mathcal{F}\left\{ \exp\left(i\left[U_s\phi_P(\mathbf{x}) + \phi_E(\mathbf{x}) \right]\right)\right\}\right|^2. \tag{S2}$$

We optimize ϕ_P against the power-normalized target by minimizing mean-squared error with Adam [1]. This is plain first-order optimization of a differentiable far-field model; it uses neither Gerchberg–Saxton alternation nor a learned image-formation model.

B. Modulator model: continuous against quantized

Two distinct modulator models appear in the manuscript, and the distinction matters when reproducing the figures. For the expander comparison (Fig. 2 of the primary manuscript) we use the continuous optimized phase, which models an ideal phase-only spatial light modulator (SLM), so that the difference between the diffractive optical element (DOE) and the metasurface is the expander's chromatic behavior alone and is not confounded by modulator quantization. For the experimental comparison (Fig. 3) the phase is instead quantized by circular phase distance to the 16 experimentally-measured LUT entries of Fig. S1 for the active wavelength, so that the simulation propagates the phase states the hardware actually displays.

This choice does not drive the reported DOE-versus-metasurface difference: the ghost fractions are analytic [Eq. (S4)] and independent of how the modulator phase is quantized, so the Fig. 2 comparison, which uses the continuous optimized phase, and the experimental comparison, which uses the LUT-quantized phase, isolate the same chromatic effect.

C. Chromatic model of the reference DOE

The DOE baseline is a fixed-depth surface-relief element whose phase is set by the material dispersion,

$$\phi_{\mathrm{DOE}}(\lambda) = \frac{2\pi\,[\,n(\lambda)-1\,]\,h}{\lambda}, \qquad h = \frac{\lambda_0}{2\,[\,n(\lambda_0)-1\,]}, \tag{S3}$$

with the etch depth h fixed to give exactly π at the design wavelength λ_0. We use $\lambda_0 = 520\,\mathrm{nm}$, which is the fair baseline against a metasurface targeting the same primary. We take $n(\lambda)$ from the stamping-resin indices measured by Tseng *et al.* [2], namely $n = 1.5223$, 1.5159 and 1.5081 at 450, 517 and 660 nm, fitted with a second-order Cauchy model in $1/\lambda^2$ that reproduces the three measured points exactly.

D. Unscattered fraction

For a thin binary mask of fill fraction p whose two states have complex transmissions t_A and t_B, the unscattered (zero-order) fraction is the squared magnitude of the spatial mean of the transmittance,

$$|c_0(\lambda)|^2 = \left|\,p\,t_B(\lambda) + (1-p)\,t_A(\lambda)\,\right|^2, \tag{S4}$$

which for lossless states separated by a phase step ϕ and a balanced mask reduces to $\cos^2(\phi/2)$ [3]. Eq. (S4) is a scalar thin-element calculation: it gives the consequences of a phase profile but cannot predict one. The DOE curve in Fig. 2(d) of the primary manuscript is therefore fully determined by Eq. (S3), whereas the metasurface curve is the full-wave simulated zero order of the fabricated meta-atom pair (Sec. 2). Evaluating Eq. (S4) with the supplied t_A, t_B and $p = 0.496$ gives 0.36%/1.83%/2.79% at the primaries, against the full-wave 0.27%/1.31%/2.75%; the scalar estimate tracks the full-wave zero order across the band except inside the 560–595 nm resonance, where it overshoots to 26.0% against the full-wave 2.8% [Fig. S2], bounding the validity of the thin-element picture there. This unscattered term is the light that bypasses the expansion and reaches the sensor as an un-expanded ghost copy of the hologram.

