

Multisource Holography: Supplemental Material

GRACE KUO, Reality Labs Research, Meta, USA
 FLORIAN SCHIFFERS, Reality Labs Research, Meta, USA
 DOUGLAS LANMAN, Reality Labs Research, Meta, USA
 OLIVER COSSAIRT, Reality Labs Research, Meta, USA
 NATHAN MATSUDA, Reality Labs Research, Meta, USA

This supplemental material includes additional derivations, implementation details, and experimental results. Please also see the supplemental videos for visualizations of the focal stack experiments and simulations.

S1 DERIVATION: SHIFT INVARIANCE WITH 1 SLM

Equation 9 in the main text describes the relationship between a hologram with on-axis illumination and one with a tilted plane wave illumination, specifically that the tilted plane wave results in a translation. Here, we provide a derivation of this relationship.

Instead of the angular spectrum (ASM) propagation kernel defined in the main text, here we use Fresnel propagation, which is the paraxial approximation to ASM. Fresnel propagation can be written as

$$\mathcal{P}_z\{\mathbf{s}(\vec{x})\} = \mathcal{F}^{-1}\{\mathcal{F}\{\mathbf{s}(\vec{x})\} \odot \mathcal{H}_z(\vec{u})\}, \quad (1)$$

$$\mathcal{H}_z(\vec{u}) = \exp(j\pi\lambda z\|\vec{u}\|^2). \quad (2)$$

As in the main text, we define $\mathbf{g}_{z,0}(\vec{x})$ to be the electric field after propagating a distance z given on-axis illumination:

$$\mathbf{g}_{z,0}(\vec{x}) = \mathcal{P}_z\{\mathbf{s}(\vec{x})\} \quad (3)$$

Similarly, $\mathbf{g}_{z,m_i}(\vec{x})$ is the electric field after propagating distance z given tilted plane wave illumination with slope \vec{m}_i :

$$\mathbf{g}_{z,m_i}(\vec{x}) = \mathcal{P}_z\{\mathbf{p}(\vec{x}; \vec{m}_i) \odot \mathbf{s}(\vec{x})\} \quad (4)$$

$$\mathbf{p}(\vec{x}; \vec{m}_i) = e^{j(\vec{x} \cdot \vec{m}_i)}. \quad (5)$$

For notational convenience, we let $\mathcal{G}_{z,0}(\vec{u})$, $\mathcal{G}_{z,m_i}(\vec{u})$, and $\mathcal{S}(\vec{u})$ be the Fourier transforms of $\mathbf{g}_{z,0}(\vec{x})$, $\mathbf{g}_{z,m_i}(\vec{x})$, and $\mathbf{s}(\vec{x})$, respectively.

Combining Eqs. 1, 4, and 5 yields:

$$\mathbf{g}_{z,m_i}(\vec{x}) = \mathcal{P}_z\{\mathbf{s}(\vec{x})e^{j(\vec{x} \cdot \vec{m}_i)}\} \quad (6)$$

$$\mathbf{g}_{z,m_i}(\vec{x}) = \mathcal{F}^{-1}\{\mathcal{F}\{\mathbf{s}(\vec{x})e^{j(\vec{x} \cdot \vec{m}_i)}\} \odot \mathcal{H}_z(\vec{u})\} \quad (7)$$

$$\mathcal{G}_{z,m_i}(\vec{u}) = \mathcal{F}\{\mathbf{s}(\vec{x})e^{j(\vec{x} \cdot \vec{m}_i)}\} \odot \mathcal{H}_z(\vec{u}) \quad (8)$$

$$\mathcal{G}_{z,m_i}(\vec{u}) = \mathcal{S}(\vec{u} - \frac{\vec{m}_i}{2\pi}) \odot \mathcal{H}_z(\vec{u}) \quad (9)$$

$$\mathcal{G}_{z,m_i}(\vec{u}) = (\mathcal{S}(\vec{u}) \odot \mathcal{H}_z(\vec{u} + \frac{\vec{m}_i}{2\pi})) * \delta(\vec{u} - \frac{\vec{m}_i}{2\pi}) \quad (10)$$

Authors' addresses: Grace Kuo, Reality Labs Research, Meta, USA; Florian Schiffers, Reality Labs Research, Meta, USA; Douglas Lanman, Reality Labs Research, Meta, USA; Oliver Cossairt, Reality Labs Research, Meta, USA; Nathan Matsuda, Reality Labs Research, Meta, USA.

Next, we use Eq. 2 to simplify the propagation kernel term in Eq. 10:

$$\mathcal{H}_z\left(\vec{u} + \frac{\vec{m}_i}{2\pi}\right) = \exp\left(j\pi\lambda z\|\vec{u} + \frac{\vec{m}_i}{2\pi}\|^2\right) \quad (11)$$

$$= \exp\left(j\pi\lambda z\left(\|\vec{u}\|^2 + \frac{1}{\pi}\vec{u} \cdot \vec{m}_i + \frac{1}{4\pi^2}\|\vec{m}_i\|^2\right)\right) \quad (12)$$

$$= \mathcal{H}_z(\vec{u}) \odot \exp(j\lambda z\vec{u} \cdot \vec{m}_i) \odot c \quad (13)$$

where $c = \exp(j\lambda z\|\vec{m}_i\|^2/4\pi)$ is a constant phase offset. As constant phase offsets cannot be measured, we drop c going forward.

Substituting Eq. 13 back into Eq. 10 gives:

$$\mathcal{G}_{z,m_i}(\vec{u}) = (\mathcal{S}(\vec{u}) \odot \mathcal{H}_z(\vec{u}) \odot \exp(j\lambda z\vec{u} \cdot \vec{m}_i)) * \delta(\vec{u} - \frac{\vec{m}_i}{2\pi}) \quad (14)$$

$$= (\mathcal{G}_{z,0}(\vec{u}) \odot \exp(j\lambda z\vec{u} \cdot \vec{m}_i)) * \delta(\vec{u} - \frac{\vec{m}_i}{2\pi}) \quad (15)$$

where we've taken advantage of the relationship $\mathcal{G}_{z,0}(\vec{u}) = \mathcal{S}(\vec{u}) \odot \mathcal{H}_z(\vec{u})$. Finally, we take the inverse Fourier transform of both sides to get:

$$\mathbf{g}_{z,m_i}(\vec{x}) = \left(\mathbf{g}_{z,0}(\vec{x}) * \delta\left(\vec{x} + \frac{\lambda z}{2\pi}\vec{m}_i\right)\right) \odot e^{j(\vec{x} \cdot \vec{m}_i)} \quad (16)$$

$$= \mathbf{g}_{z,0}\left(\vec{x} + \frac{\lambda z}{2\pi}\vec{m}_i\right) \odot e^{j(\vec{x} \cdot \vec{m}_i)} \quad (17)$$

This gives the relationship in the main text showing that the electric field is a shifted version of that on-axis, up to a carrier wave, $e^{j(\vec{x} \cdot \vec{m}_i)}$.

S2 DERIVATION: FOCAL STACK BLUR PARAMETERS

The amount of blur needed for a holographic display depends on how the system is configured, particularly the focal length of the eyepiece which sets the field of view and eyebox size of the display. If f is the focal length of the eyepiece, then

$$\text{FoV} = 2 \tan^{-1}\left(\frac{w}{2f}\right), \quad (18)$$

$$\text{eyebox} = 2f \tan(\theta), \quad (19)$$

where w is the SLM width and θ is the maximum diffraction angle defined as

$$\theta = \sin^{-1}\left(\frac{\lambda}{2p}\right), \quad (20)$$

for an SLM with pixel pitch p . For our display, we assume an 8 μm pixel, matching the Holoeye Pluto-2.1 SLM used in our experiments.

Since we target natural defocus blur that matches what the user would see in the real world, we set the focal length of the eyepiece such that the eyebox size is similar to the size of a human pupil,

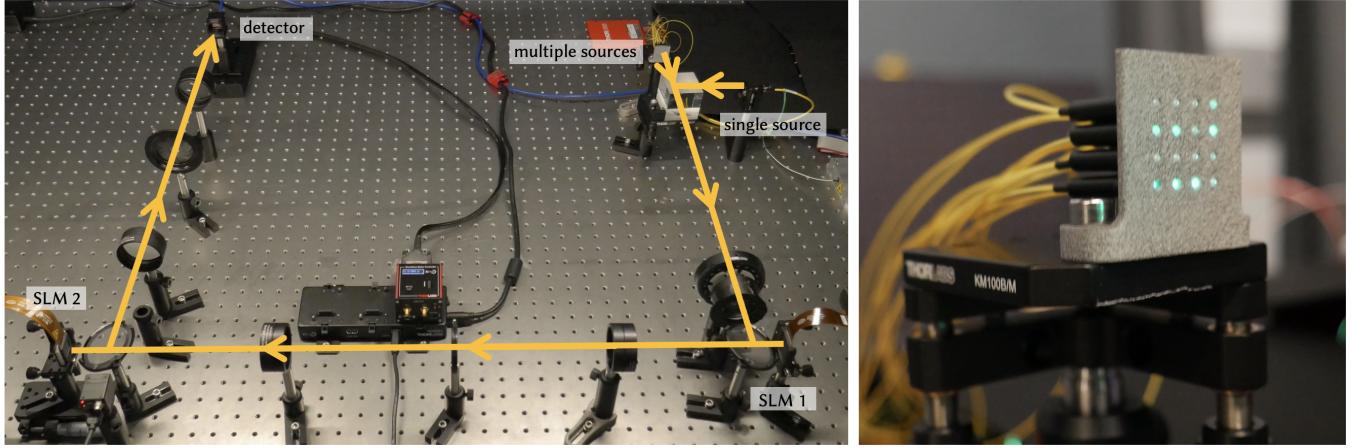


Fig. S1. Experimental Prototype Photos: (Left) Photograph of benchtop experimental setup with beam path drawn. (Right) Close-up of the experimental multisource illumination module. This consists of an array of 4×4 sources that were split from a single SLED. The fiber tips are held in place by a 3D printed mount.

chosen to be 3.5 mm. Assuming a nominal wavelength of 520 nm, this yields an eyepiece focal length of $f = 53.8$ mm.

We target a focal range of 4 diopters, from optical infinity to 25 cm to enable proper focal cues for near objects within arms reach. Based on the thin lens equation, this requires 9.5 mm of axial range at the SLM. We round to 10 mm for our experiments and simulations, which corresponds to a volume from 23.4 cm to optical infinity.

To determine the amount of blur to render, we use the maximum diffraction angle, which defines the cone of light that focuses to a point in the image plane when the eyebox is uniformly filled. Based on the geometry of this cone, the circle of confusion diameter in units of pixels is described by

$$c = 2 \frac{\delta z}{p} \tan(\theta), \quad (21)$$

where δz is the axial distance from the focal plane.

