Pancake Multisource Holography

Sagnik Ghosh, Praneeth Chakravarthula, and Aswin C. Sankaranarayanan

Abstract—A common approach to increasing the étendue of a holographic display is to use an array of point light sources as the input to the system, along with two spatial light modulators placed in physically separate planes to break the memory effect; such designs invariably increase the footprint and complexity of the system, making them inconsistent with the eventual goal of a compact display. In this work, we propose a compact, high-étendue holographic display that uses a single phase-only modulator, and still breaks the memory effect. Our proposed design achieves this by routing the incident light through a single modulator twice before directing it to the observer. Such polarization-based optical folding, or pancake optics, has previously been used to achieve a compact form factor, whereas we repurpose it primarily to break the memory effect. Through simulations and a hardware prototype, we demonstrate that our design is comparable to prior multisource holographic display designs in visual quality, despite having fewer degrees of freedom, while maintaining a compact physical footprint.

Index Terms—Holographic displays, étendue expansion, pancake optics, spatial light modulators, computational photography

1 Introduction

The human visual system relies on a set of essential cues: wide field of view, high-quality noise-free visuals, and defocus cues that we desire in a near-eye display for a fully immersive experience. Conventional display designs like multi-focal displays and vari-focal displays rely on complex optical designs to mimic defocus cues, making them impractical for near-eye display implementations. In contrast, holographic displays natively support defocus cues by precisely shaping the input wavefront, and their simplicity endows them with one of the most compelling features of a near-eye display, compactness. Hence, they are widely regarded as the favored option for near-eye displays in the near future.

However, holographic displays suffer from an inherent tradeoff between the field of view and the eyebox extent that they can support. The product of the field of view and the eyebox extent, known as the étendue of the display, is constrained by the number of pixels of the phase modulator used to shape the hologram. Due to hardware constraints, increasing the pixel count of the modulator is not a viable pathway to increasing the étendue of the display. Various designs have been proposed to increase the étendue of a holographic display. Broadly, these designs can be categorized into two classes: introducing a static scattering element [3], [4], [5], [6], [7], or using a high-étendue input illumination implemented using an array of point light sources [1], [2], [8]. Introducing a high-resolution static scattering element like a binary phase mask is prone to generating noisy visuals, with severity increasing as the target étendue-expansion factor grows.

While using a high-étendue multisource backlight addresses the issue of limited étendue without introducing substantial noise, it is far from a panacea, as prior multisource designs have resorted to excessively bulky setups due to the introduction of a second modulator, contrary to the initial goal of compact holographic displays. In prior multisource holographic display designs, while the second modulator adds to the system footprint, it plays a key role; it breaks the shift-invariance of the system, also known as the memory effect in the context of Fourier optics, without which the generated visuals would suffer from severe ghosting artifacts.

We observe that the shift-invariance, or the memory effect of the system, is broken only as long as the light is modulated at two distinct planes along its path, regardless of whether the modulation patterns at the two planes are distinct. Hence, a system with the same modulator placed at two distinct planes along the light path would successfully break the memory effect. Using polarization-based optical folding, commonly referred to as pancake optics, our proposed design folds the light path of prior multisource designs [1], [2] such that the two modulators collapse into a single modulator in the folded design. The folded design thus eliminates the need for a second modulator while preserving the ability to modulate the wavefront at two distinct planes, thereby breaking the memory effect. An additional benefit of the folded design is that it naturally favors the use of reflective phase modulators, which are more widely adopted in comparison to transmissive modulators due to manufacturing simplicity.

Our contributions can be summarized as follows:

- We propose a multisource design that breaks the memory effect with a single reflective phase SLM in a smaller form factor using pancake optics.
- We validate through simulations that using the same modulation pattern for both modulation stages does not cause a significant drop in visual quality.
- We build a hardware prototype of our design and show experimentally captured results.

2 Related Work

2.1 Holographic Displays and the Étendue Bottleneck

Holographic displays have emerged as a compelling platform for near-eye displays because they directly reconstruct optical wavefronts, thereby providing correct accom-

- *SG and ACS are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213.*
PC is with the Department of Computer Science, University of North Carolina, Chapel Hill

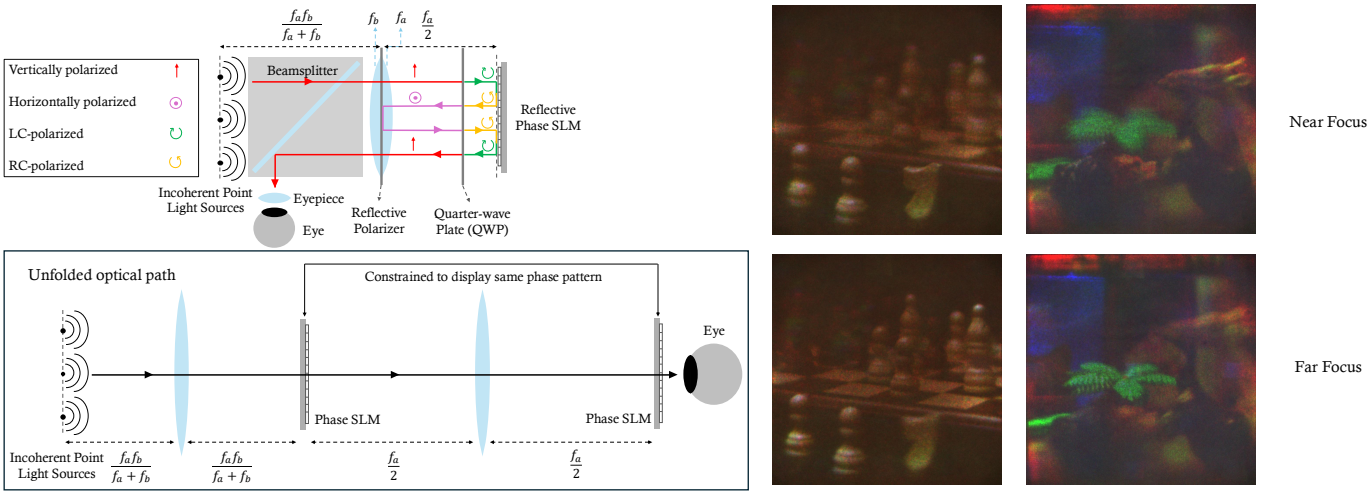


Fig. 1: We propose *Pancake Multisource Holography*, a technique that uses polarization-based optical folding, or pancake optics to implement a compact holographic display. Our optical design (top left) shows how we can break the memory effect with a single reflective phase spatial light modulator (SLM) by redirecting light to be incident on the SLM twice, thereby eliminating the need for multiple SLMs. The unfolded version of the proposed system (bottom left) resembles prior high-étendue multisource holographic display designs [1], [2], and highlights the optical path compression achieved by our folded configuration. This folding, however, imposes the constraint that the same phase modulation pattern is applied in both modulation stages. Despite this harsh constraint, our system can produce 3D holograms with quality and étendue comparable to existing multisource display designs, as seen in the near-focus (top right) and far-focus (bottom right) captures from our hardware prototype.

modation cues, natural depth perception, and true three-dimensional imagery [9], [10]. Recent advances in neural hologram synthesis [11], [12], end-to-end optimization frameworks [13], improved light efficiency and brightness [14], [15], and increasingly compact optical architectures [16], [17], [18] have significantly improved the feasibility of holographic near-eye displays. Despite this progress, current systems remain fundamentally constrained by the limited étendue of spatial light modulators (SLMs).

2.2 Étendue Expansion for Holographic Displays

A wide range of techniques have been proposed to effectively expand or redistribute the available étendue in holographic display systems.

Static masks. One approach relies on designing the optical configuration itself to expand system étendue, commonly achieved by introducing a high-resolution static mask [3], [4], [5], [6], [7] whose pitch is significantly smaller than the pixel pitch of the SLM. Park et al. [19] demonstrated an analogous concept using a non-periodic photon sieve, a static mask of randomly positioned sub-wavelength pinholes placed directly on a transmissive LCD panel, to achieve wide-angle diffraction and large-area holographic display in an ultrathin form factor.

Complex illumination. An alternative approach expands étendue by designing incident illumination that is more complex than a single collimated beam of light. Monin et al. [20] demonstrated per-pixel angular steering by introducing multiple binary tilt layers prior to SLM modulation, enabling exponential étendue scaling with the number of layers. However, realizing these tilt layers requires multiple cascaded relays, which significantly increases optical complexity and limits practicality in compact near-eye displays.