E. Multi-level fixed-depth alternatives

A conventional element with more phase levels could in principle suppress the RGB zero order further, and we quantify that route with the resin dispersion of Eq. (S3). The standard multi-level element [4] divides the 0 to 2π range at the design wavelength into N equal steps, $h_j = j\lambda_0/[N(n(\lambda_0)-1)]$. For an equal-area mask the unscattered fraction is then a Dirichlet kernel in the step phase $\Delta(\lambda) = 2\pi r(\lambda)/N$,

$$|c_0(\lambda)|^2 = \frac{1}{N^2}\frac{\sin^2[N\Delta(\lambda)/2]}{\sin^2[\Delta(\lambda)/2]}, \qquad r(\lambda) = \frac{[\,n(\lambda)-1\,]/\lambda}{[\,n(\lambda_0)-1\,]/\lambda_0}, \tag{S5}$$

where the dispersion ratio r does not depend on the level count. Because N enters Eq. (S5) only through Δ, adding levels helps at first but saturates quickly [Fig. S3, Table S3]: four levels cut the worst channel from 11.9% to 8.0%, eight levels reach 7.2%, and further levels converge to the continuous blazed limit $\mathrm{sinc}^2[r(\lambda)]$, which is 1.9%/0%/7.0% at the three primaries. That floor is set by the material dispersion alone, so no fixed-depth element quantizing a single 2π range reaches the fabricated metasurface's 2.75% worst channel, whatever its level count. Escaping the floor requires abandoning the 2π constraint: allowing the N depths to be chosen freely, four levels at $\{4.05, 4.49, 8.60, 9.03\}$ μm null the zero order to 0.001% at all three primaries, and a deep binary etch reaches 3.3% at $h = 17.61$ μm, the multi-order and harmonic regime [5–7]. Table S3 places the two families on a common footing: every standard element, and the deep binary, exceeds the metasurface's 2.75% already at zero depth error, so only the four-level design improves on it, and only while its two etch steps hold to roughly ± 56 nm. What these designs trade is fabrication. The four-level solution is a two-mask binary-optics element whose levels differ by 0.43 and 4.54 μm, so it needs two aligned exposures and a micrometer-deep etch, and the deep binary needs a 17.6 μm etch; both encode the achromatic behavior in depth, which etch rate and time set. Depth-to-pitch ratios of 0.4 to 1.6 on 10.8 μm cells also leave the thin-element picture underlying Eq. (S4), so the sub-percent figures would not survive rigorous modeling. The metasurface encodes the same behavior laterally, in pillar diameters defined by a single electron-beam exposure at one fixed 640 nm etch depth. The fabricated element's measured worst channel is 9.1% (Sec. A), and the nominal-depth columns of Table S3 no longer favor it: the blazed limit reaches 7.0%, eight levels 7.2% and the deep binary 3.3%. Because every fixed-depth entry sets its step by etch depth, the tolerance column is the comparison that matches how these elements are made, and there the measured 9.1% sits below the 20.7%, 16.1%, 15.6% and 12.8% of the binary, four-level, eight-level and deep-binary designs at ± 50 nm per etch step. Only the free four-level design stays below it, at 2.1%, and it does so by holding two aligned exposures and a 9 μm etch to that tolerance across the full aperture.

Table S3. Unscattered fraction $|c_0|^2$ of fixed-depth alternatives. The 450 and 660 nm columns are the outer primaries, where the leakage is worst; the "Worst" column is the worst of the three primaries at zero depth error, and the last column repeats it when every etch step carries an independent ± 50 nm error, near the tolerance at which the best conventional design reaches the metasurface's 2.75%. Standard multi-level elements quantize 0 to 2π at 520 nm, where they null exactly, and saturate at the blazed limit, so none of them reaches the fabricated metasurface whatever its level count. Free-depth designs reach a lower nominal floor, at the fabrication cost discussed in the text; the metasurface sets its step by lateral pillar geometry, so a depth tolerance does not apply to it in the same way.

Element	Etch depth	450 nm	660 nm	Worst	At ± 50 nm
Standard, 0 to 2π at 520 nm					
Binary ($N = 2$)	0.50 μm	7.0%	11.9%	11.9%	20.7%
Four-level	0.76 μm	2.6%	8.0%	8.0%	16.1%
Eight-level	0.88 μm	2.1%	7.2%	7.2%	15.6%
Blazed limit ($N \to \infty$)	1.01 μm	1.9%	7.0%	7.0%	–
Free depths					
Deep binary	17.61 μm	3.3%	3.3%	3.3%	12.8%
Four-level	4.05–9.03 μm	0.001%	0.001%	0.001%	2.1%
Metasurface (full-wave)	0.64 μm	0.27%	2.75%	2.75%	–

F. Parameters

Table S4 lists the parameters used for the simulation figures. One display operation is applied to the simulated intensity panels and is cosmetic and outside the model: a dark subtraction removes the median background pedestal and clips at zero, matching the dark-subtracted experimental captures. Each color channel averages $K = 8$ multiplexed solutions; because the DOE-versus-metasurface difference is the analytic chromatic ghost of Eq. (S4), which the averaging does not remove, it persists through the multiplexing.