Once again, using a nominal wavelength of 520 nm, the maximum diffraction angle is $\theta = 1.86^\circ$, which yields a blur diameter of 8.1 pixels per millimeter of defocus. We use these parameters for all of our target focal stacks.

S3 SIMULATION IMPLEMENTATION DETAILS

The equations in the main paper describing the image formation model are written in continuous space, but when solving for the SLM patterns, we need to discretize. We choose to conduct simulations at $2 \times$ the resolution of the SLM in each direction (i.e. there are 2×2 pixels in the simulation for each SLM pixel) to prevent aliasing. Aliasing can occur without upsampling because the maximum spatial frequency of the signal doubles when taking the absolute value squared to get intensity. Therefore, a complex signal at the maximum frequency the SLM can produce (for example: +1, -1, +1, -1, ...) cannot be represented after converting to intensity (now: +1, +1, +1, ...). To maintain that spatial frequency in the final intensity image, the original complex signal needs to be sampled at $2 \times$ the SLM resolution.

Aliasing from this effect is usually ignored in holographic display papers, and the effect is not usually substantial. However, we found with multisource holography that the aliasing can make it appear as though there is an angularly-varying response with only 1 SLM (i.e. Eq. 9 in the main text is not true with aliasing). Since the aliasing isn't physical, this can lead to incorrect conclusions, particularly for closely spaced sources. Therefore, we upsample in all of our simulations.

However, upsampling makes higher orders apparent in the simulation, but we assume these will be filtered out in the final setup. Therefore, in our simulations, we implement a low pass filter that removes the higher orders after each SLM. In addition, we would like our target image to have the same number of degrees of freedom as the SLM to avoid having a poorly posed optimization problem. Therefore, after taking the intensity, we downsample the model output before calculating the loss.

We note that it might seem redundant to upsample and then later downsample, but the model output is slightly different. This is most important for simulating accurate behavior when the sources are spaced closely together; without upsampling, sources close together do not create a blurry output due to the non-physical aliasing, which makes this situation appear better than it is in practice.

S4 EXPERIMENT PHOTOS

Figure S1 shows photographs taken of our experimental benchtop prototype. As described in the main text, the prototype uses two SLMs. There is a $4f$ system between the SLMs and a second $4f$ system between the final SLM and the detector. Irises at the Fourier plane of each $4f$ system block higher orders. Although we collect all our calibration data with the final camera, another detector in the front left corner can optionally be used to collect calibration data before the beam interacts with the second SLM.

The right image of Fig. S1 shows a close-up of the multisource illumination module, which is made of 16 fiber tips held together in a 3D printed housing. As described in the main text, we use a low

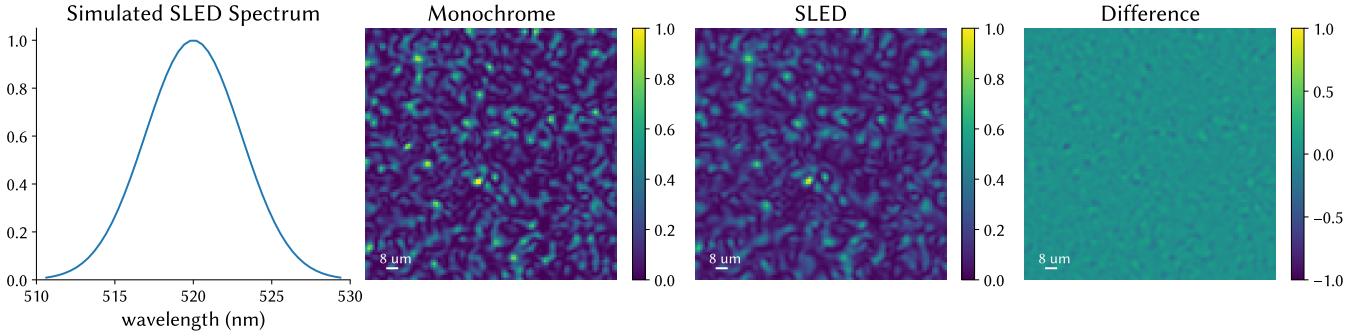


Fig. S2. Effect of SLED on Speckle: Speckle was generated by simulating a random phase distribution on the SLM and propagating by 20 mm, and we show this speckle under both monochromatic illumination and SLED illumination. To simulate the SLED, we assume a Gaussian spectral profile (shown on the left) and sum over 50 discrete wavelength samples. The broader bandwidth of the SLED creates slight blurring of the intensity compared to monochromatic light.



Fig. S3. Effect of SLED on Hologram: Here we simulate a 2D hologram optimized assuming monochromatic illumination. (Left) Simulation of the optimized SLM pattern displayed with ideal monochromatic illumination. In this case there is no model mismatch and the reconstruction matches the target well. (Right) Simulation of the optimized SLM pattern displayed with the SLED illumination. Since the pattern is not optimized with the SLED bandwidth, the mismatch creates artifacts and reduces the PSNR compared to the ideal case. We expect similar artifacts could be present in our experimental results since we do not model the SLED during optimization. However, some artifacts may be diminished through our active camera-in-the-loop calibration.

coherence SLED as the light source and split the output into the 16 fibers. Due to the low coherence, the outputs from each fiber are incoherent with each other, as required by our method.

S5 MODELING SPECTRAL BANDWIDTH

Section 3 of the main text assumes that each source is monochromatic but in our experimental setup we use a SLED which has nonnegligible bandwidth. Here, we show how our model can be extended to sources with broader spectral bandwidth. Given a source with emission spectrum, $q(\lambda)$, we can describe the intensity at the image plane as

$$I_{z,q}(\vec{x}) = \int I_z(\vec{x})q(\lambda)d\lambda, \quad (22)$$

where $I_z(\vec{x})$ is the intensity assuming monochromatic illumination defined in the following equations of the main text: Eq. 1 for single source, Eq. 8 for multisource with one SLM, and Eq. 11 for multisource with two modulators. Here we assume the spectrum is normalized such that $\int q(\lambda)d\lambda = 1$.

To calculate Eq. 22 in practice, one can take the approach of Peng et al. [2021] and discretize the continuous emission spectrum:

$$I_{z,q}(\vec{x}) \approx \sum_n I_z(\vec{x})q\left(\lambda^{(n)}\right), \quad (23)$$

where $\lambda^{(n)}$ is a discrete wavelength sample. For computational efficiency, a subset of discrete wavelengths can be stochastically sampled during each iteration. However, this still increases computational cost. Since multisource holography is already a computationally intensive method, in our results we chose to optimize assuming monochromatic light.

Figures S2 and S3 illustrate the effect of the SLED on speckle and optimized holograms. In Fig. S2, we simulate a random phase pattern on the SLM with both monochromatic and SLED illumination, and generate speckle by propagating the beam by 20 mm. For the SLED, we assume a Gaussian spectrum with standard deviation of 3 nm. We simulate the SLED bandwidth using Eq. 23 with 50 discrete samples over a 19 nm range. To visualize the subpixel differences between the illumination sources, we conduct the simulation at 4×

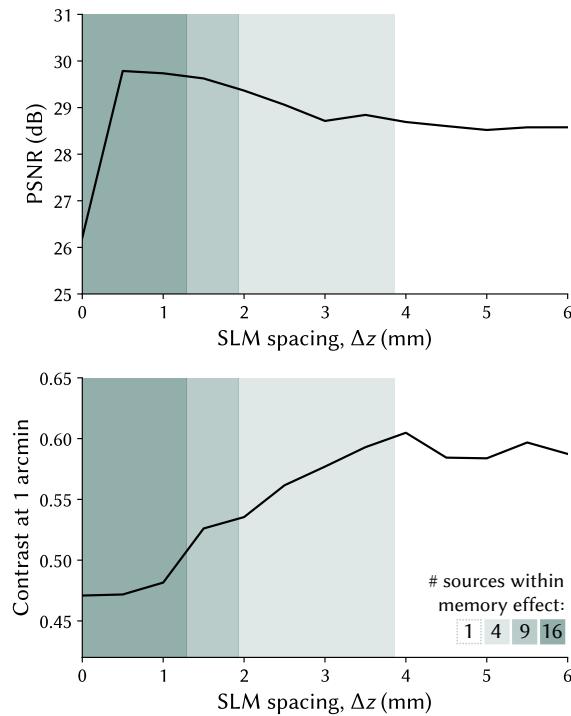


Fig. S4. Analysis of SLM Spacing: In simulation, we calculate PSNR (top) and contrast at 1 arcmin (bottom) as a function of the gap between the two SLMs, Δz . Here, we simulate 4×4 sources spaced $\Delta m = 25$ rad/mm apart. The number of sources within the memory effect region is indicated by the background color of the plot. When the SLMs are closer together, more sources fall within the memory effect region and therefore cannot be controlled independently, resulting in blurring or replicas in the output hologram. Although PSNR does a poor job of capturing this artifact, our contrast metric shows worse high resolution contrast when the SLMs are close together. As Δz increases, contrast rises until all of the sources are outside the memory effect region.

the SLM resolution in each direction. As shown in Fig. S2, the extra bandwidth of the SLED creates a small but visible blur in the speckle image, which is clearly evident in the difference image on the right.

Next, we examine how the SLED impacts a hologram by optimizing an SLM pattern to produce a 2D image at a 20 mm propagation distance. Here, we optimize assuming monochromatic illumination (520 nm) in the forward model. Figure S3 compares the ideal case (reconstruction given ideal monochromatic illumination) to the realistic case with SLED illumination. When the illumination matches the model, the reconstruction does not have noticeable speckle or other artifacts. However, when the SLED is used, there is additional speckle from the model mismatch.

This scenario is analogous to our experiments where we optimize assuming monochromatic light (for computational tractability) but display on a system with SLED illumination. Therefore, we expect there may be artifacts in the experimental results due to the SLED, and modeling the SLED in the optimization (using Eq. 23) could improve the experimental results. However, since we use active

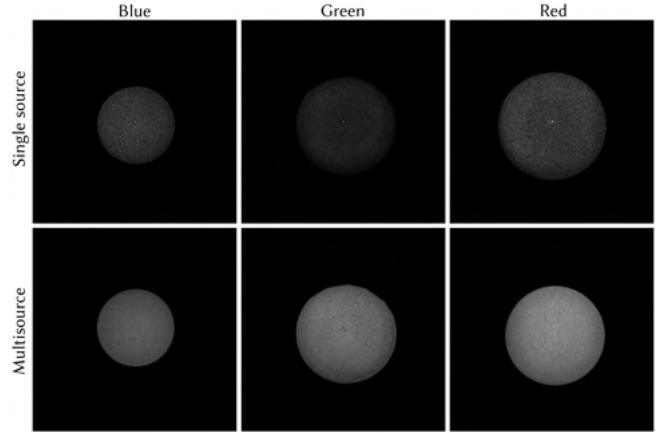


Fig. S5. Simulated Eyebox of Experimental SLM Patterns: Here we show the simulated eyebox for both single and multisource holography that is created when we display the two SLM patterns. Except for the DC-term which is almost unavoidable due to field fringing, the eyebox is uniformly filled in both experiments. The multisource holography eyebox appears smoother due to its reduced speckle.

camera-in-the-loop (ACiTL) in our experiment, some of these artifacts may have been corrected as ACiTL uses the true experimental system when updating the SLM pattern and can therefore account for the SLED. We also note that the SLED was used for both our multisource and single source experiments, so any artifacts due to illumination will appear in all results.