Temporal steering. Another family of methods achieves étendue expansion by temporally steering the incident illumination [7], [21], effectively sweeping a small pupil across a larger eyebox over time. While this increases viewing tolerance, it consumes the temporal dimension, which could otherwise be used to improve image quality.

Pupil replication and eyebox redistribution. Several methods expand the effective eyebox by optically replicating or redistributing the exit pupil. Chang et al. [22] multiplexed holograms with multiple off-axis plane waves to duplicate the Maxwellian-view pupil into a spatial array, expanding the eyebox while maintaining a compact lensless configuration. Chae et al. [23] replaced the conventional eyepiece with a lens array to generate a sparse eyebox, redistributing the limited étendue over a larger area at the cost of spatial continuity and brightness uniformity.

Coherent multi-source illumination. Lee et al. [24] and Watanabe et al. [25] explored mutually coherent multi-source illumination, which requires fewer degrees of freedom than mutually incoherent sources but sacrifices the despeckling benefits of incoherent illumination. More recently, Xia et al. [26] explored multi-illumination-interfered neural holography with coherent multi-angle illumination and a single SLM for eyebox expansion, whereas our approach uses mutually incoherent sources and a second modulation stage to break the memory effect.

2.3 Multisource holographic configurations

Two-SLM design. Multisource holographic display configurations typically employ an array of point light sources to illuminate downstream optics that generate the hologram. For such configurations, Jo et al. [2] argue that a single

plane of modulation would create multiple repeated copies of an image (a consequence of the memory effect), and suggest the use of a second modulation stage in the form of a static binary amplitude mask in the Fourier plane of the SLM. However, the static nature of the binary mask limits the available degrees of freedom and degrades visual quality. Using light sources that are closely spaced, Kuo et al. [8] show that a similar design but with two SLMs can be used to achieve significant speckle reduction. Recent dual-SLM designs such as HoloChrome [27] have also explored polychromatic illumination for speckle reduction, which is complementary to our focus on compact two-stage modulation for multisource étendue expansion.

Fourier-plane amplitude SLM. A more recent work by Chao et al. [1] replaced the static binary mask of Jo et al. [2] with a Fourier-plane amplitude SLM, significantly improving visual quality and comprehensively demonstrating view-dependent effects.

Our approach. Our proposed design falls within this class of multisource holographic configurations and similarly uses an array of mutually incoherent sources as input illumination. However, our primary focus is on enabling a compact prototype by careful redirection of light using polarization-based optical folding (pancake optics) to provide multiple planes of modulation using a single reflective phase SLM. We discuss polarization-based optical folding and its role in our design in the following subsection.

2.4 Polarization-Based Optical Folding

Polarization-based optical folding, commonly referred to as pancake optics, uses a combination of polarizing beam splitters, quarter-wave plates, and reflective polarizers to fold the optical path. Light bounces back and forth between the components multiple times before exiting, effectively achieving a long optical path length in a thin form factor. This architecture has been widely adopted in near-eye displays as a way to reduce overall thickness.

Narasimhan [28] demonstrated an ultra-compact pancake lens system for VR headsets that exploits curved aspherical surfaces within the folded cavity to achieve optical foveation alongside significant volume reduction. Maimone et al. [29] combined holographic optical elements with polarization-based optical folding and directional laser backlighting to realize VR displays with thicknesses under $9\,\mathrm{mm}$ and fields of view exceeding $90°$, demonstrating that pancake architectures can be integrated with holographic optics without sacrificing display performance. Bang et al. [30] proposed a thin VR display combining Fresnel lenses with a lenslet array in a folded optical path inspired by pancake optics, achieving a wide field of view in a flat form factor, though their system was limited to 2D image generation. Cakmakci et al. [31] extended the pancake design to optical see-through augmented reality for the first time, using volume holographic elements to replace conventional refractive components within the folded cavity.

Across these prior works, polarization-based optical folding has been employed to improve form factor and display performance. In contrast, our design leverages the folded optical path for a different purpose: by redirecting light back onto the same reflective phase SLM, the folding provides a second plane of modulation that breaks the memory effect inherent to multisource holographic configurations (Section 2.3), while simultaneously retaining the compactness benefits of the pancake architecture.

3 Method

3.1 Single-source holography

The most rudimentary holographic display configuration consists of a collimated beam of light illuminating a phase SLM, followed by the user's eye, such that the field formed on the retina is the optical Fourier transform of the field at the SLM. More formally, a thin lens of focal length f performs an optical Fourier transform: the field at the back focal plane is related to the field u at the front focal plane by

$$\frac{1}{j\lambda f}\,U\!\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right) \tag{1}$$

where U is the Fourier transform of u and λ is the wavelength of illumination. In the specific case where the field at the front focal plane is a phase SLM illuminated by a collimated beam, the Fourier transform formed by the eye lens of focal length f_{eye} is uniquely controllable over an extent of $f_{\text{eye}}\lambda/p$ along each axis, where p is the pixel pitch of the SLM, as illustrated in Figure 2(a). Hence, the FoV can be derived as

$$\text{FoV}_x = \text{FoV}_y = \frac{\lambda}{p} \tag{2}$$

where p is the pixel pitch of the SLM and λ is the wavelength of the input illumination. Further, the extent of the eyebox is simply the physical extent of the SLM,

$$\text{Eyebox} = H_{\text{SLM}} \times W_{\text{SLM}} \tag{3}$$

where H_{SLM} and W_{SLM} are the height and width of the SLM, respectively. The étendue G of the display is defined as the product of its field of view and eyebox,

$$G = \left(\frac{\lambda}{p}\right) \times \left(\frac{\lambda}{p}\right) \times (H_{\text{SLM}} \times W_{\text{SLM}}) = \lambda^2 N_x N_y, \tag{4}$$

where N_x and N_y are the number of pixels of the SLM along each axis.

3.2 Multisource holographic designs

Multisource holography addresses the issue of limited étendue by introducing additional collimated beams incident at different angles on the SLM. This is achieved by collimating an array of point light sources spaced apart as illustrated in Figure 2(b). In this configuration, the image plane is related to the source plane by a 4F-system, with the SLM modulating each source wavefront by u_{SLM} at the Fourier plane. The wavefront at the image plane due to source s_i is

$$u_i(x,y) = U_{\text{SLM}}\left(\frac{x+x_i}{\lambda f_1},\frac{y+y_i}{\lambda f_1}\right) \tag{5}$$

where U_{SLM} is the Fourier transform of u_{SLM}, and x_i and y_i are proportional to the location of the source s_i in the source plane.

Although introducing additional sources leads to an increase in FoV (with the eyebox remaining the physical

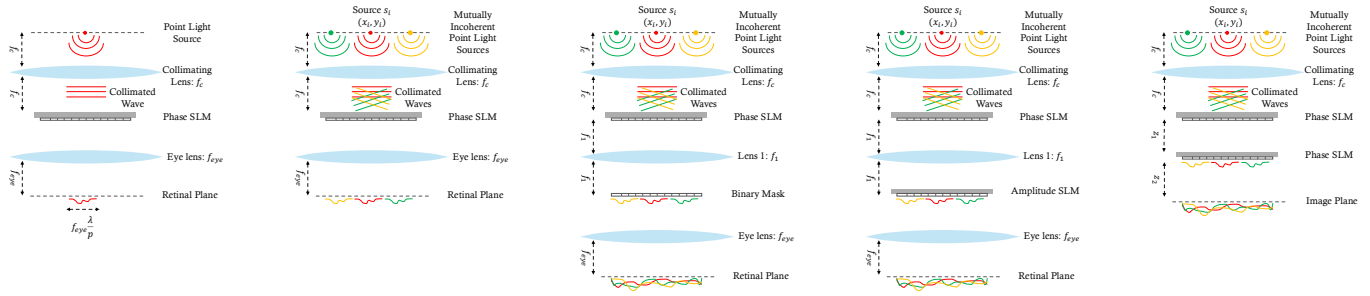


Fig. 2: **Multisource holographic display configurations.** (a) Single-source holography. (b) Multisource configuration with a single modulation stage, which does not break the memory effect and therefore produces repeated copies of each source wavefront at the image plane. (c) Configuration proposed by Jo et al. [2], which introduces a random binary amplitude mask in the Fourier plane to decorrelate the wavefronts from different sources at the eyebox plane, thereby breaking the memory effect; as a result, the individual source wavefronts are no longer repeated copies. (d) Configuration proposed by Chao et al. [1], which replaces the binary amplitude mask with an amplitude SLM in the Fourier plane, leading to a significant improvement in image quality compared to the design of Jo et al. [2]. (e) Configuration proposed by Kuo et al. [8], in which light propagates from the first SLM to a second SLM over a finite distance, breaking the memory effect; the closely spaced multisource array in this design is primarily used for incoherent speckle averaging and is less effective for etendue expansion.

extent of the SLM), Equation (5) indicates that the wavefront associated with each source is merely a shifted copy. In general, this effect is observed in optical systems with a single scattering element and is referred to as the memory effect. In order to break the memory effect, prior methods introduce another scattering element, like a static binary mask [2], an amplitude SLM [1], or a phase SLM [8]. While this effectively breaks the correlation between the different source wavefronts, introducing an additional SLM or a binary mask adds to the cost and complexity of the system.