4. EXTENDED SIMULATION RESULTS

Table S2 collects the simulated quantities reported in the primary manuscript. The unscattered fractions follow from Eq. (S3) and Eq. (S4) and are analytic, so they are independent of grid size and modulator quantization.

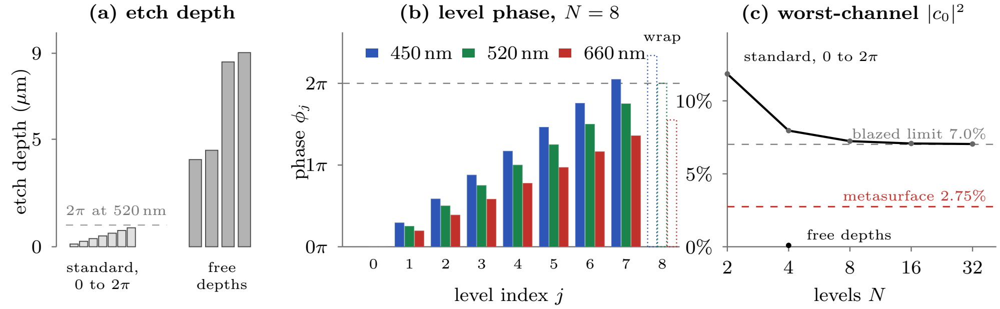

Fig. S3. Two families of fixed-depth element. (a) The standard multi-level element divides one 2π range at 520 nm into equal steps, an etch depth below 1 μm; free-depth designs use etch depths an order of magnitude larger. (b) Equal steps place the level phases on a uniform ladder of slope $\Delta(\lambda) = 2\pi r(\lambda)/N$, so the wrap step lands on 2π, and the zero order nulls, only at the design wavelength; dispersion makes it overshoot at 450 nm and fall short at 660 nm. (c) Adding levels samples the same phase range more finely, so the leakage saturates at the blazed limit, still above the fabricated metasurface.

Table S4. Simulation parameters.

Parameter	Value	Meaning
Grid $R \times C$	800×1280	Native addressed mirror grid; zero-padded to a square 1280×1280 FFT window
Adam iterations	2000	Per solution (run to convergence)
Learning rate	0.1	Adam
Multiplexing K	8	Solutions averaged per color (one video frame's experimental drive budget)
Étendue expansion factor s	1, 2, 4	Modulator samples grouped per axis
DOE design wavelength	520 nm	Sets h in Eq. (S3)
Mask fill fraction p	≈ 0.496	Measured from the expander pattern
Relay iris	passes super-pixel zone	Set per configuration (0.30 c/s at 660 nm for $s = 2$, 0.15 for the $s = 4$ comparison); blocks the $1/s$-spaced replica orders
Aperture apodization	10% taper	Cosine (soft-edged illumination)

The image contrast quoted in Figs. 2(a) and 2(b) is the mean reconstructed intensity on the target support (pixels above one fifth of the target peak) divided by the mean on the pixels the target leaves black, with the background restricted to the un-expanded ghost support. That support is the relay-iris disk: the iris clips the specular channel of a mismatched expander before the mask, so the ghost's far field is confined to the disk, one physical aperture and hence one wavelength-independent circle in camera coordinates. At $s = 2$ the disk contains the entire displayed zone, so the restriction is inactive and the $s = 2$ rows of Table S2 are the plain zone-wide ratios; at $s = 4$ the displayed zone is an order of magnitude larger in area than the disk, and the zone-wide background is dominated by the mode-deficit speckle the two expanders share (both arms read 3.1:1 zone-wide), which would dilute the chromatic signal the metric exists to isolate. The contrast is measured on the raw multiplexed intensities. We do not score the dark-subtracted panels, since the subtraction removes the very background pedestal the ratio measures, and applying it first inflates the contrast more than threefold. The binary color-ring target makes the ratio threshold-insensitive, and a ratio is unchanged by the per-channel display normalization. This is why contrast separates the two expanders where PSNR does not: the composed PSNR restricts its error to the target support, excluding the distinguishing background haze, and the per-arm content normalization removes the DOE's power deficit, so PSNR does not separate the expanders.