S6 ANALYSIS OF SLM SPACING

In Section 4.3 of the main text, we analyze the impact of the number of sources and source spacing on multisource holography. Here, we provide a similar analysis on the spacing between the SLMs, Δz , shown in Fig. S4. As in the main text, we compute both PSNR and contrast at 1 arcmin (the highest spatial frequency we can show with the SLM, with a period of 2 SLM pixels). We fix the number of sources at a 4×4 grid and we set the spacing between sources to be 25 rad/mm. Based on Eq. 13 of the main text, we calculate the number of sources within the memory effect region, which is indicated by the background color in Fig. S4.

As with our analysis of source spacing in the main text, we expect that when sources are within the memory effect region of the two SLMs, they cannot be controlled independently, making the image generation problem poorly posed. However, in this scenario, PSNR can be a misleading metric since replicas caused by the different sources still reduce speckle (which is heavily penalized by PSNR). PSNR is not sensitive to the blurring due to the different sources, resulting in a higher PSNR when the SLMs are close together. However, this blurring is undesirable since it prevents us from producing high resolution features in the image. As in the main text, our contrast metric captures this trend, with lower contrast when more sources are within the memory effect region. Once all the sources are outside the memory effect, the maximum contrast is reached.

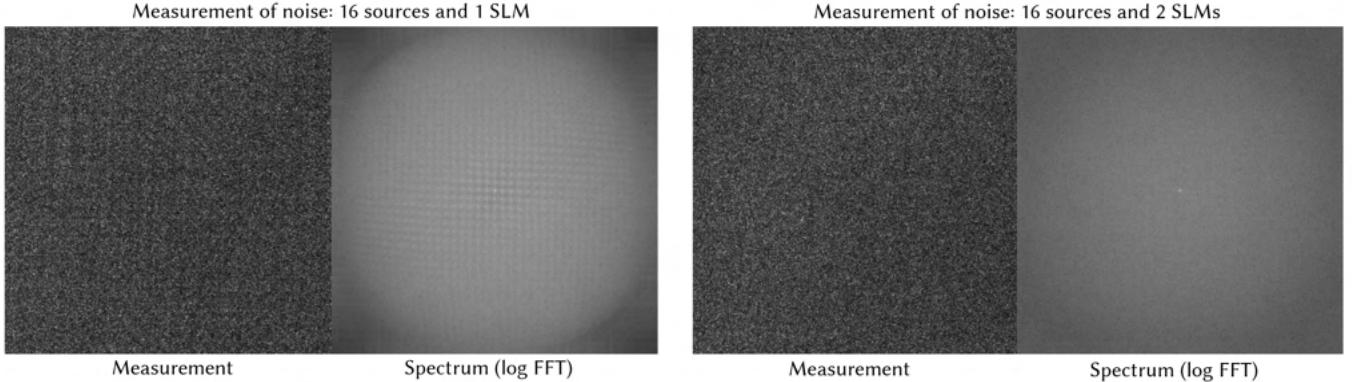


Fig. S6. Breaking the Memory Effect: Both images show the experimental measurement when uniform white noise is displayed on the SLMs using our multisource setup with 4×4 sources. For the left image, only the first SLM was modulated, and the second SLM was kept constant, simulating the case with only one SLM. In this scenario there is structure in the noise from the locations of the sources. This structure is even more evident in the spectral plot which shows a grid pattern. In contrast, for the right image, both SLMs were active and displaying uniform random noise, eliminating the pattern in the measurement and its Fourier transform.

S7 ADDITIONAL EXAMPLES

S7.1 Full Bandwidth Holograms

To demonstrate that our experimental holograms are truly random phase, we show at the eyebox which is calculated using our calibrated model. Figure S5 illustrates the simulated eyebox for the flower scene of Fig. 7 in the main test, comparing both multisource and single source holography. These eyeboxes are computed by passing the SLM patterns through the physically-calibrated model until the final aperture is reached. At this point, we calculate the magnitude of the eyebox, and for the multisource setup, an average is calculated from all 16 modes.

The simulation reveals a near-uniform distribution of the eyebox across the entire available area, thereby confirming the SLM patterns are random phase. It is worth noting the visibility of the DC peak in the eyebox, with a single peak for the single source setup and one peak per source for the multisource setup. This peak is predominantly attributable to field fringing in the liquid crystal-on-silicon (LCoS) modulator. Although we account for field fringing in the model, it still creates unmodulated light that contributes to the DC term.

The DC peak could potentially be eliminated, for instance, by incorporating a DC block into the relay system or using different SLM technologies, such as a MEMS-based devices [Choi et al. 2022; Ouyang et al. 2022]. MEMS SLMs do not exhibit DC peaks resulting from field fringing since they use micro-mirrors to modulate the light instead of liquid crystals.

S7.2 Memory Effect Demonstration

Figure S6 shows an example of how the memory effect manifests in our experimental setup with 16 sources arranged in a grid. On the left, we display a random pattern on just one of the SLMs while the other is kept constant. This mimics the case of a single SLM, and in this scenario, the 16 independent sources create correlated outputs, as explained in Sec. S1. This correlation manifests as recurring structures within the captured images, distinctly visible in the Fourier

transform spectral image of the captured data. However, when both SLMs are displaying random patterns (right), the correlations vanish and the spectrum no longer has a grid structure.

S7.3 Temporal Multiplexing with More Frames

In the main text, we demonstrate that multisource holography outperforms temporal multiplexing with 6 frames, both quantitatively and qualitatively. In Fig. S7 we compare multisource against temporal multiplexing with more frames. Here, we simulate temporal multiplexing assuming a single phase only SLM, and the frames are jointly optimized as described in the main text. For multisource holography, we simulate 4×4 sources to match the experimental setup.

With 16 sources, multisource holography performs similarly to temporal multiplexing with 8 frames. We note that multisource has a quarter of the degrees of freedom compared to temporal multiplexing (1 frame with 2 SLMs vs. 8 frames with 1 SLM), so even though multisource has more incoherent averages (16 vs. 8), the two methods perform similarly. As the number of frames increases, temporal multiplexing begins to exceed multisource quantitatively. This is also visible in the reduced speckle, which is most apparent in smooth regions of the image. However, this speckle reduction comes at the direct cost of temporal bandwidth.

S7.4 Comparison of Configurations in 2D

The 2D results in the main paper were optimized using both SLMs to present a fair comparison in terms of degrees of freedom for single source and multisource. To demonstrate that, in the single source case, this is equivalent to the traditional holography setup with only 1 SLM, we show an example where we optimized for a 2D pattern using only 1 SLM (the second SLM in the system), displayed in Fig. S8 (top row). As described in the main text, multisource shows blurring and ghosts when optimized with only a single SLM.

All our results in the main text use the “active” camera-in-the-loop (ACiTL) approach of Peng et al. [2020] to fine-tune the SLM patterns



Fig. S7. Temporal Multiplexing with More Frames: In simulation, we compare multisource holography with 16 sources to temporal multiplexing with 8, 16, and 32 frames. Multisource holography performs similarly to temporal multiplexing with 8 frames: even though there are more incoherent averages in our approach, multisource holography has fewer degrees of freedom compared to temporal multiplexing, which limits the amount of despeckling. As the number of frames increase, temporal multiplexing outperforms multisource, but at the expense of temporal bandwidth.

for a particular image. Fig. S8 also shows examples without this extra calibration. The results without ACiTL still use the calibrated model, which is fit completely offline and generalizes to new content, but they do not use the online calibration. Although the image quality is better for both single and multisource with ACiTL, it is clear that multisource outperforms single source even without ACiTL. ACiTL is particularly helpful in cleaning up artifacts at the edges of the field of view where the calibrated model is less accurate. We anticipate that improvements to the calibration pipeline could enable the quality of the ACiTL results without active fine-tuning.

Finally, we test the ability of our model to generalize by displaying 2D images at several depths. Our model was calibrated using only data collected at $z = 20$ mm, but, as shown in Fig. S8, the model performs very similarly at $z = 15$ mm and $z = 25$ mm, even without ACiTL. By 30 mm there is some degradation in the images without ACiTL, showing that the model is best within a few millimeters of the calibration plane. However the 10 mm range is sufficient to cover the focal stack of interest.

Close-up crops that better capture speckle are shown for all these comparisons in Fig. S9.

S7.5 Focal Stacks with and without ACiTL

Figures S10, S11, and S12 show experimental comparisons with and without ACiTL on three focal stacks examples. As with the 2D results above, here, all results use the calibrated model that is fit

offline. The ACiTL results are additionally fine-tuned in the loop, using the gradients of the calibrated model for backpropagation. The no ACiTL results are not fine-tuned experimentally.

Once again, ACiTL offers an improvement in image quality, although we point out that the multisource result without ACiTL is higher quality than single source result with ACiTL indicating the magnitude of improvement provided by multisource holography. We again expect that improvements to the offline calibration could enable the quality of the ACiTL results without needing active fine-tuning.

S7.6 Raw Image Captures

To demonstrate that there is no substantial post processing in the results in the main text, we show examples of the raw image captures in Fig. S13, without any alignment or cropping. The only processing is converting the three monochrome captures into one color image. Images in the main text have been warped into the frame of the SLM using the TPS deformation model and the edges are cropped. As you can see in Fig. S13 the image quality is the same without this warping.

S8 CALIBRATED MODEL

Next we discuss some additional details of the calibrated model used for our experimental results, and we visualize the parameters that the model learns. Recall that we calibrate separate models for each

color and each setup type (single source vs. multisource), giving a total of 6 different models that are all independent.

S8.1 Warping between SLMs

A key part of making our 2 SLM setup work in practice is the deformation map between the two SLMs, which we'll call SLM1 and SLM2. This involves mapping the complex field after SLM1 into the coordinate system of SLM2. We implement this warping using the thin plate spline (TPS) model from Kornia [Riba et al. 2020].