Étendue and degrees of freedom. Jo et al. [2] and Chao et al. [1] place a second modulator at the Fourier plane of the first to break the memory effect, as illustrated in Figure 2(c) and (d). The image plane in these configurations is related to the SLM by a 4F-relay with magnification f_{eye}/f_1; consequently, the field of view is

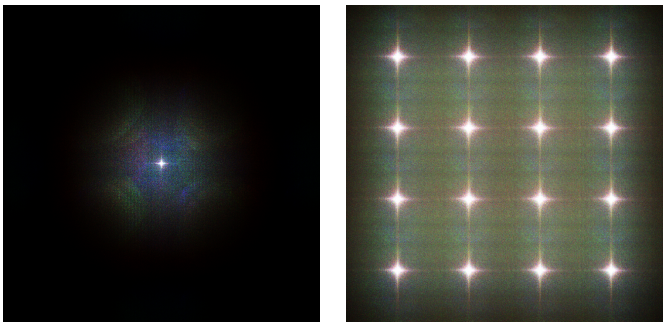
$$\text{FoV}_x = \frac{W_{\text{SLM}}}{f_1}, \quad \text{FoV}_y = \frac{H_{\text{SLM}}}{f_1} \tag{6}$$

where W_{SLM} and H_{SLM} are the physical dimensions of the first SLM. The eyebox plane of this design corresponds to the plane of the second SLM. From Equation (5), it is evident that for a single source, the eyebox extent equals the extent of its shifted Fourier spectrum, which is $f_1\lambda/p$ along each direction, where p is the pixel pitch of the first SLM. The étendue is therefore

$$G = \left(\frac{W_{\text{SLM}}}{f_1} \times \frac{H_{\text{SLM}}}{f_1}\right) \times \left(\frac{f_1\lambda}{p} \times \frac{f_1\lambda}{p}\right) = \lambda^2 N_x N_y, \tag{7}$$

which remains directly proportional to the number of pixels of the first SLM. By introducing multiple sources, the eyebox can be extended to span the full physical extent of the second SLM, as illustrated in Figure 3. This decouples the étendue from the SLM resolution, or equivalently, the degrees of freedom of the system, while simultaneously breaking the memory effect due to two-stage modulation.

Visual quality and degrees of freedom. The total degrees of freedom of a two-SLM multisource system is simply the sum of the resolutions of the two SLMs, and one might expect the visual quality to be proportional to this aggregate. However, Figure 4, which plots PSNR as a function of the resolutions of the two SLMs for the case of a Fourier phase SLM (the unfolded equivalent of the proposed design), suggests that visual quality is not merely proportional to the aggregate degrees of freedom afforded by the two SLMs. In particular, the resolution of the first SLM has little bearing on visual quality. Similarly, Chao et al. [1] report that aggressively binning the Fourier amplitude SLM to mitigate alignment sensitivity does not noticeably degrade visual quality. Taken together, these observations suggest that the degrees of freedom associated with each of the SLMs play asymmetric roles in forming the visual, and a large fraction of the combined degrees of freedom is superfluous.



(a) Single source. (b) 4×4 source array.

Fig. 3: **Eyebox energy distribution.** The energy distributions shown in (a) and (b) were averaged over four lightfield scenes and cropped to an extent of $8.64\,\text{mm} \times 8.64\,\text{mm}$ around the optical axis.

Reliance on transmissive SLMs in two-SLM multisource designs. Transmissive SLMs, in contrast to reflective

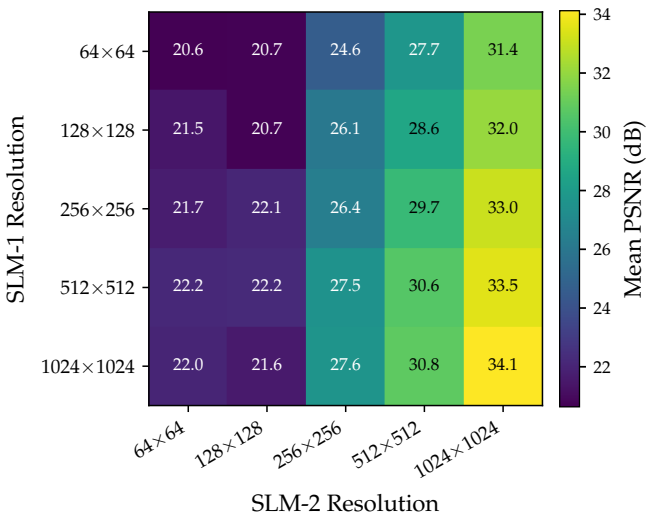


Fig. 4: **PSNR (dB) as a function of SLM resolution for a multisource holographic display.** The multisource configuration used is that of Chao et al. [1] with a phase SLM at the Fourier plane, which is also the unfolded equivalent of the proposed design (Figure 1). PSNR values were obtained by optimizing and averaging over four focal stack scenes, using a 4×4 source array and a Fourier-transforming lens with focal length $62.5\,\mathrm{mm}$.

SLMs, house their driving circuitry between the pixels, greatly reducing diffraction efficiency. As reflective SLMs are the preferred choice for optical setups, the complexity of the final hardware prototypes of prior multisource designs is further exacerbated in the process of accommodating reflective SLMs. While their conceptual designs (Figures 2(c–e)) are not egregiously bulky with the use of transmissive SLMs, the equivalent hardware prototypes have large footprints due to the use of multiple 4F relays.

Motivation for proposed design. The above analysis reveals that the degrees of freedom of a two-SLM multisource holographic display are not fully utilized in forming the visual, and the reliance on transmissive SLMs in such designs further constrains practical adoption. Together, these considerations motivate a multisource holographic display design that employs the available degrees of freedom more judiciously, while dispensing with transmissive SLMs altogether. In the following subsection, we describe the proposed design, which uses a single reflective SLM to break the memory effect, thereby eliminating the need for a second modulator. Despite operating with half the aggregate degrees of freedom of existing two-SLM designs, the proposed system achieves comparable visual quality with a more compact form factor.

3.3 Pancake Multisource Holography

In the context of near-eye displays, pancake optics has been used solely to miniaturize viewing optics using polarization components. In the proposed design, in addition to miniaturization, it serves to obviate the need for a second modulator typically used in multisource holographic designs. Specifically, by using a reflective wire grid polarizer and a quarter wave plate in the configuration shown in Figure 1, the incident light from the multisource array traverses the cavity between the reflective polarizer and the SLM three times before reaching the user's eye. In the process, it gets modulated by the same SLM twice at different planes along its light path, breaking the memory effect.

The quarter wave plate transmits linearly polarized light of orthogonal orientations to circularly polarized light of opposite handedness and vice versa. At the first interaction with the SLM, the incident light is right circularly polarized and, upon reflection from the SLM, it flips handedness. As the orthogonal linear polarization gets reflected by the wire grid polarizer, it enables a second interaction with the SLM. Note that this design requires an SLM that modulates phase while being agnostic to the polarization state of the incident light; such a capability is provided by recent advances in micromirror-based phase modulation. The complete light path color-coded by polarization is illustrated in Figure 1.

The wavefront at the second interaction with the SLM is related to the wavefront at the first interaction by an optical Fourier transform with a lens of focal length $f_a/2$. Hence, it can be shown that the wavefront due to a single source s_i after the second interaction with the SLM is

$$u_{2i}(x,y) = U_{\text{SLM}}\left(\frac{x+x_{2i}}{\lambda f_a/2}, \frac{y+y_{2i}}{\lambda f_a/2}\right) \cdot u_{\text{SLM}}(x,y), \quad (8)$$

where U_{SLM} is the Fourier transform of u_{SLM}, and x_{2i} and y_{2i} are proportional to the position of source s_i in the source plane.