The second $|c_0|^2$ row is worth noting: a fixed-depth DOE designed at 660 nm instead, as in the random-DOE baseline of Ref. [2], leaks 51.2% of the blue power into the un-expanded ghost; we report the 520 nm design as the more conservative comparison. Figure S4 renders that 660 nm design directly in the $s = 4$ configuration of the primary manuscript's Fig. 2: at 450 nm the un-expanded ghost carries half the power and dominates the frame, the chromatic breakdown visible in the color binary-DOE baselines of Ref. [2], while at 660 nm the element is exact. The $s = 2$ metasurface contrast row uses the fabricated design's simulated complex response with holograms optimized against it; the ideal-π model gives slightly less (11.6:1/16.0:1/17.4:1), the fabricated response's specular remnant acting as additional controllable low-frequency modes.

The comparison of the primary manuscript's Figs. 2(a) and 2(b) operates at $s = 4$, with the relay iris stopped down to pass the $s = 4$ superpixel zone (Table S4). Figure S5 retains the same comparison at $s = 2$, the configuration of the experimental captures of Fig. 3; its per-wavelength contrasts are the $s = 2$ rows of Table S2, and the panel labels are the corresponding full-color composite values. At $s = 4$ the unscattered channel concentrates into the small central iris disk instead of spreading over the whole displayed zone, so even the metasurface's residual zero order becomes faintly visible in the red channel, consistent with the 2.75% at 660 nm of Fig. 2(d) and with no visible counterpart at 450 or 520 nm.

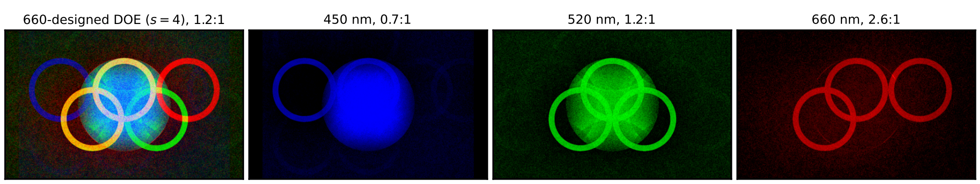

Fig. S4. The 660 nm-designed fixed-depth binary DOE of Table S2. Rendered in the $s = 4$ configuration with the same π-designed holograms as the primary manuscript's Fig. 2: full-color composite and per-wavelength channels. At 450 nm the analytic 51.2% unscattered fraction arrives as the un-expanded central ghost and outshines the target support (contrast 0.7:1); at 660 nm, the design point, the element is exact.

Increasing the étendue expansion factor trades reconstruction quality for support: for the fixed color-ring target of Fig. 2(c) the $K = 8$ PSNR, measured on the dark-subtracted multiplexed reconstructions, falls from 20.6 to 16.7 and 13.8 dB at $s = 1$, 2 and 4, while multiplexing holds the speckle contrast σ/μ near 0.5 across the sweep, so PSNR is the sharper indicator of the mode-budget cost.

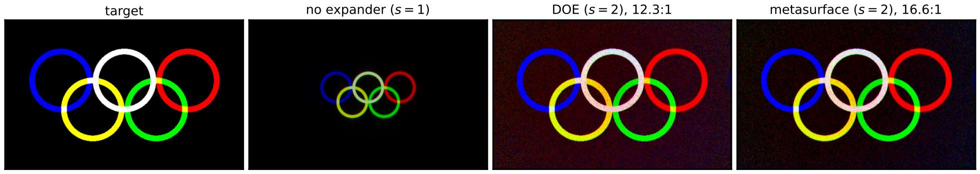

Fig. S5. The expander comparison of the primary manuscript's Fig. 2(a) at $s = 2$. The configuration of the experimental captures: the target, no expander ($s = 1$), the 520 nm-designed binary DOE, and the fabricated metasurface, $K = 8$ multiplexed solutions per color.