To fit the TPS model, we first optimize each SLM pattern separately to generate a grid of dots at the camera. We use an asymmetric dot pattern to ensure that orientation changes due to the beamsplitter (which flips the image left-right) and $4f$ system (which flips the image in both directions) are captured correctly. We detect the dot centers, which are in the camera coordinate space, and fit a TPS model from SLM1 to the camera and a second TPS model from SLM2 to the camera. Next, we use linear algebra to compute the TPS transformation between SLM1 and SLM2. Finally the TPS transform parameters are fine-tuned with gradient descent along side the other parameters in the learned model.

The TPS transformation between SLM1 and SLM2 is visualized in Fig. S14. Note that because of a slight misalignment (there's some rotation and translation), we cannot make use of the full SLM field of view. We therefore restrict our evaluations to the valid region where both SLM1 and SLM2 can modulate the wavefront.

The dot patterns used for fitting the TPS show interesting behavior in the multisource case. Figure S15 shows the captured image with multisource illumination using the same dot patterns and compares these images to our model output. In these images, the relative locations and intensities of the sources are clearly visible, and our model is able to match the measurement well even though this data is very different than the training data used to fit the model.

S8.2 Global Learned Apertures

In the main text, we describe how our model uses a spatially varying aberration function during propagation. In practice, we split this aberration map into two components, a global aperture that is applied to every field location, and a spatially varying component that is applied locally after.

The global aberration function is parameterized by Zernike coefficients (40 coefficients) to describe the phase, and a binary circular aperture where the size and location of the aperture are learned. This parameterization allows us to resize the aperture to any arbitrary crop shape on the fly, allowing our model can be evaluated with different simulation sizes. The global aberration functions are visualized in Fig. S16, where “aperture 1” is the aperture between the two SLMs and “aperture 2” is the aperture between SLM2 and the camera.

S8.3 Spatially Varying Aberrations

To model the influence of spatially varying aberrations, we learn a location-dependent 2D pupil function for both relay systems in addition to the global aperture described above. The total aperture is the product of the two and they could be learned together, but we

choose to learn them separately to provide more flexibility in the software.

The spatially varying aberrations are parameterized with a 128×128 complex valued function that is upsampled with bilinear interpolation to the full pupil size. This provides more total flexibility than the more restrictive parametrization of the global apertures. We can compute aberration maps at arbitrary locations through bilinear interpolation of the learned aperture functions.

In Fig. S17 and S18, we visualize our learned spatially varying aberration functions at a grid of 6×11 pupil positions. We expect that spatially varying aberrations should change smoothly, which we see in the visualization, suggesting that our model is learning something physically meaningful.

S8.4 SLM Lookup Table

Fig. S19 shows the learned lookup table (LUT) for the two SLMs used in the prototype. Each LUT contains 256 values representing the conversion from digital input to phase. Since the two SLMs are not the same, they have different LUTs: the first SLM's LUT is approximately linear, while the second has some non-linear behavior that's captured by our model.

S8.5 Calibration Patterns

Figure S20 shows examples from our training dataset of the SLM patterns and captured images pairs, which we use for calibration. To produce the SLM patterns, we start with uniform random phase and apply a Gaussian kernel with variable standard deviation to generate variable amounts of blur in the SLM pattern. After blurring, SLM patterns are normalized to cover the full phase range (0-255).

The blurring creates images with different feature sizes, which provide diversity of data needed to fit different parameters in the model. For example, the low frequency images, which were blurred with a larger Gaussian kernel, are useful for calibrating the TPS warping. The higher frequency images are needed to calibrate the aberrations, field fringing (cross-talk) kernel, and LUT.

REFERENCES

- Suyeon Choi, Manu Gopakumar, Yifan Peng, Jonghyun Kim, Matthew O’Toole, and Gordon Wetzstein. 2022. 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*. 1–9.
- Bing Ouyang, Sue Gong, Tom Lawrence, and James Hall. 2022. Evaluating Texas instruments Phase Light Modulator (PLM). In *Ocean Sensing and Monitoring XIV*, Vol. 12118. SPIE, 150–157.
- Yifan Peng, Suyeon Choi, Jonghyun Kim, and Gordon Wetzstein. 2021. Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration. *Science Advances* 7, 46 (2021), eabg5040.
- Yifan Peng, Suyeon Choi, Nitish Padmanaban, and Gordon Wetzstein. 2020. Neural holography with camera-in-the-loop training. *ACM Transactions on Graphics (TOG)* 39, 6 (2020), 1–14.
- Edgar Riba, Dmytro Mishkin, Daniel Ponsa, Ethan Rublee, and Gary Bradski. 2020. Kornia: An open source differentiable computer vision library for Pytorch. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3674–3683.

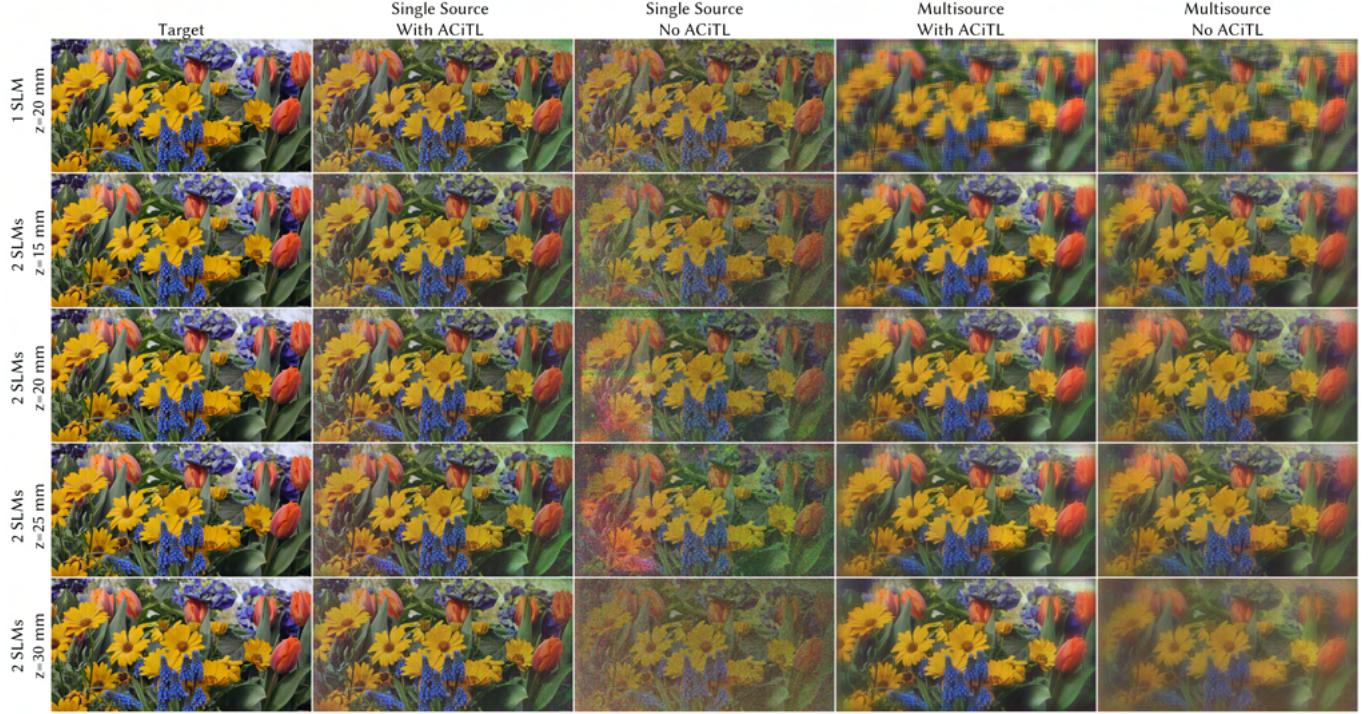


Fig. S8. Experiment (2D Comparison): In this figure, we compare several different setup configurations on 2D image content. The first row shows results where only one SLM is active and the other is kept constant. In particular, you can see that while the single source model can achieve good image quality, the multisource image looks blurry due to the infinite memory effect explained in the main paper. In the last 4 rows, both SLMs are active and content is displayed at different depth planes. Even though the model was only trained for one specific depth (20 mm) it generalizes well to other locations, especially in the desired propagation range of ± 5 mm. Finally, we also show results with and without the active camera-in-the-loop (ACiTL) approach of Peng et al. [2020]. Although ACiTL improves image quality for both single and multisource holograms, it's still evident that multisource results in reduced speckle even without ACiTL. (See Fig. S9 for detailed crops.)