An important consideration is that the phase patterns at both interactions with the SLM are constrained to be the same. While this has no bearing on breaking the memory effect, it ostensibly constrains the space of possible images or focal stacks generated by the design. Through simulations (see Section 4), it can be shown that the holograms generated by this design span a rich space of visuals, and the quality is comparable to prior multisource designs with multiple modulators.

4 Analysis

4.1 Field-of-view and eyebox extent

The proposed design (Figure 1), when unfolded, is equivalent to a two-SLM Fourier modulation multisource system ([1], [2]), and therefore inherits the field-of-view and eyebox analysis of Section 3.2. Specifically, the effective focal length of the Fourier-transforming lens between the two modulation planes in the proposed design is $f_a/2$; substituting into Equation (6) gives

$$\text{FoV}_x = \frac{2W_{\text{SLM}}}{f_a}, \quad \text{FoV}_y = \frac{2H_{\text{SLM}}}{f_a} \quad (9)$$

where W_{SLM} and H_{SLM} are the physical dimensions of the SLM. Similarly, the eyebox extent is simply the physical extent of the SLM.

4.2 Effect of number of sources

In a multisource holographic display where the illumination sources are mutually incoherent, increasing the number of sources leads to improved image quality due to incoherent noise averaging. However, this improvement exhibits

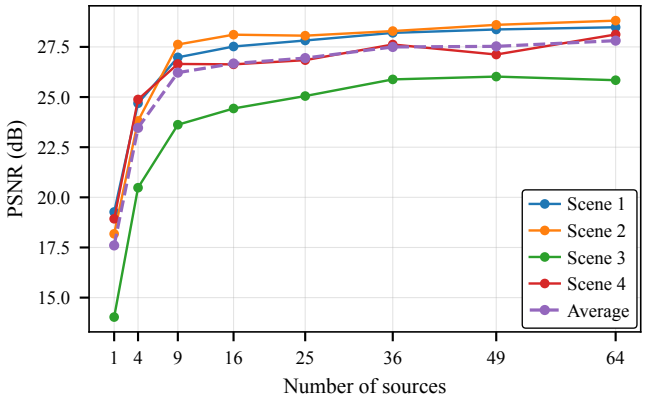


Fig. 5: **Effect of number of sources on visual quality.** Our method was simulated using different number of sources across four focal stack scenes. The visual quality plateaus after around sixteen sources. These simulations used an SLM resolution of 800×800, lower than that used in the quantitative comparisons of Section 4.4, due to the computational cost of simulating up to 64 mutually incoherent sources; this accounts for the relatively lower PSNR values compared to those reported there.

diminishing returns beyond a certain point. Figure 5 shows that this trend is also observed for our proposed multisource display design. In particular, there is a large improvement in image quality when increasing the number of sources from one to four, and from four to nine. The improvement from nine to sixteen sources is noticeably smaller, and beyond sixteen sources the increase in image quality becomes marginal, despite the substantially higher computational cost required. Hence, nine and sixteen sources represent reasonable operating points for our system, and we use sixteen sources for all simulations as this balances image quality and computational cost.

4.3 Higher-order Diffraction

As discussed by Chao et al. [1], higher-order diffraction is unavoidable in multisource systems because higher diffraction orders (HDOs) overlap with useful shifted spectra and cannot simply be separated using a conventional blocking filter. One key benefit they identify for dynamic Fourier-plane modulation is that it accounts for HDOs in the forward model and content-adaptively modulates the Fourier spectrum. The Fourier-plane modulation in our design is also dynamic, although it is implemented using the same SLM as the first modulation stage; therefore, our method inherits the same ability to account for and modulate higher-order diffraction terms rather than requiring them to be blocked.

4.4 Simulated comparisons

We now provide a comprehensive comparison of our proposed compact high-étendue design to prior high-étendue display designs. The design closest to ours is that of Chao et al. [1], with two key differences: it uses an amplitude SLM in the Fourier plane, and its two modulation stages

TABLE 1: **Quantitative comparison in simulation.** We compare to: static phase mask with low-pass filter [4], multi-source with static binary mask in Fourier plane [2], dynamic Fourier modulation [1], and a variant of that technique with a phase SLM in the Fourier plane, and our proposed technique including a time-multiplexed variant with four temporal frames. Each entry reports average PSNR (dB) / SSIM across four scenes for focus stack and light field guidance.

Method	Focus stack	Light field
Kuo et al. [4]	22.53 / 0.464	15.85 / 0.173
Jo et al. [2]	23.81 / 0.515	19.26 / 0.237
Chao et al. [1]	32.81 / 0.896	27.30 / 0.777
Chao et al. [1] (Phase SLM)	35.07 / 0.905	26.06 / 0.651
Proposed	32.24 / 0.837	25.07 / 0.656
Proposed (time multiplexed)	35.00 / 0.907	27.71 / 0.794

are independent. We also compare to a slight variant of this technique, where the amplitude SLM is replaced with a phase SLM, thereby creating a close competitor to our technique where the two planes of modulation are independent. We use this to quantify the drop in performance due to the use of a single SLM. In addition, we compare to Jo et al. [2] and Kuo et al. [4], where fixed binary amplitude and binary phase masks are used, respectively, to increase étendue. Finally, for the proposed method, we also consider a time-multiplexed variant [32] that allows for speckle reduction via averaging across multiple patterns shown on the SLM in rapid succession, within the persistence of vision of the eye. For each method, we consider two content optimization techniques for estimating the SLM patterns, based on target data that is either a focus stack or a light field. Additional simulation details, including optimization details for both focal stack and light field scenes, are provided in the supplementary material.

The simulated diagonal field-of-view for all methods is fixed to be $8.04°$. The phase SLM specifications were simulated to match the TI DLP6750Q1EVM, which has a resolution of 800×1280 pixels and a pixel pitch of $10.8\,\mu\text{m}$. The amplitude SLM used in [1] and the binary amplitude mask used in [2] were both simulated using a resolution of 500×500 with a physical extent of $10\,\text{mm} \times 10\,\text{mm}$. We used a 4×4 source array with a source spacing of $1.94\,\text{mm}$. All methods are implemented using PyTorch.

Table 1 summarizes the quantitative comparison of various methods across four scenes, and for focus stack and light field targets. We first observe that all methods perform significantly better on focal stack scenes than on light field scenes. This can be attributed to the fact that, in the focal stack setting, the entire eyebox contributes to image formation, whereas in the light field case only a part of the eyebox is used to form each view. It is also evident that methods that rely on static masks [2], [4] perform considerably worse than approaches that employ dynamic modulation at all modulation planes. From a light propagation perspective, the Fourier phase SLM variant of Chao et al. [1] can be viewed as equivalent to our method (unfolded) without the