5. EXTENDED EXPERIMENTAL RESULTS

A. Temporal multiplexing

Speckle is the dominant residual degradation and is deterministic for a given hologram, so it averages down over independent realizations. We load eight separately optimized solutions of Eq. (S2), one per subframe of the video frame, and the camera integrates their intensities; this reuses the same hardware, introduces no learned model, and is distinct from the passive spatial expansion itself. Table S5 reports the measured reduction at 520 nm and $s = 2$, over four representative letter glyphs and over the full evaluation set; Fig. S6 below shows the corresponding captures for the three-color session.

Table S5. Measured speckle reduction, one solution against eight diverse solutions (520 nm, $s = 2$, equal brightness).

Metric	One solution	Eight solutions	Change
Contrast σ/μ, four glyphs	≈ 0.69	≈ 0.32–0.37	about $2\times$
Contrast σ/μ, full set	0.72	0.36	$2\times$
SSIM against target	0.76	0.92	$+0.16$
PSNR	15.2 dB	19.2 dB	$+4.0$ dB

Fully uncorrelated realizations would predict a contrast of $0.72/\sqrt{8} \approx 0.25$; the measured 0.36 indicates residual correlation between the solutions plus a shared background of uncontrolled-mode energy, common to every solution, that does not average down with K.

The three-color session of Sec. C repeats the sweep at $K = 1$, 8 and 24 on the additive-circles target at $s = 2$, on the composed three-color luminance [Fig. S6]: σ/μ falls from 0.65 to 0.44 at $K = 8$ but only to 0.41 at $K = 24$, the static defocused zero order and the uncontrolled-mode floor setting the asymptote. The primary manuscript's Fig. 3 nevertheless shows $K = 24$, which costs nothing within the fixed 60 ms exposure; the $K = 1$ frame demonstrates that a single subframe carries the full expanded support, only its speckle changing with K.

B. Comparison with prior étendue-expansion demonstrations

Published étendue-expansion results generally look cleaner than the raw captures shown here; the difference is one of presentation rather than physics. Kuo *et al.* [8] low-pass filter their captures with a fifth-order Butterworth filter whose cutoff matches the number of controllable modes, relocating the uncontrolled-mode energy to spatial frequencies a viewer cannot resolve, a reasonable basis for a display evaluated by eye. Tseng *et al.* [2] average 20 modulator patterns from different random seeds to reduce speckle.

Our own captures are likewise temporally multiplexed, at a depth the hardware fixes: the primary manuscript's Fig. 3 assigns the full 24-plane sequence to the sequentially-captured active laser ($K = 24$), against the 20 patterns of Ref. [2]. The modulator class also differs: both prior works drive near-continuous-phase (8-bit) liquid-crystal SLMs, whereas our patterns are 4-bit PLM states whose 24-plane sequence displays within a single video frame (Sec. 1), so the multiplexing spends subframes the drive emits anyway, at no additional display time, while working against a coarser phase quantization. Beyond that, the composites apply one mean-match gain per channel (the channel balance, Sec. C); it enters the PSNR evaluation unchanged and is applied identically

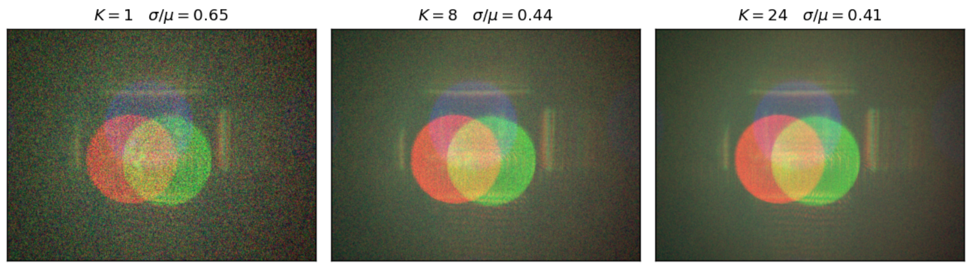

Fig. S6. Measured three-color reconstruction of the additive-circles target at $s = 2$. For $K = 1$, 8 and 24 multiplexed subframes, all displayed with the same normalization scalar. The expanded support is complete in the single subframe; averaging reduces the speckle contrast σ/μ from 0.65 to 0.44 and 0.41.

