

Stochastic Light Field Holography

Florian Schiffers, Praneeth Chakravarthula, Nathan Matsuda, Grace Kuo,
Ethan Tseng, Douglas Lanman, Felix Heide, Oliver Cossairt

Abstract—The Visual Turing Test is the ultimate goal to evaluate the realism of holographic displays. Previous studies have focused on addressing challenges such as limited étendue and image quality over a large focal volume, but they have not investigated the effect of pupil sampling on the viewing experience in full 3D holograms. In this work, we tackle this problem with a novel hologram generation algorithm motivated by matching the projection operators of incoherent (*Light Field*) and coherent (*Wigner Function*) light transport. To this end, we supervise hologram computation using synthesized photographs, which are rendered on-the-fly using Light Field refocusing from stochastically sampled pupil states during optimization. The proposed method produces holograms with correct parallax and focus cues, which are important for passing the Visual Turing Test. We validate that our approach compares favorably to state-of-the-art CGH algorithms that use Light Field and Focal Stack supervision. Our experiments demonstrate that our algorithm improves the viewing experience when evaluated under a large variety of different pupil states.

Index Terms—Computational Display, Holography, Light Field, Wigner Distributions, Near-Eye Display, VR/AR

1 INTRODUCTION

IN 1972, the Cartier jewelry store on 5th Avenue in New York City displayed an analog hologram of a hand-holding jewelry on their window storefront. This hologram was so realistic that an elderly woman passing by attempted to attack the virtual arm floating in mid-air [1]. Large-format analog holograms are known for their incredible realism and ability to give the impression of a complete 3D picture frozen in time. For static objects, full-color holography has been shown to provide realism on par with the visual inspection of the actual object. As such, holography is often seen as the most likely approach to pass the Visual Turing Test [2]. However, the dynamic display of high-quality digital holographic 3D imagery has proven challenging as a result of the limited system étendue. Étendue is determined by the Space-Bandwidth-Product (SBP) of the modulator, which is equal to the product of the maximum diffraction angle supported by the modulator and its modulation area. Most existing holographic systems rely on phase-only spatial light modulators (SLM), where this angle is limited by the pixel size – today’s fabrication techniques achieve $8\mu\text{m}$ pixel pitch for phase SLMs with no more than 4° of diffraction while limiting the eyebox to the size of the device. As a result, the available SBP mandates a *trade-off between FOV and eyebox size that impacts the number of achievable angular views*. The SBP required for holographic near-eye displays with a large eye box (roughly 10mm) and large FoV (100 degree), is still so high that pupil steering using eye tracking might be necessary unless étendue-expansion approaches [3], [4], [5], [6], [7] make significant progress.

Problem Statement. Even with novel designs, achieving and utilizing a large enough SBP to fully realize the potential of holographic displays remains a challenge. Specifically,

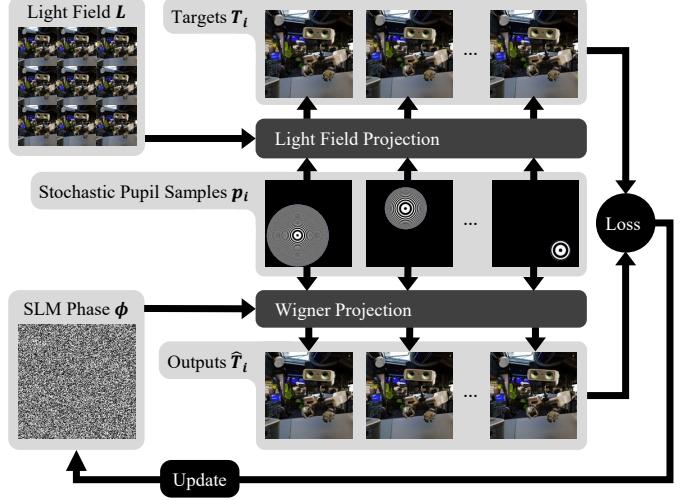


Fig. 1: *Stochastic Light Field Holography Algorithm*: We wish to infer an SLM phase image ϕ that produces a hologram with viewable intensity matching the projections rendered from a target Light Field L . To do this, a series of pupil samples with randomly generated position, diameter, and focus, p_i are passed to two equivalent image formation models: a Light Field projection operator, which produces individual target images T_i , and a Wigner projection operator, which produces individual output estimates \hat{T}_i , for each of the pupil samples. These target and output estimates are compared using a photo-consistency loss function, which is in turn used to update the SLM phase image with a gradient descent step.

• F. Schiffers is with the Computer Science Department at Northwestern University
 • F. Schiffers, N. Matsuda, G. Kuo, D. Lanman and O. Cossairt are with Reality Labs Research, Meta
 • P. Chakravarthula, E. Tseng and F. Heide are with Princeton University

a key question is how to compute holograms such that they provide an optimal image for *any* possible viewpoint. This problem is amplified for large SBP systems with large eyebox size as the pupil can move significantly within. Pupil movement, including changes in location, size, and defo-