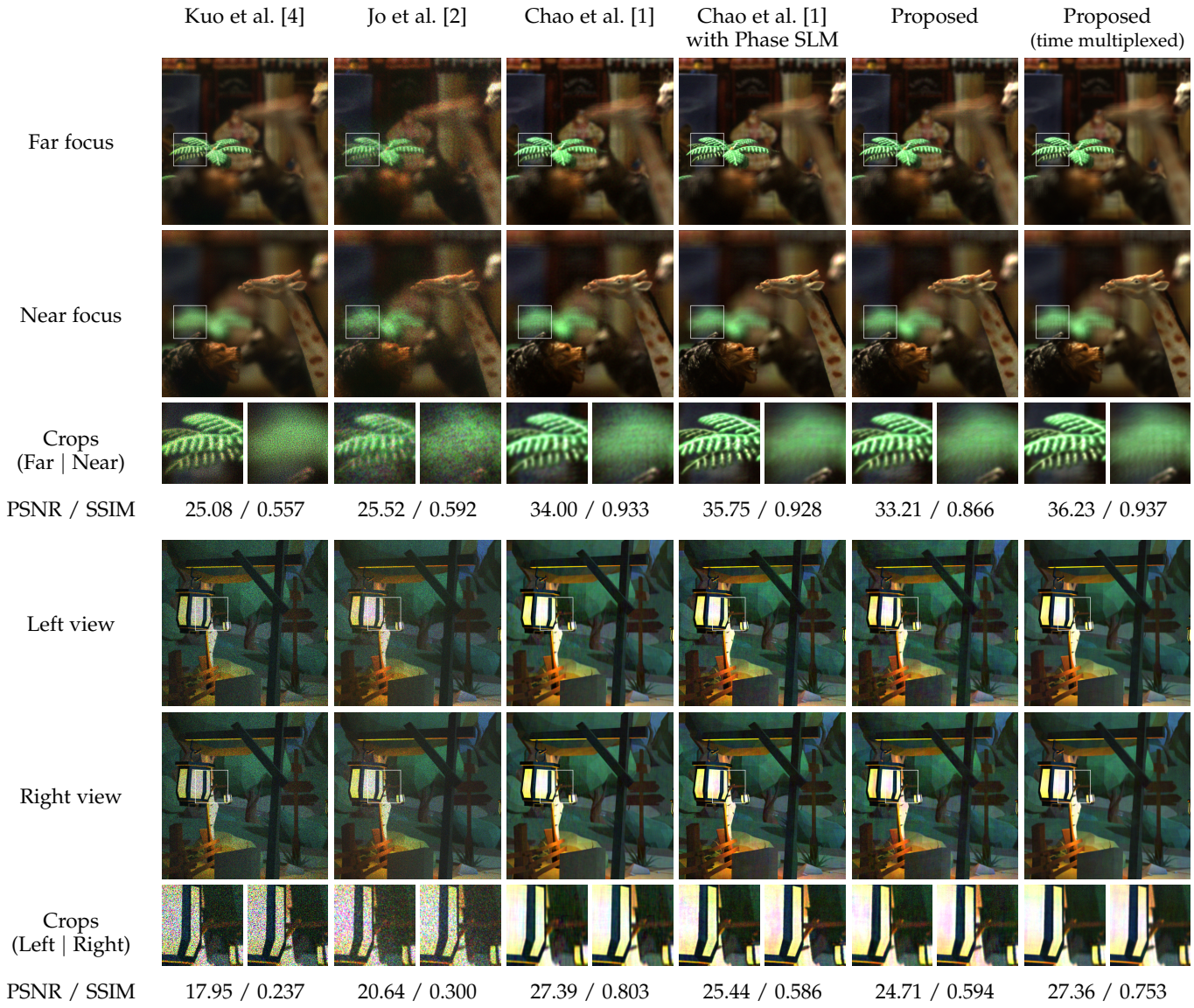


Fig. 6: Qualitative comparison on a focal stack scene (top) and a light-field scene (bottom) in simulation. Full views show the evaluation region (boxed); crops show zoomed-in details. Metrics are reported as PSNR(dB) / SSIM.

constraint that the phase patterns at the two modulation stages are identical, which gives it additional degrees of freedom. Despite this strong constraint, the table shows that our method performs only slightly worse than its unfolded counterpart without the constraint, while significantly outperforming methods based on static masks. This indicates that our approach inherits the benefits of dynamic two-stage modulation, even though the same phase pattern is used at both modulation stages. Figure 6 shows qualitative comparisons on a focal stack and a lightfield scene. Additional simulated focal stack and light field results are provided in the supplementary material.

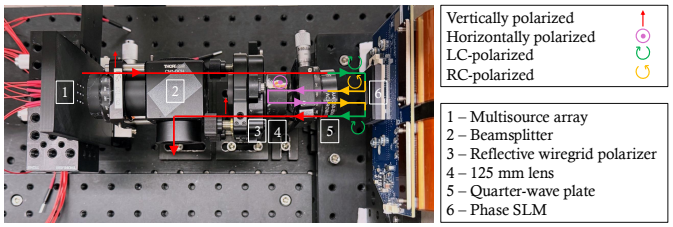
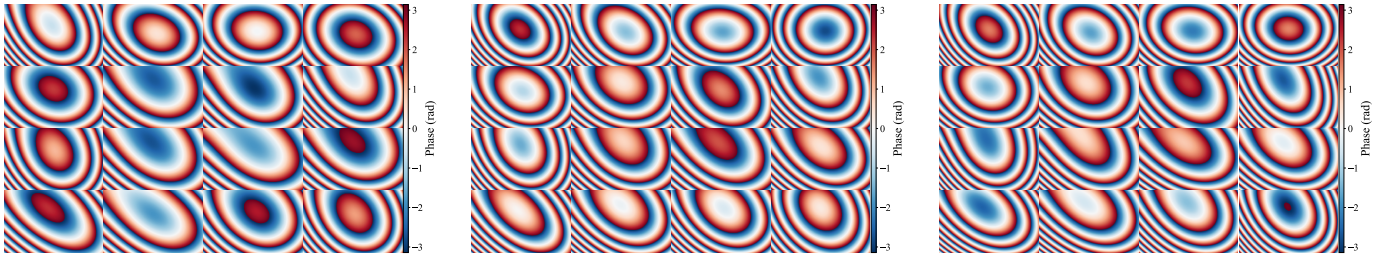


Fig. 7: Hardware prototype.

5 Hardware Prototype

5.1 Hardware prototype specifications

Figure 7 presents the hardware prototype used to capture 3D scenes, serving as a proof of concept for our approach. Our hardware prototype uses three fiber-coupled laser light sources for the red, green, and blue channels (Thorlabs LP637-SF70, LP520-SF40, and LP488-SF20G), which are combined using an RGB combiner (Thorlabs 46HF) and subsequently routed through a custom fiber splitter. The splitter is designed such that each output path is approximately 10 cm longer than the previous one, exceeding the coherence lengths of the input laser sources and ensuring mutual incoherence across the multisource array.



(a) Red (637 nm) (b) Green (520 nm) (c) Blue (488 nm)

Fig. 8: Calibrated Zernike phase corrections for all sources across (a) red, (b) green, and (c) blue wavelengths.

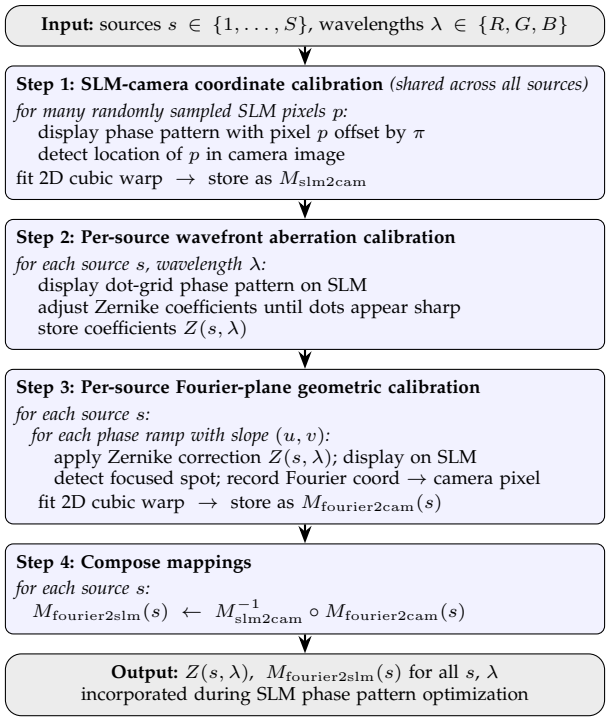


Fig. 9: Flowchart of the calibration procedure.

The splitter outputs are terminated with fiber tips and held in a 3D-printed mount to form a 4×4 point-source array. This array is followed by a reflective wire-grid polarizer and a 125 mm focal length lens that collimates the input point light sources. The same lens also serves as the Fourier-transforming lens between the first and second modulation planes, with an effective focal length of 62.5 mm. An achromatic quarter-wave plate (Thorlabs AQWP10M-580) is used in the setup, and the phase SLM used is the TI DLP6750Q1EVM, which has a resolution of 800×1280 pixels and a pixel pitch of 10.8 µm, chosen for its polarization-independent operation. The SLM supports three-channel, four-frame time-multiplexing at 120 Hz.

5.2 Calibration procedure

In our optical setup, the input point light sources are first collimated and incident on the SLM, which serves as the first modulation plane. After this initial modulation by the SLM, the resulting wavefront undergoes a Fourier transform that is formed physically at the same SLM plane during the second interaction with the SLM. Consequently, both the SLM modulation of the incident field and the corresponding Fourier-domain field associated with each source are defined on a common SLM plane. For calibrating the light in its first and second interaction with the SLM, we adjust the quarter-wave plate so that a portion of the light from the first modulation passes through the optical path without undergoing the Fourier transform, allowing the SLM pixels themselves to be imaged directly at the camera. By directly imaging this plane, we can calibrate source-dependent Fourier-domain aberrations as well as the geometric mapping between the per-source shifted Fourier transforms and the SLM pixel coordinates. The full procedure, which proceeds in four steps and whose outputs are incorporated during the optimization of SLM phase patterns for different 3D scenes, is summarized in Figure 9.