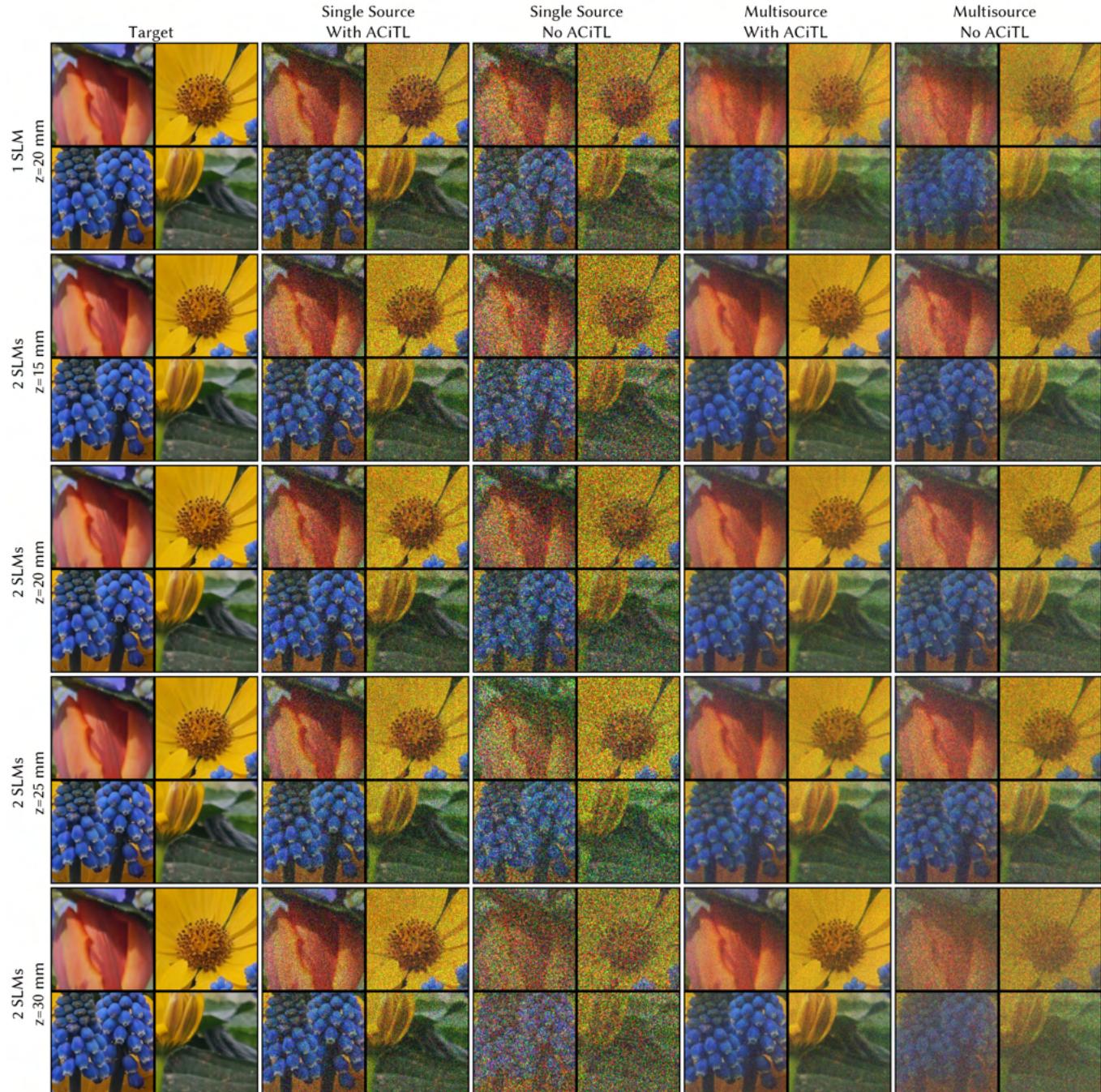


Fig. S9. **Experiment (2D Comparison):** Close-up crops of the same images show in Fig. S8.

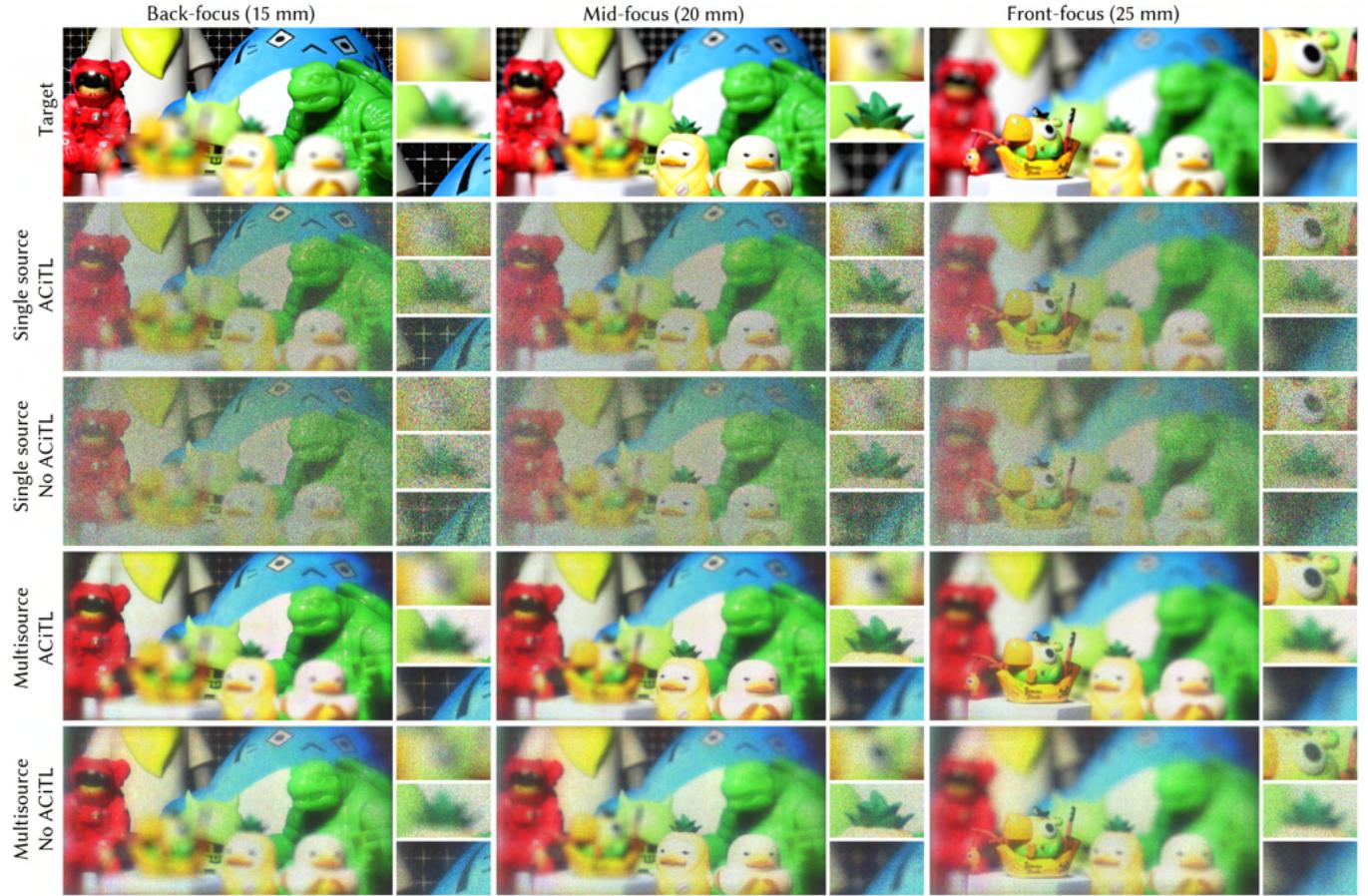


Fig. S10. Experiment (Focal Stack | ACiTL vs Learned Model): This figure shows a focal stack scene with natural defocus blur for both single and multisource optimized with ACiTL (using the proxy gradients of the calibrated model) vs. with the calibrated model only without ACiTL. While image quality with the learned model is already good, the ACiTL shows some clear improvement, especially in terms of noise. However, the multisource experimental capture without ACiTL is less noisy than the single source capture with ACiTL, highlighting the improvement that multisource holography provides.

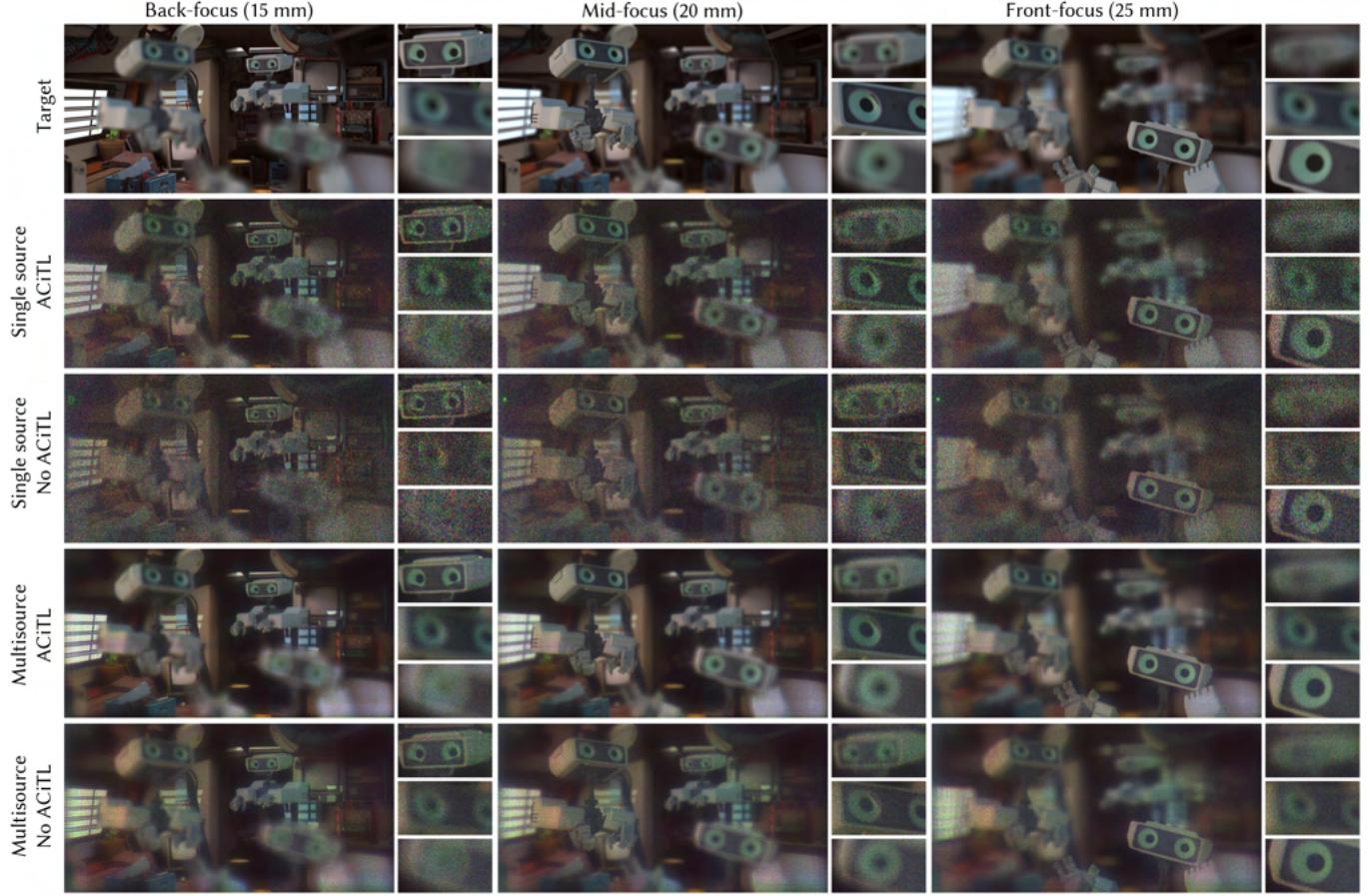


Fig. S11. Experiment (Focal Stack | ACiTL vs Learned Model): This figure shows a focal stack scene with natural defocus blur for both single and multisource optimized with ACiTL (using the proxy gradients of the calibrated model) vs. with the calibrated model only without ACiTL. While image quality with the learned model is already good, the ACiTL shows some clear improvement, especially in terms of noise. However, the multisource experimental capture without ACiTL is less noisy than the single source capture with ACiTL, highlighting the improvement that multisource holography provides.