to the simulated composites, so simulation and measurement remain directly comparable. We apply no spatial filtering, no DC compensation, and no spatially varying correction of any kind; the display's reduced red modulation depth is documented by the measured LUTs of Fig. S1, independently of the composites' channel balance.

C. Three-color operation

The three-color captures of the primary manuscript's Fig. 3 come from a single capture session: the ten targets reported here, displayed at $s = 2$ through the passive metasurface and captured at all three primaries as sequential monochrome exposures from the single fixed mirror bias, with one 60 ms exposure and one laser-power setting throughout, so no per-channel gain enters at capture. The camera sits 16.8 mm off the Fourier-lens focal plane and every hologram pre-compensates that defocus, spreading the residual zero order into a low-level background; per-wavelength magnification is equalized in the hologram by pre-scaling each target by $450/\lambda$ (flat to below 1%). Display processing is the evaluation pipeline of the next subsection: per-pixel dark subtraction, the fixed left-right mirror of the folded path, registration onto the hologram grid via the calibrated scale of Eq. (S1) with sub-pixel cross-correlation, and one mean-match gain per channel into target units. That gain is the channel balance, a digital operation compensating the strongly unequal native laser powers, applied identically to the simulated composites. Captures exist at the two multiplexing depths of the drive hardware, $K = 8$ and $K = 24$ (Sec. 1); Fig. 3 and Fig. S7 show the $K = 24$ set, and Table S6 quantifies both depths.

Figure S7 shows the five subjects beyond those of the primary manuscript's Fig. 3, spanning text, color shapes, neutral logos and natural photographs; every subject reconstructs at the expanded support from the same fixed bias and hardware, and each is shown beside the scalar model's rendering of the same displayed patterns.

D. Simulation against measurement

The displayed patterns are known bit-exactly, so the captures of Sec. C admit a direct quantitative comparison with the forward model. For every subject and primary we propagate the displayed LUT-quantized patterns through Eq. (S2) with the same defocus pre-compensation, average the same K subframes, and stack the primaries into a color composite; the measured counterpart is processed as in Sec. C. Both are scored identically: each channel is scaled into target units by matching its mean over the content region (a least-squares slope would be biased low by speckle), and the color PSNR is evaluated against the target composite over the central content region, on the linear composites before the display gamma; the printed panels of Fig. 3 and Fig. S7 follow the same convention.

Table S6 reports the resulting color PSNRs. The simulation exceeds the measurement by two to five decibels on most subjects; the gap collects what the scalar model omits: residual stray background, imperfect registration between modulator and metasurface, and drive nonidealities, decomposed in Sec. E. A static four-dot alignment pattern measures 48.8 dB against 55.1 dB simulated, bounding the capture and evaluation chain itself, and $K = 24$ matches or improves on $K = 8$ for every subject, with the same saturation beyond eight subframes as Fig. S6.

The same composites bound the chromatic leakage of the fabricated mask. We compute the