Per-source wavefront aberration calibration. Due to source-dependent optical imperfections, the Fourier transform associated with each point light source is aberrated and does not form a sharply focused pattern. We model these aberrations using a single phase correction term per source, parameterized by a Zernike basis defined on the SLM plane. A phase pattern that is optimized to produce a grid of dots in the Fourier plane is displayed on the SLM. The Zernike coefficients are adjusted until the grid corresponding to that source appears sharp, correcting the wavefront aberrations. The resulting Zernike coefficients are stored for each source and wavelength combination and applied to all subsequent phase patterns associated with that source. Figure 8 visualizes the calibrated Zernike phase corrections for all sources and wavelengths.

SLM-camera coordinate calibration. Before calibrating source-specific Fourier-domain geometry, we compute a global geometric mapping between SLM pixel coordinates and camera pixel coordinates. This mapping is obtained by randomly sampling individual SLM pixels and displaying phase patterns in which the selected pixel has a phase offset of approximately π relative to the remaining pixels, allowing its location to be detected in the captured camera image. Repeating this process across many sampled pixels provides a set of SLM-camera correspondences, which are used to fit a two-dimensional cubic warp. This global SLM-camera mapping is shared across all sources and serves as a common spatial reference for subsequent per-source geometric calibration steps.

Per-source Fourier-plane geometric calibration. We calibrate geometric distortions in the Fourier-domain represen-

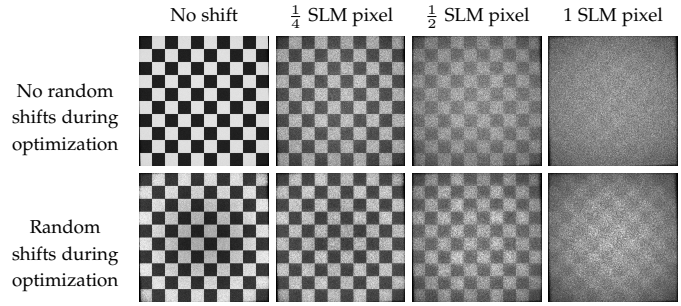


Fig. 10: **Effect of micron-level shifts between the source Fourier transforms and the SLM.** Progressive deterioration in image quality as the relative shift $(\Delta x, \Delta y)$ increases from zero to one full SLM pixel ($10.8\,\mu$m). The bottom row shows improved robustness obtained by incorporating random shifts during optimization.

tation associated with each source following the first modulation. For a single source, we sequentially display a set of phase ramps on the SLM, each pre-corrected using the corresponding source's Zernike coefficients. Each phase ramp produces a focused spot at a known Fourier coordinate determined by the horizontal and vertical slopes of the ramp. These spots are detected in the camera images, and a two-dimensional cubic warp is fit between the expected Fourier coordinates and the observed camera pixel locations. This process is repeated independently for each source, resulting in a per-source mapping that captures source-dependent geometric distortions in Fourier coordinates.

Composition of mappings. The calibrated geometric mappings are composed to obtain a correspondence between Fourier-domain coordinates and SLM pixel coordinates for the second modulation. Specifically, the per-source Fourier-to-camera mapping is composed with the inverse of the global SLM-camera mapping to produce a source-dependent Fourier-to-SLM coordinate transformation. This composed mapping is incorporated during the optimization of SLM phase patterns for different 3D scenes, ensuring that wavefront aberrations and geometric distortions are accounted for during the second interaction with the SLM.

5.3 Sensitivity to alignment

A central challenge in calibrating our setup is its extreme sensitivity to micron-scale misalignment between the per-source Fourier transforms and the SLM pixel grid. Even very small relative shifts cause significant degradation in image quality. This effect is illustrated in the top row of Figure 10, where a relative shift of approximately half an SLM pixel in both directions ($5.4\,\mu$m) results in a substantial loss of image quality, while a shift of one full SLM pixel in each direction causes the image to nearly disappear.

Introducing random Fourier-domain shifts during the optimization of the SLM phase pattern improves robustness to such misalignment. As shown in the second row of Figure 10, optimization with random shifts improves robustness to increasing misalignment. In practice, we observe that performing an additional per-source fine-tuning step improves image quality by compensating for residual bias in the calibrated Fourier-to-SLM mapping.

A similar sensitivity to alignment has been reported by Chao et al. [1], who attribute the gap between simulated and real-world visual quality in their system to challenges in aligning the per-source Fourier transform with the Fourier-plane modulator. Because their system employs a separate amplitude SLM in the Fourier plane, they address this sensitivity by aggressively binning the pixels of that SLM.

We attribute the discrepancy between simulated and real results in our system to the same micron-level sensitivity to alignment. This challenge is more pronounced in our case, as we use the same SLM for both modulation stages and therefore cannot aggressively bin pixels without greatly sacrificing visual quality. We believe that more precise alignment strategies and improved engineering practices could further reduce this gap, bringing the visual quality of real-world results closer to that observed in simulation.

5.4 Experimental results

This section presents experimental focal stack results captured using our hardware prototype. The results are summarized in Figure 11, which shows two focal stack scenes evaluated under four configurations: a single source and a 4×4 source array, each captured with a single frame and with four-frame time multiplexing. For our hardware prototype, the measured diagonal FoV is approximately 11.35° for both the single-source and 4×4 source-array configurations, while the eyebox size increases from approximately $1.94\,\mathrm{mm} \times 1.94\,\mathrm{mm}$ for a single source to $5.92\,\mathrm{mm} \times 5.92\,\mathrm{mm}$ for the 4×4 source array. This corresponds to an étendue expansion factor of approximately $9.3\times$ when moving from a single source to a 4×4 source array. The SLM phase patterns for all experimental results are optimized in the same way as described for the simulated results in Section 4.4, additionally incorporating the calibrated wavefront aberration corrections and geometric mappings described above.

The single-source far-field configuration exhibits noticeably higher contrast than the results captured using our prototype. This is expected, as the single-source far-field setup does not require the elaborate micron-scale calibration needed by our method. However, this configuration also has a field of view that is less than half of that achieved by our system, and suffers from visible speckle noise. In addition, the single-source far-field results show a prominent bright DC spot, clearly visible in the figure, which arises from the limited diffraction efficiency of the SLM. In our method, this DC component undergoes an additional Fourier transform, causing it to defocus and reduce contrast in the captured results.

The 4×4 source-array results are visibly less noisy than the corresponding single-source results obtained with our system. This improvement is consistent across both single-frame and time-multiplexed captures and can be attributed to incoherent speckle averaging across multiple mutually incoherent sources. Time multiplexing further improves visual quality for both configurations by averaging over multiple independently optimized phase patterns.

Overall, these experimental results demonstrate that the benefits of multisource illumination observed in simulation carry over to real hardware captures. In particular, the