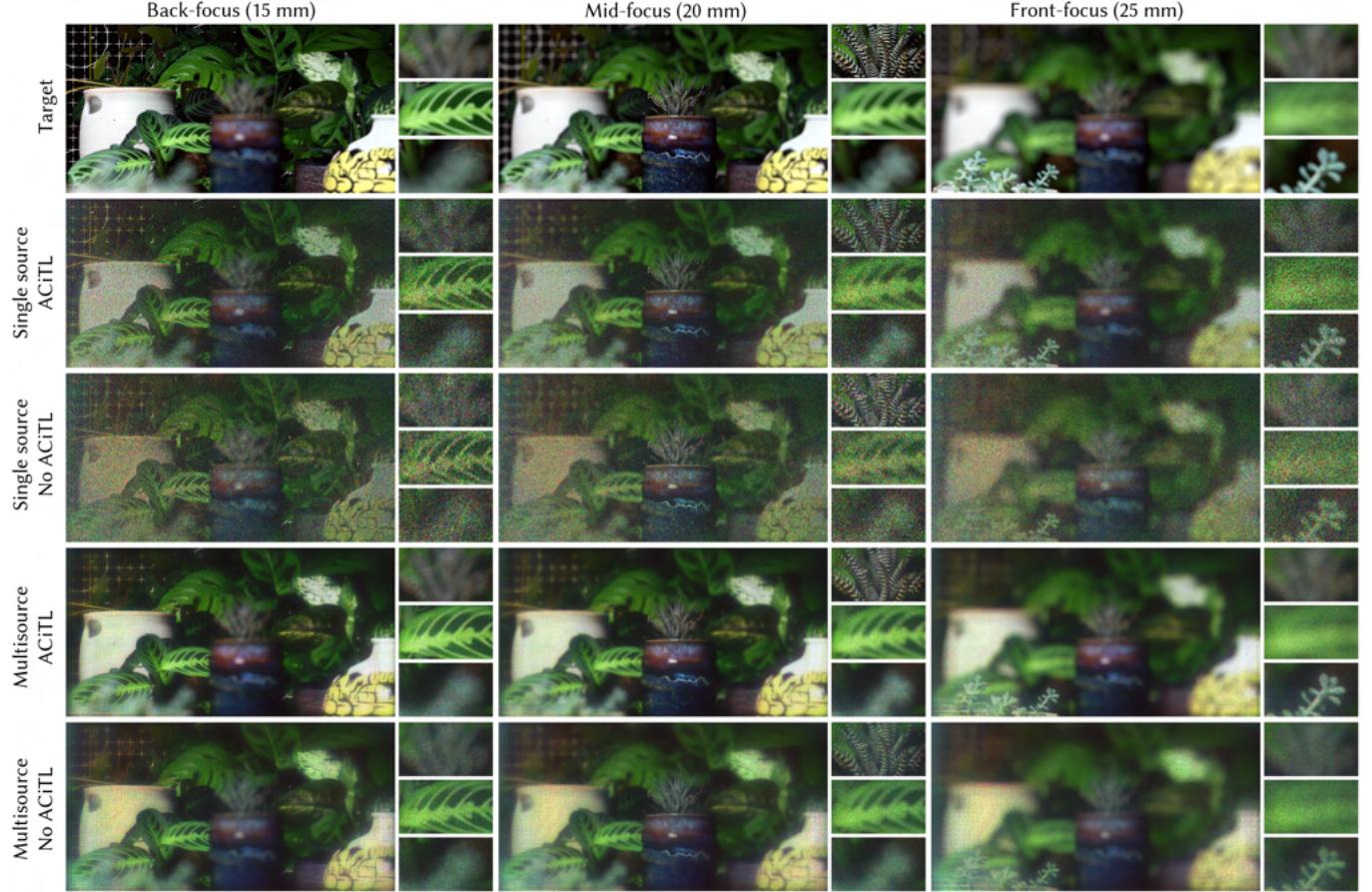


Fig. S12. Experiment (Focal Stack | ACiTL vs Learned Model): This figure shows a focal stack scene with natural defocus blur for both single and multisource optimized with ACiTL (using the proxy gradients of the calibrated model) vs. with the calibrated model only without ACiTL. While image quality with the learned model is already good, the ACiTL shows some clear improvement, especially in terms of noise. However, the multisource experimental capture without ACiTL is less noisy than the single source capture with ACiTL, highlighting the improvement that multisource holography provides.

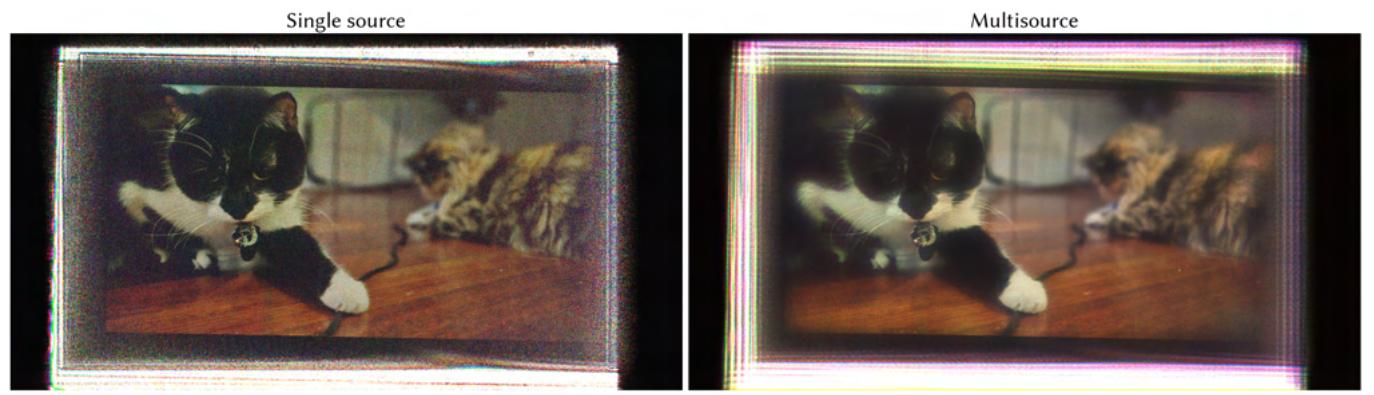


Fig. S13. Raw Camera Capture: We show examples for both single source and multisource of the raw camera images before any alignment or cropping is applied. The only processing is combining the three monochrome captures into a color image.

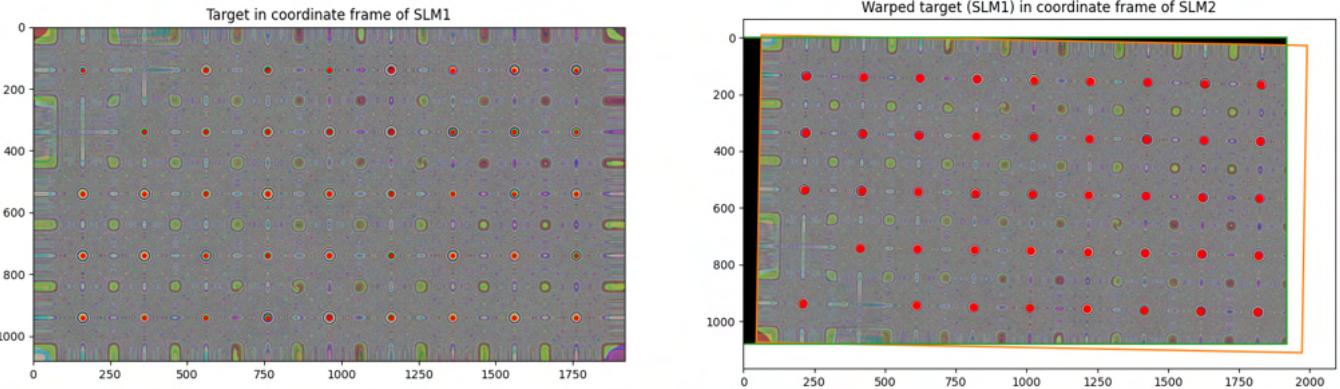


Fig. S14. Visualization of Warping Between SLMs: We visualize the transformation between the two SLMs, which is modeled using a thin plate spline (TPS) deformation model. Using a displayed grid of dots, we calculate a TPS mapping from SLM1 to the detector and from SLM2 to the detector. Then, we convert these mappings into a TPS from SLM1 to SLM2. Our dot pattern is asymmetric which helps ensure the orientation is correct.

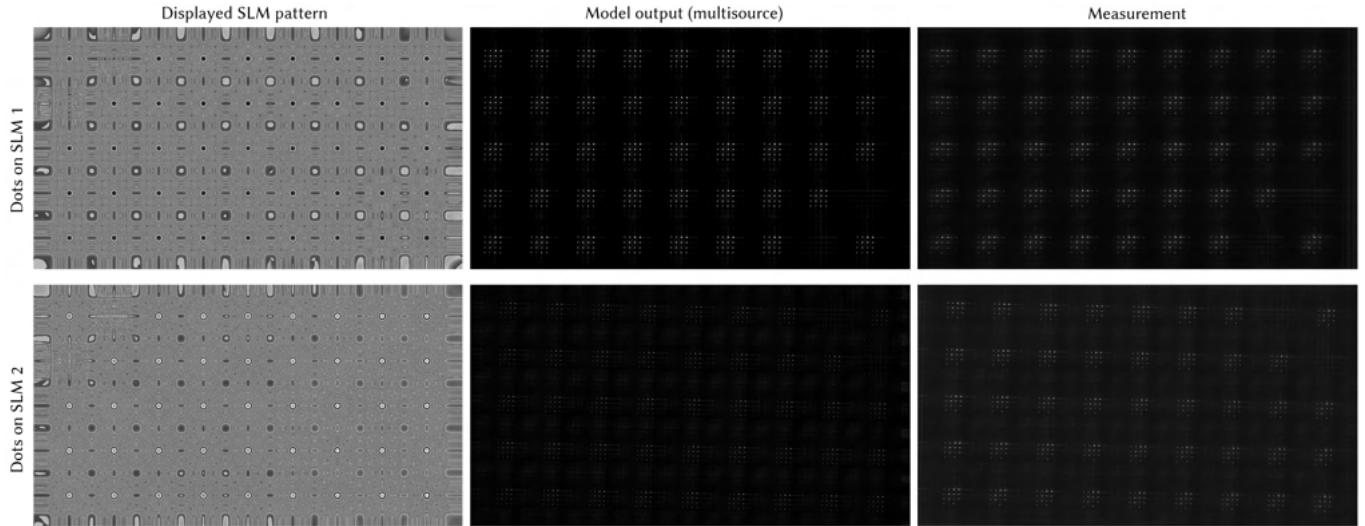


Fig. S15. Experimental Image of Dots with Multisource Illumination: If we optimize an SLM pattern to create a grid of dots with a single source, when we illuminate the same pattern with our multisource configuration, we see an image of the source at each dot location. Here, we see a 4×4 grid of sources at each location where there is a quadratic phase in the SLM pattern. Comparing the experimental measurement to our model output, one can see the the model is able to learn the locations of the sources and their relative intensities fairly accurately.