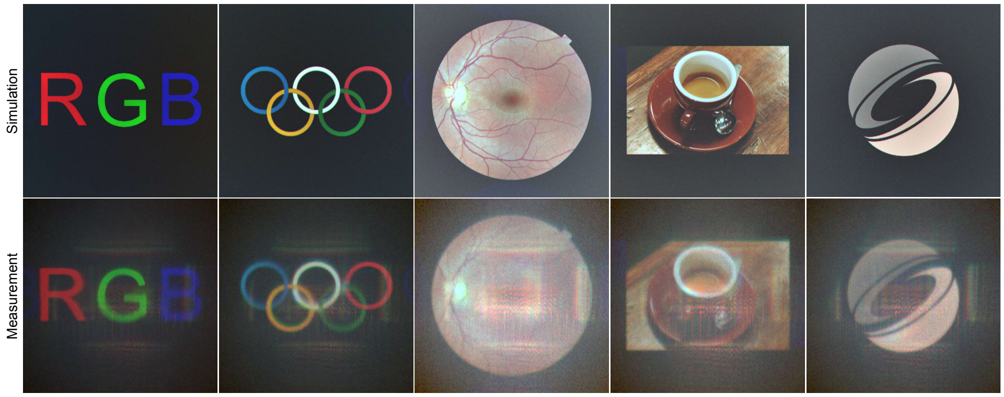

Fig. S7. The remaining five subjects of the capture session. In the layout of the primary manuscript's Fig. 3: the scalar model's rendering of the displayed $K = 24$ patterns (top) above the measured $K = 24$, $s = 2$ reconstruction (bottom), processed identically per subject and at a common magnification.

image contrast of Sec. 4 per channel over the content region, restoring the pedestal subtracted from the measured channels during registration so that measurement and simulation are treated symmetrically. Because every hologram is optimized against the nominal π mask, any chromatic deviation of the element enters the measurement only, as an unscattered ghost that raises the target-black background. The measured-to-simulated contrast ratio is near constant across the primaries (medians near 0.25 at all three; subject means 0.21, 0.19 and 0.18 at 450, 520 and 660 nm), and its mild wavelength dependence follows the documented red stray light rather than the 520 nm-centered signature of a fixed-depth element, which in the same simulated metric loses a further 1.24× at 450 and 1.7× at 660 nm (the $s = 2$ rows of Table S2). The ratio bounds the chromatic leakage below the fixed-depth scale; it does not resolve the few-percent zero order of the full-wave model, which is degenerate with the red stray light.

Two natural subjects run against the PSNR ordering: the Hubble field measures 25.1 dB against 24.0 dB simulated, and the retina's $K = 24$ composite comes within a decibel of its simulation. This is a property of the metric rather than a failure of the model: the scalar render retains the full-bandwidth speckle and overshoots the brightest cores, which dominate its error on smooth low-contrast content, while the measured channel arrives low-pass filtered by the sensor chain. Structurally the simulation stays far closer to the target (correlation above 0.95 against 0.6 to 0.8), so a measurement scoring at or above its simulation there is consistent.

Table S6. Simulated against measured color PSNR at $s = 2$ (dB).

Subject	Simulated ($K = 8$)	Measured ($K = 8$)	Measured ($K = 24$)
circle	16.5	13.3	13.4
color logo	18.1	13.2	13.5
tiger	18.0	14.2	14.3
letter	22.4	17.9	18.0
hubble	24.0	25.1	25.2
coffee	20.3	17.6	17.8
retina	21.0	19.5	20.1
rgb stripes	14.9	10.9	11.0
olympics	21.6	17.2	17.3
logo	18.0	14.2	14.2

E. Origin of the residual background

The measured composites carry a broad background that the model does not fully reproduce, and Table S7 separates its sources. Scoring the reconstructions with the contrast of Sec. 4, the measured background contrast is 2.3, 2.5 and 3.2 at 450, 520 and 660 nm in the median over the subjects of the primary manuscript's Fig. 3, against 8.8, 9.2 and 11.8 for the matched simulation of the same displayed patterns. The background therefore sits between a third and a half of the signal level, and the measurement carries about four times more of it than the forward model predicts.