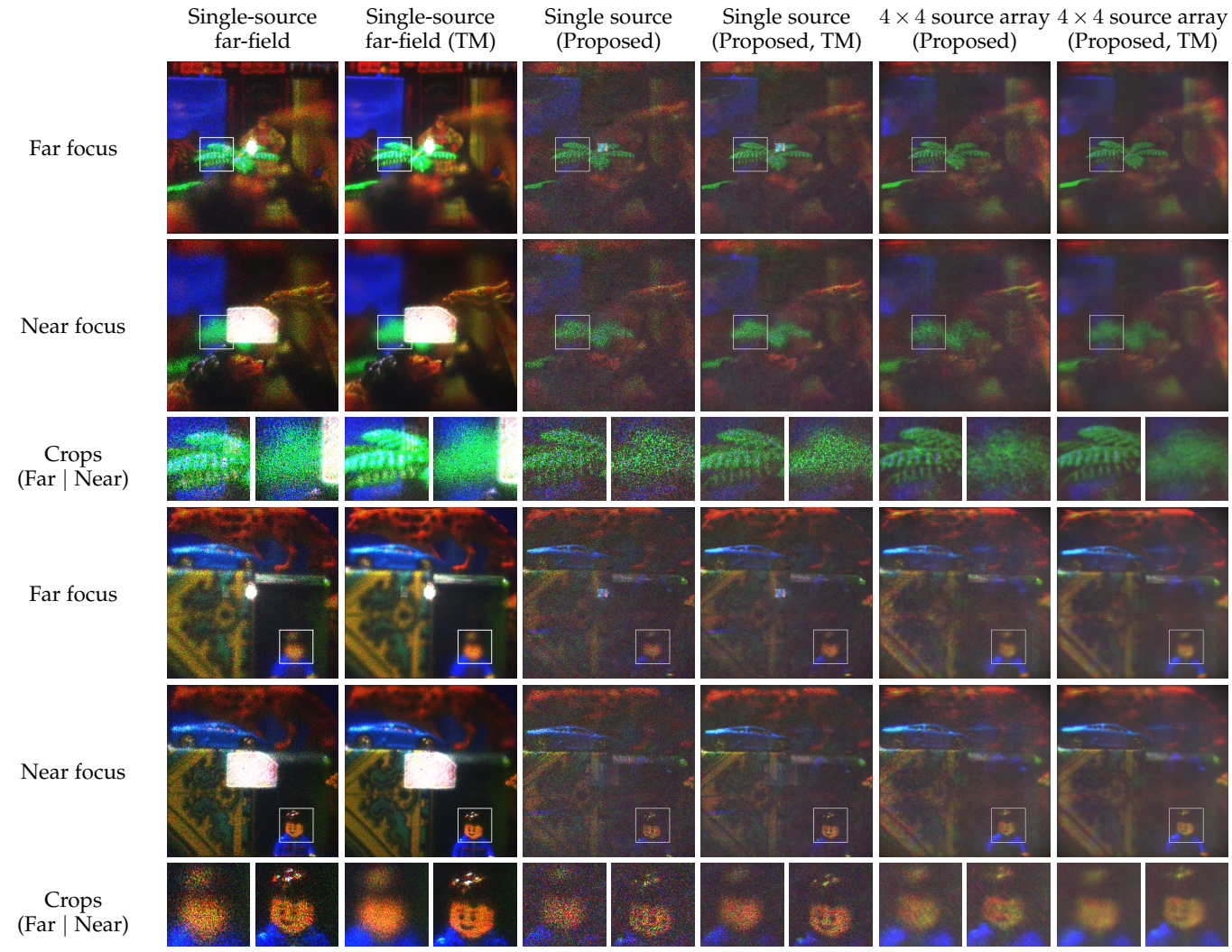


Fig. 11: Qualitative comparison of real captured focal stack results across two scenes, showing far and near focus planes along with zoomed-in crops. We compare a single-source far-field baseline against our proposed system using a single source and a 4×4 source array, each captured with a single frame and with four-frame time multiplexing (TM). Full views show the evaluation region (boxed); crops show zoomed-in details.

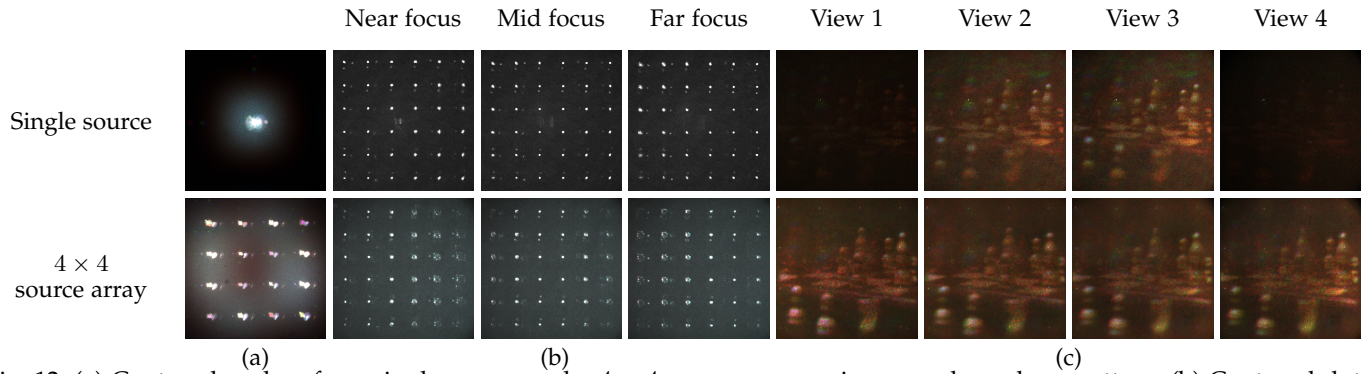


Fig. 12: (a) Captured eyebox for a single source and a 4×4 source array using a random phase pattern. (b) Captured dot grid scene composed of columns at increasing depths. (c) Captured light field scene at four viewpoints for the same two configurations. All captures use a $3.5\,\mathrm{mm}$ aperture (adjustable iris on a translation stage) placed in front of a static camera.

4×4 source array provides reduced speckle noise and a larger eyebox compared to a single source. Additional experimentally captured focal stack results are provided in the supplementary material.

5.5 Eyebox and view-dependence analysis

The primary motivation for introducing multiple illumination sources is to increase the extent of the eyebox. Figure 12(a) shows that the eyebox extent increases sub-

stantially when moving from a single source to a 4×4 source array. This expanded eyebox manifests as a more pronounced defocus blur produced by the system. This effect is evident in the zoomed-in crops in Figure 11, particularly for the first scene. Specifically, in Figure 12(b), which shows a dot-grid scene with each column placed at increasing depths, the difference in defocus behavior between the two configurations is clearly observed. Upon closer inspection, the defocus pattern in the 4×4 source-array configuration resembles the spatial arrangement of the 4×4 source array itself, directly illustrating the dependence of defocus blur size on the extent of the eyebox.

The other implications of increasing the eyebox extent are improved robustness to pupil shifts and the emergence of view-dependent effects. This is demonstrated in Figure 12(c). For the left and right viewpoints, the pupil samples only a small portion of the eyebox in the single-source configuration, causing the scene to be barely visible. In contrast, for the 4×4 source array, the pupil remains within the eyebox extent, and the scene remains clearly visible. Moreover, for the 4×4 source array, as the pupil moves from left to right, different parts of the scene shift relative to one another, clearly exhibiting view-dependent behavior. This highlights the importance of the second modulation stage, which decorrelates wavefronts at the eyebox plane and enables view-dependent effects. Additional experimentally captured light field results are provided in the supplementary material.

6 Limitations and Conclusion

In this work, we present a compact design for implementing multisource holography with multiple planes of modulation to break the memory effect, all while using a single reflective phase SLM. Unlike much prior work in this space, which, despite conceptually compact designs, results in extended tabletop prototypes, we show that folded optics enables us to repurpose a single SLM and achieve results comparable to the state of the art.

Our proposed work has a number of limitations that arise from practical limitations of off-the-shelf devices. Specifically, we require high-contrast reflective polarizers and high-precision quarter-wave plates to ensure purity of polarization states in the folded configuration. The reflective wire-grid polarizer and achromatic quarter-wave plate used in our prototype are imperfect, causing light from the first interaction with the SLM to leak into the final image and reducing contrast. Our design also requires polarization-insensitive phase modulators; the only commercially available model has a large pixel pitch and a low bit depth of 4 bits, both of which reduce the quality of images generated by our prototype. Perhaps the biggest limitation of our approach is the need for extremely precise calibration, where the wavefront associated with each source must be aligned to the SLM at the second modulation stage to sub-pixel precision. This contributes to a gap between simulated and real results that is common across prior multisource holographic designs; Chao et al. [1] report a similar gap despite their pixel-binning strategy. We explored CITL-style learned calibration strategies [33] to learn source-dependent aberrations and micron-scale per-source warps directly from captured data, but observed limited improvement. We believe this is because the calibration problem requires jointly estimating source-dependent phase aberrations and spatially-varying sub-pixel alignment at the second modulation stage, while image quality drops precipitously within only a few microns of the optimum. Outside this small neighborhood, the loss landscape does not appear to provide a useful supervisory signal for calibration. Closing this gap may require dedicated calibration methods capable of robustly modeling spatially-varying sub-pixel warps and source-dependent aberrations jointly, which we see as a promising direction for future work that would benefit prior multisource designs as well. Finally, our folded multisource modulation strategy is complementary to recent waveguide holography systems for compact AR/MR displays [17], [18]; integrating it with such architectures could remove the spacing constraint imposed by our current free-space Fourier lens.

Acknowledgments

SG gratefully acknowledges support from the William J. Happel Fellowship in Electrical & Computer Engineering. PC is supported by DARPA Program No. HR0011262E003. The views contained in this article are those of the authors and not of the funding agency. The authors gratefully acknowledge Matthew O'Toole and Dorian Chan for valuable insights and help in calibrating the PLM.

References

[1] B. Chao, M. Gopakumar, S. Choi, J. Kim, L. Shi, and G. Wetzstein, "Large étendue 3d holographic display with content-adaptive dynamic fourier modulation," in *SIGGRAPH Asia 2024 Conference Papers*, 2024, pp. 1–12.