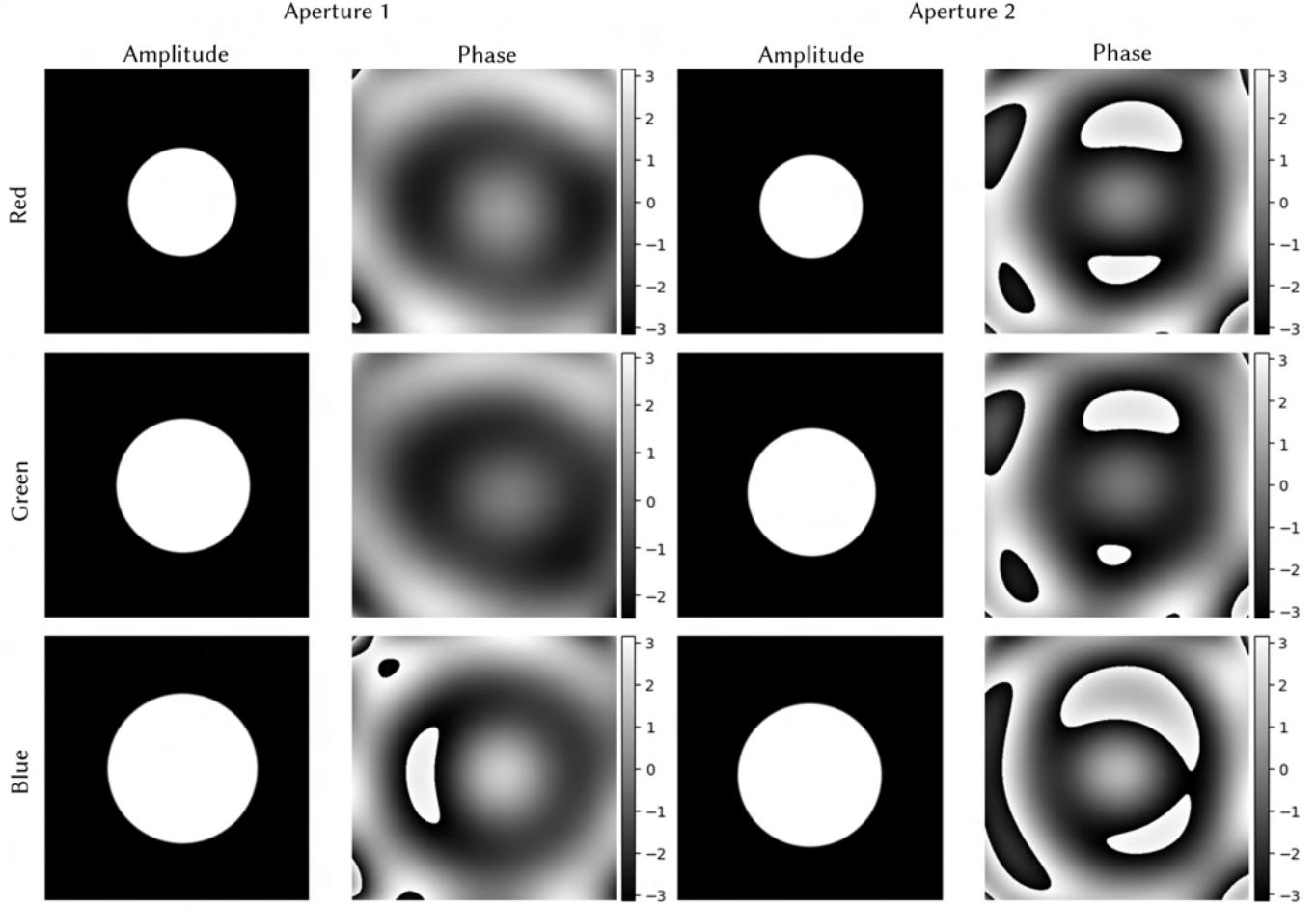


Fig. S16. Calibrated Global Apertures: The global apertures learned by our model, visualized here for each color, are applied in a non-spatially varying manner and are parameterized by a circle in amplitude and Zernike coefficient in phase. We learn a separate aperture for each propagation: “Aperture 1” is applied during propagation between the two SLMs and “Aperture 2” is applied during propagation from the second SLM to the sensor. Note that the same physical aperture appears to be different sizes for different wavelengths since we visualize the apertures in relative frequency coordinates which scale with wavelength.

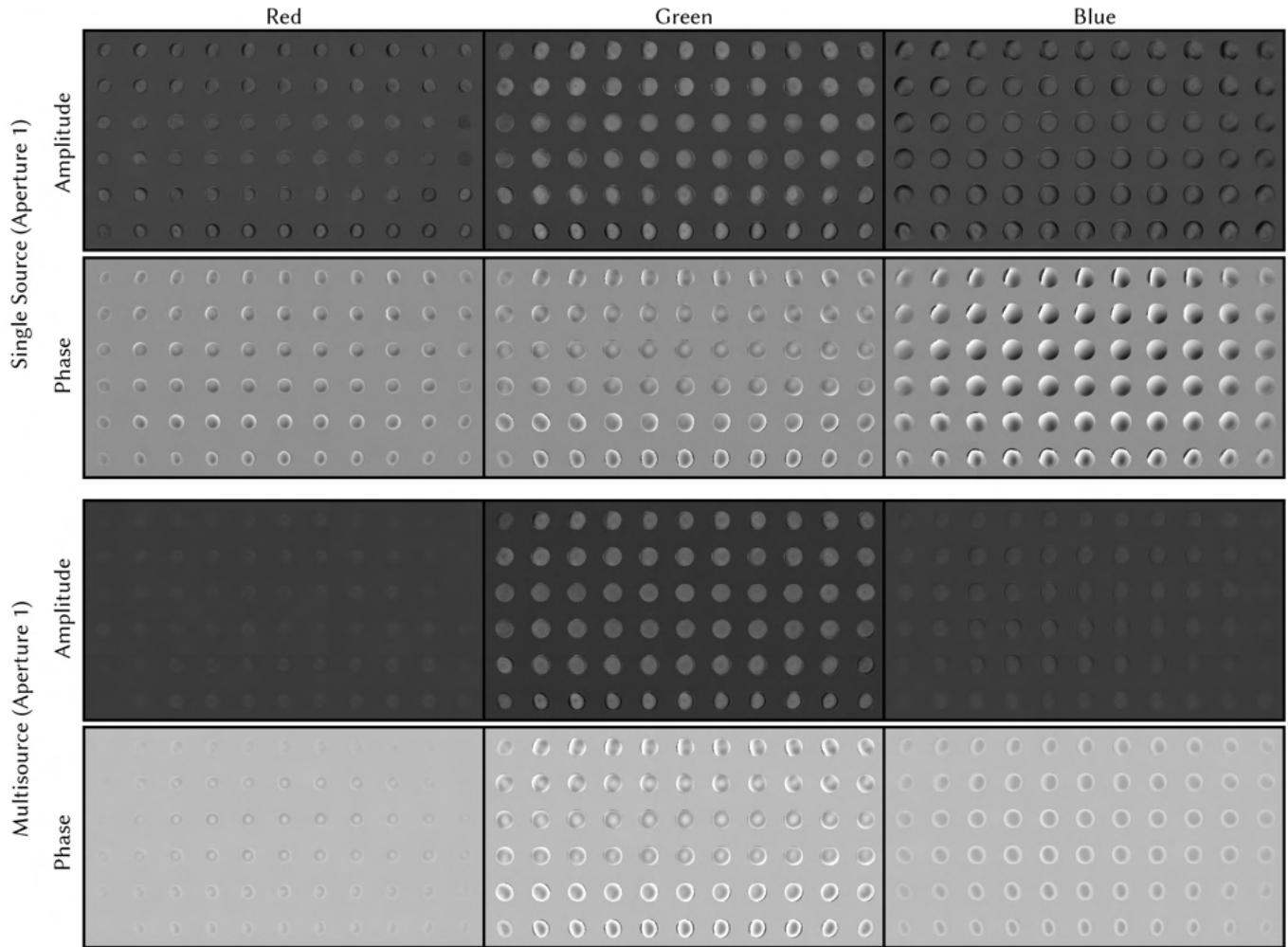


Fig. S17. Spatially Varying Aberrations for Aperture 1: We visualize the spatially varying aberration map for “Aperture 1” (between SLM1 and SLM2), evaluated at 6×11 locations in the field of view of the SLM. The spatially varying aberrations change smoothly over the field of view, which matches our physical intuition of how aberrations vary with field position.