The defocus is not the cause. The holograms pre-focus the content at the sensor, which sits $\Delta z = 16.8\,\text{mm}$ off the Fourier-lens focal plane, so every unmodulated term stays focused at the focal plane and reaches the sensor as a geometric blur of $N_j p\, \Delta z/(f\delta)$ pixels per axis, with $\delta = 2.4\,\mu\text{m}$ the sensor pitch. That blur is 645×403 pixels at every primary, the wavelength cancelling against the sample-to-pixel scale, a compact footprint near the center of the frame. Two unmodulated terms share it: the element's unscattered order, set by the full-wave $|c_0|^2$ of Sec. 2, and the modulator's own zero order leaking past the DC stop, which is the term the defocus was introduced to spread. We bound their sum from the captures themselves. The measured background inside the footprint runs 1.3 to 1.5 times the level outside it, while the matched simulation, which carries no leak, runs 0.7 to 0.8; the excess places the defocused zero order at about 9% of the background averaged over the scored region, at all three primaries. Removing it entirely would raise the scored contrast by at most a tenth, to 2.6, 2.7 and 3.6, against the 8.8, 9.2 and 11.8 of the simulation, so the offset is not what separates measurement from model. We do not divide this figure between the element and the modulator's leaked order: the direct measurement of Sec. A puts the element's unscattered fraction well above the simulated value, and the bound above covers both terms without separating them. The pedestal is optical rather than electronic, the dark frames placing the sensor floor near 64 counts against measured pedestals of 1600 to 3000 counts.

What remains divides into a modeled and an unmodeled part. The modeled part is the mode deficit itself: at $s = 2$ only $1/s^2$ of the output modes are controllable, so the rest carry light that lands as a floor common to every subframe, which is why the speckle contrast stops improving between $K = 8$ and $K = 24$ (Sec. 5). The simulation contains this term, and it propagates the displayed LUT-quantized patterns, so neither the four-bit quantization nor the reduced red modulation depth belongs to the remaining factor of about four. What is left unmodeled is mask-to-mirror registration error, which diffracts light into superpixel-replica orders measured near 10% locally against approximately 2% in simulation, together with stray light and LUT calibration error. A chromatic error of the phase step is disfavored: the measured-to-simulated ratio spans 0.28, 0.26 and 0.24, a spread of under a fifth, with none of the wavelength-selective signature a mismatched step would impose, and its mild fall toward the red matches the deficit already reported in Sec. C. Wide-angle scatter from the element is not separable from the mode-deficit floor in these captures, both arriving as a diffuse background, so the flat ratio bounds only its chromatic part. The signal-relative background is worst in the blue channel, which the model also predicts.

Table S7. Background contrast of the Fig. 3 subjects. Measured against the matched simulation of the same displayed patterns at $K = 24$, with the median of the per-subject ratios, which bounds what the forward model does not capture.

	450 nm		520 nm		660 nm	
Subject	sim.	meas.	sim.	meas.	sim.	meas.
circle	8.8	2.5	11.5	3.0	15.6	4.0
rgb stripes	5.4	1.6	8.4	2.4	9.4	2.2
tiger	10.3	2.3	9.2	2.5	11.8	3.2
color logo	5.1	1.5	7.2	1.9	8.5	2.0
hubble	24.5	3.0	36.8	4.4	47.5	6.1
median	8.8	2.3	9.2	2.5	11.8	3.2
ratio	0.28		0.26		0.24	

REFERENCES

1. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, (2015). arXiv:1412.6980.
2. E. Tseng, G. Kuo, S.-H. Baek, *et al.*, "Neural étendue expander for ultra-wide-angle high-fidelity holographic display," Nat. Commun. **15**, 2907 (2024).
3. J. W. Goodman, *Introduction to Fourier Optics* (W. H. Freeman, 2017), 4th ed.
4. G. J. Swanson, "Binary optics technology: the theory and design of multi-level diffractive optical elements," Tech. Rep. Technical Report 854, MIT Lincoln Laboratory, Lexington, Massachusetts (1989).
5. D. Faklis and G. M. Morris, "Spectral properties of multiorder diffractive lenses," Appl. Opt. **34**, 2462–2468 (1995).
6. D. W. Sweeney and G. E. Sommargren, "Harmonic diffractive lenses," Appl. Opt. **34**, 2469–2475 (1995).
7. T. R. M. Sales and G. D. Raguin, "Multiwavelength operation with thin diffractive elements," Appl. Opt. **38**, 3012–3018 (1999).
8. G. Kuo, L. Waller, R. Ng, and A. Maimone, "High resolution étendue expansion for holographic displays," ACM Trans. Graph. **39**, 66 (2020).