[2] Y. Jo, D. Yoo, D. Lee, M. Kim, and B. Lee, "Multi-illumination 3d holographic display using a binary mask," *Opt. Lett.*, vol. 47, no. 10, pp. 2482–2485, May 2022. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-47-10-2482

[3] E. Buckley, A. Cable, N. Lawrence, and T. Wilkinson, "Viewing angle enhancement for two-and three-dimensional holographic displays with random superresolution phase masks," *Applied optics*, vol. 45, no. 28, pp. 7334–7341, 2006.

[4] G. Kuo, L. Waller, R. Ng, and A. Maimone, "High resolution étendue expansion for holographic displays," *ACM Trans. Graph.*, vol. 39, no. 4, Aug. 2020. [Online]. Available: https://doi.org/10.1145/3386569.3392414

[5] S. Monin, A. C. Sankaranarayanan, and A. Levin, "Analyzing phase masks for wide étendue holographic displays," in *2022 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2022, pp. 1–12.

[6] E. Tseng, G. Kuo, S.-H. Baek, N. Matsuda, A. Maimone, F. Schiffers, P. Chakravarthula, Q. Fu, W. Heidrich, D. Lanman *et al.*, "Neural étendue expander for ultra-wide-angle high-fidelity holographic display," *Nature communications*, vol. 15, no. 1, p. 2907, 2024.

[7] J.-H. Park and B. Lee, "Holographic techniques for augmented reality and virtual reality near-eye displays," *Light: Advanced Manufacturing*, vol. 3, no. 1, pp. 137–150, 2022.

[8] G. Kuo, F. Schiffers, D. Lanman, O. Cossairt, and N. Matsuda, "Multisource holography," *ACM Trans. Graph.*, vol. 42, no. 6, Dec. 2023. [Online]. Available: https://doi.org/10.1145/3618380

[9] B. Javidi, A. Carnicer, A. Anand, G. Barbastathis, W. Chen, P. Ferraro, J. Goodman, R. Horisaki, K. Khare, M. Kujawinska *et al.*, "Roadmap on digital holography," *Optics express*, vol. 29, no. 22, pp. 35 078–35 118, 2021.

[10] D. Pi, J. Liu, and Y. Wang, "Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display," *Light: Science & Applications*, vol. 11, no. 1, p. 231, 2022.

[11] Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural holography with camera-in-the-loop training," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 6, p. 185, 2020.

[12] L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3d holography with deep neural networks," *Nature*, vol. 591, no. 7849, pp. 234–239, 2021.

[13] P. Chakravarthula, Y. Peng, J. Kollin, H. Fuchs, and F. Heide, "Wirtinger holography for near-eye displays," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, p. 213, 2019.

[14] B. Chao, M. Gopakumar, S. Choi, and G. Wetzstein, "High-brightness holographic projection," *Optics Letters*, vol. 48, no. 15, pp. 4041–4044, 2023.

[15] K. Kavaklı, L. Shi, H. Urey, W. Matusik, and K. Akşit, "Multi-color holograms improve brightness in holographic displays," in *SIGGRAPH Asia 2023 Conference Papers*, 2023, pp. 1–11.

[16] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 85, 2017.

[17] M. Gopakumar, G.-Y. Lee, S. Choi, B. Chao, Y. Peng, J. Kim, and G. Wetzstein, "Full-colour 3d holographic augmented-reality displays with metasurface waveguides," *Nature*, vol. 629, no. 8013, pp. 791–797, 2024.

[18] S. Choi, C. Jang, D. Lanman, and G. Wetzstein, "Synthetic aperture waveguide holography for compact mixed-reality displays with large étendue," *Nature Photonics*, vol. 19, no. 8, pp. 854–863, 2025.

[19] J. Park, K. Lee, and Y. Park, "Ultrathin wide-angle large-area digital holography using a photon sieve," *Nature Communications*, vol. 10, no. 1, pp. 1–8, 2019.

[20] S. Monin, A. C. Sankaranarayanan, and A. Levin, "Exponentially-wide étendue displays using a tilting cascade," in *2022 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2022, pp. 1–12.

[21] C. Jang, K. Bang, G. Li, and B. Lee, "Holographic near-eye display with expanded eye-box," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–14, 2018.

[22] C. Chang, W. Cui, J. Park, and L. Gao, "Computational holographic Maxwellian near-eye display with an expanded eyebox," *Scientific Reports*, vol. 9, no. 1, p. 18749, 2019.

[23] M. Chae, K. Bang, D. Yoo, and Y. Jeong, "Étendue expansion in holographic near eye displays through sparse eye-box generation using lens array eyepiece," *ACM Transactions on Graphics (TOG)*, vol. 42, no. 4, pp. 1–13, 2023.

[24] D. Lee, K. Bang, S.-W. Nam, B. Lee, D. Kim, and B. Lee, "Expanding energy envelope in holographic display via mutually coherent multi-directional illumination," *Scientific Reports*, vol. 12, no. 1, p. 6649, 2022.

[25] R. Watanabe, S. Kang, and Y. Sakamoto, "Binocular holographic display with a wide viewing zone using eye-tracking and multiple reconstruction lights," *Applied Optics*, vol. 64, no. 4, pp. 984–994, 2025.

[26] X. Xia, P. Mi, Y. Tao, X. Meng, W. Zhou, Y. Yu, and Y. Peng, "Multi-illumination-interfered neural holography with expanded eyebox," *IEEE Transactions on Visualization and Computer Graphics*, vol. 31, no. 11, pp. 9667–9676, 2025.

[27] F. A. Schiffers, G. Kuo, N. Matsuda, D. Lanman, and O. Cossairt, "HoloChrome: Polychromatic illumination for speckle reduction in holographic near-eye displays," *ACM Transactions on Graphics*, vol. 44, no. 3, pp. 1–18, 2025.

[28] B. A. Narasimhan, "Ultra-compact pancake optics based on ThinEyes super-resolution technology for virtual reality headsets," in *Digital Optics for Immersive Displays*, vol. 10676. SPIE, 2018, pp. 359–366.

[29] A. Maimone and J. Wang, "Holographic optics for thin and lightweight virtual reality," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, pp. 67:1–67:14, 2020.

[30] K. Bang, Y. Jo, M. Chae, and B. Lee, "Lenslet VR: Thin, flat and wide-FOV virtual reality display using Fresnel lens and lenslet array," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2545–2554, 2021.

[31] O. Cakmakci, Y. Qin, P. Bosel, and G. Wetzstein, "Holographic pancake optics for thin and lightweight optical see-through augmented reality," *Optics Express*, vol. 29, no. 22, pp. 35 206–35 215, 2021.

[32] S. Choi, M. Gopakumar, Y. Peng, J. Kim, M. O'Toole, and G. Wetzstein, "Time-multiplexed neural holography: a flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators," in *ACM SIGGRAPH 2022 Conference Proceedings*, 2022, pp. 1–9.

[33] Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural Holography with Camera-in-the-loop Training," *ACM Trans. Graph. (SIGGRAPH Asia)*, 2020.

Sagnik Ghosh is a Ph.D. candidate in Electrical and Computer Engineering at Carnegie Mellon University. He received his B.Tech in Electrical Engineering from the Indian Institute of Technology Madras in 2023. He received the William J. Happel Fellowship in Electrical & Computer Engineering in 2025. His research interests span AR/VR and computational imaging, with a focus on wide-field-of-view and holographic near-eye displays.

Praneeth Chakravarthula is an Assistant Professor at the University of North Carolina at Chapel Hill. His research focuses on novel computational imaging and display systems, at the intersection of optics, perception, computer graphics, optimization, and machine learning. Dr. Chakravarthula is a Senior Member of IEEE and Optica, recipient of the IEEE VR Best Dissertation Award, as well as multiple Best Paper and Demo Awards at premier venues such as ACM SIGGRAPH, IEEE VR, and ISMAR. He completed his postdoctoral research at Princeton University, received his Ph.D. from UNC Chapel Hill, and holds B.Tech and M.Tech degrees in Electrical Engineering with a specialization in Signal Processing from the Indian Institute of Technology (IIT) Madras.

Aswin C. Sankaranarayanan is a professor in the ECE department at CMU, where he is the PI of the Image Science Lab. His research interests are broadly in compressive sensing, computational photography, signal processing, and machine vision. His doctoral research was at the University of Maryland, where his dissertation won the distinguished dissertation award from the ECE department in 2009. Aswin is the recipient of multiple best paper awards, including at SIGGRAPH 2023 and CVPR 2019, the NSF CAREER award, and the Eta Kappa Nu (CMU Chapter) Excellence in Teaching award.