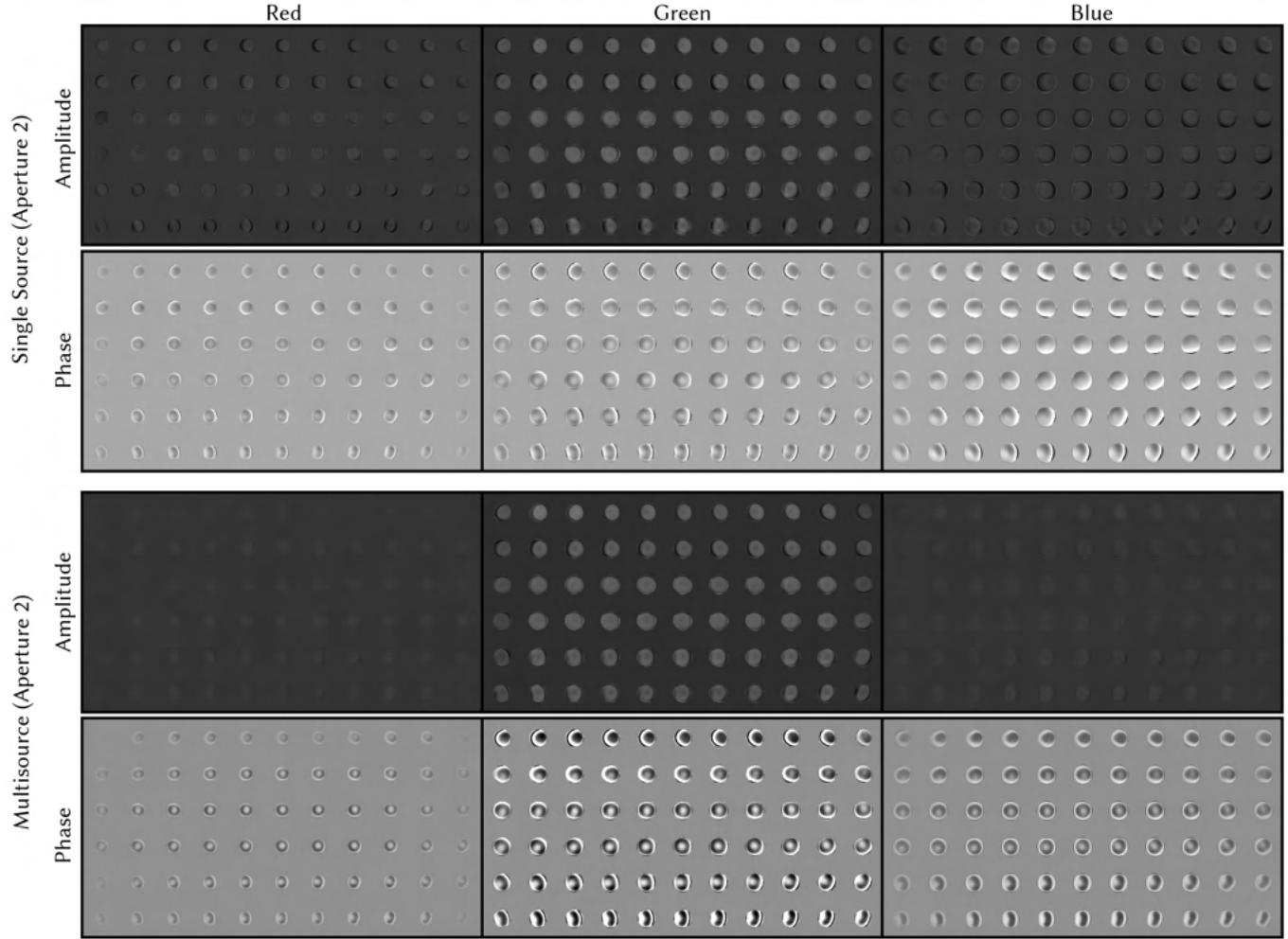


Fig. S18. Spatially Varying Aberrations for Aperture 2: We visualize the spatially varying aberration map for “Aperture 2” (between SLM2 and the sensor), evaluated at 6×11 locations in the field of view of the SLM. The spatially varying aberrations change smoothly over the field of view, which matches our physical intuition of how aberrations vary with field position.

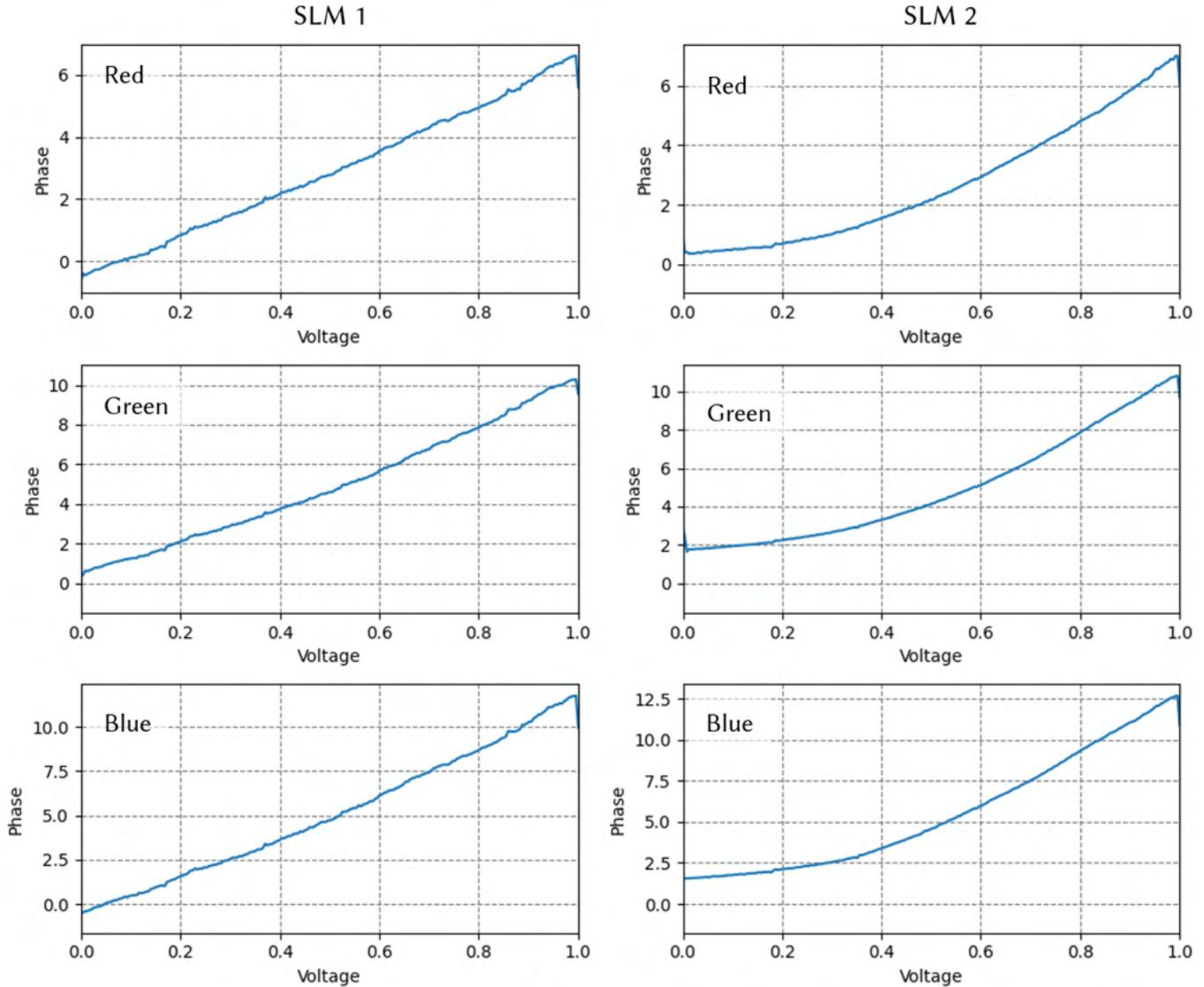


Fig. S19. Calibrated Lookup Tables: To calibrate how the SLM transforms a digital signal to a phase offset, we learn a lookup table (LUT) for each SLM. The LUT contains one phase value for each possible digital input (0-255), and is made differentiable through 1D interpolation. We calibrate a different LUT for each color. Note that SLM1 has an approximately linear LUT, but SLM2 shows non-linearities which must be accurately captured by the model to form high quality holograms.

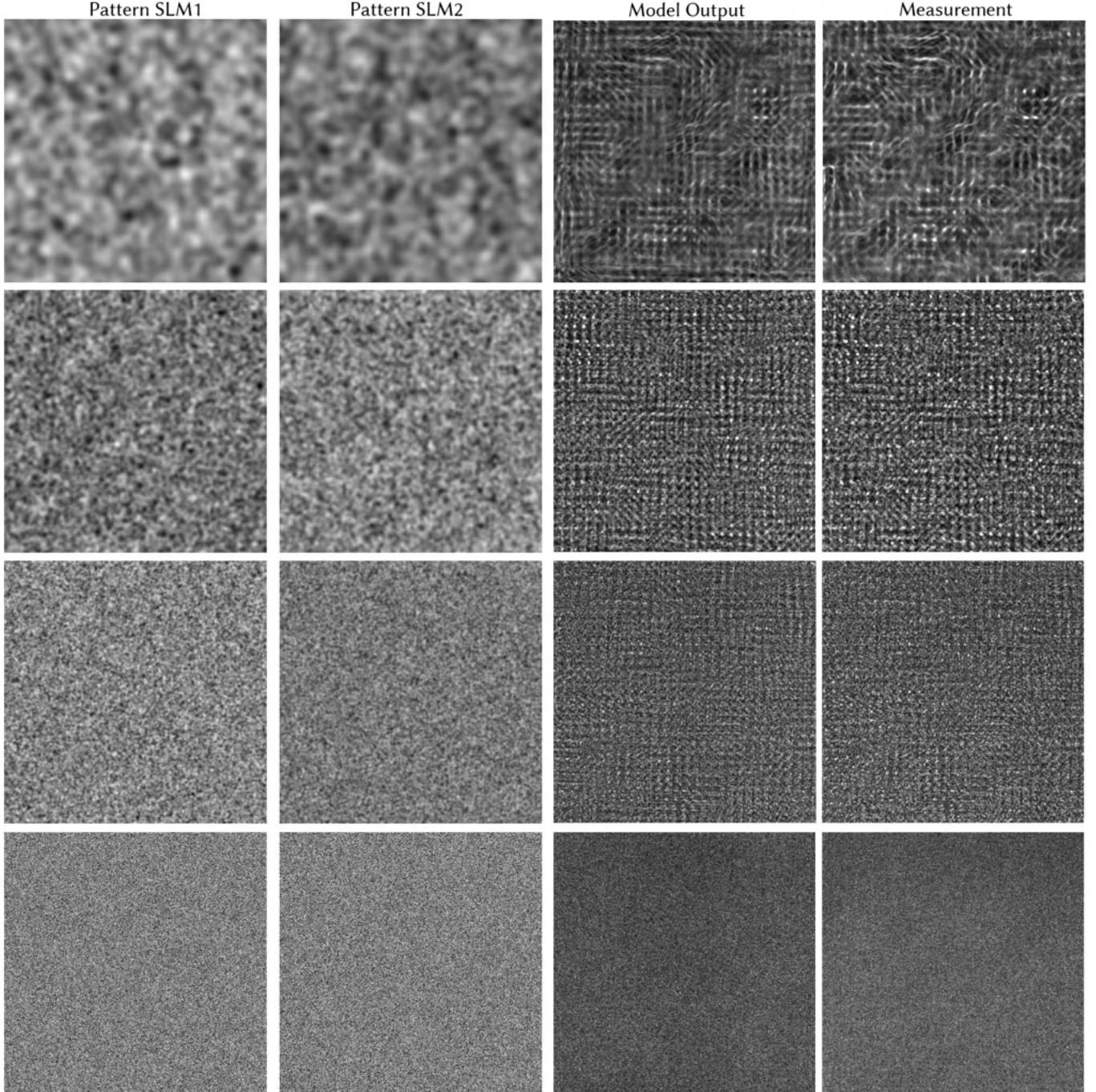


Fig. S20. Example Calibration Images: We show examples of the SLM patterns and captured images used for training our physically-based model. The SLM patterns are blurred with a Gaussian kernel of variable width to create a range of spatial frequencies in the calibration dataset. We use 7 different blur levels in total, but only visualize 4 here, and we visualize crops of 600×600 pixels in the center of the SLM. Sample experimental measurements used for training are shown in the right-most column, and the output of our calibrated model is shown adjacent. The model output and measurement are visually similar, although there are differences at the edges due to differences in boundary conditions. To prevent these from impacting model calibration, we further crop these images to a region without boundary artifacts before computing the loss.