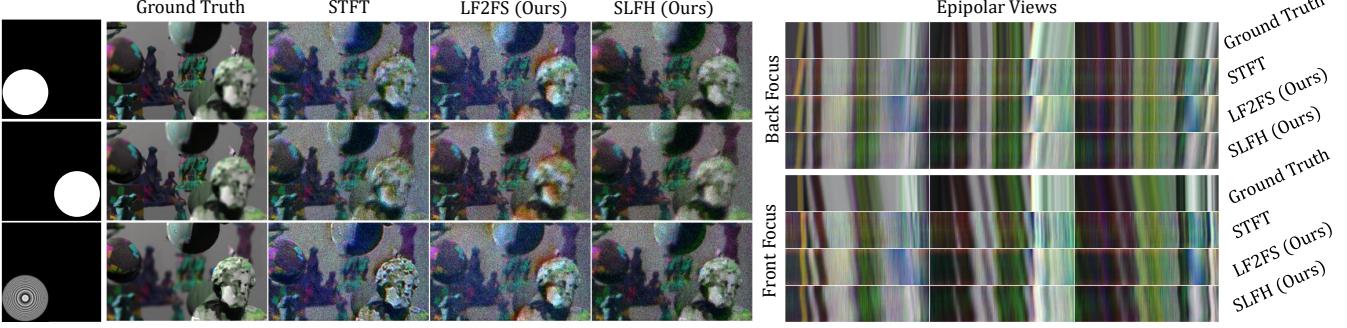


Fig. 2: *Stochastic Light Field Holography on experimental hardware.* Recent advances in holographic displays achieve high image fidelity using smooth-phase holograms. This comes at the cost of a highly concentrated eye box limiting image formation to only the central view. For a more robust visual experience, holographic displays should provide good image quality for arbitrary pupil-states (*diameter, location, and accommodation*) within the entire available eye box. With *Stochastic Light Field Holography* (SLFH), we propose a novel framework that ensures photo-consistency over the entire eye-box volume. We implement a novel *Focal Stack* supervision algorithm (LF2FS) and show that SLFH is a generalization of state-of-the-art (SOTA) *Focal Stack* and *Short-Time-Fourier-Transform* (STFT) supervision of CGH optimization algorithms, representing only a limited subset of possible pupil states. Consequently, our proposed method provides the best average image fidelity over the entire eyebox compared to SOTA. On the left, we show experimental captures from three example pupil states (*Front/Back focus for accommodation, left/right for parallax*). On the right, we show epipolar slices generated by capturing a 1D trajectory consisting of 31 small pupils (1/8th of eyebox) from left to right for both front and back focus. Our SLFH method produces the least artifacts over the full eyebox, while Focal Stack supervision produces strong color-fringing at occlusions and STFT-optimized holograms often show ringing artifacts and over-sharpened in focus images.

cus, can significantly affect image quality on holographic displays [8]. Accurately accounting for pupil movement is, as such, a critical challenge in the development of photo-realistic holographic displays, as it can greatly affect the realism and immersion of the viewing experience (see Fig. 3). Only recently, researchers have investigated this issue for 2D holography [8] and showed only very preliminary 3D results using a two-plane representation with two objects. Hence, the computation of pupil-aware 3D holograms is an open problem that we address in this paper.

Proposed Solution. In this work, we address the challenge of computing high-quality holograms that accurately reproduce vision cues during pupil movement. To do this, we propose a new Stochastic Light Field Holography (SLFH) algorithm that generates holograms using a loss function motivated by matching the photographic projection operators of incoherent and coherent light. Our approach ensures optimal performance for a random assortment of pupil states and, unlike previous work, produces 3D imagery while also avoiding overfitting to specific viewing conditions. While previous methods use a loss function based on a set of static, prerendered images, we supervise our loss function by synthesizing novel views using Light Field refocusing [9] from arbitrary pupil states. By sampling pupil states stochastically, we ensure, unlike existing work, that the optimized hologram provides correct parallax, defocus, and depth-of-field cues for all possible pupil states. In order to have a fair comparison method with multi-plane approaches [10], [11], we introduce a novel Focal Stack supervision algorithm (LF2FS) which ensures a parallax-consistent defocus allowing for a direct comparison to our main contribution, which is SLFH.

Contributions. Our software codebase for stochastic pupil sampling will be made publicly available after publication of this

manuscript. Our paper makes the following contributions:

- We are the *first* to *study* the image quality of 3D-CGH algorithms for all possible pupil states within the available eye-box volume.
- We are the *first* to *introduce* on-the-fly Light Field refocusing [12] into the holographic generation process. We *propose* a new CGH-algorithm that reproduces parallax and accommodation for arbitrary pupil states by coupling the projection operators for ray-space (*Light Field*) and wave-optics (*Wigner Distribution*). We use this to implement both a novel Focal Stack supervision algorithm (LF2FS) and our proposed SLFH method.
- We *validate* the method in simulation and demonstrate that, while previous techniques overfit to specific pupil states used during training, our method produces the most favorable display for arbitrary viewing conditions (see Fig. 4).
- We *assess* the method with an experimental prototype and demonstrate higher quality parallax information with fewer artifacts than state-of-the-art Light Field supervision CGH optimization algorithms (see Figs. 2, 6, and Supplemental Material).

2 RELATED WORK

Holography for displays relies on diffraction and interference of light to generate images. Based on the diffracted field, a hologram can be classified as a far-field Fourier hologram or a near-field Fresnel hologram. Using phase-only SLMs requires computing phase-only holograms that are capable of producing the diffraction field that can closely mimic the target image. However, the underlying phase retrieval problem is generally ill-posed and non-convex. Though introduced for Fourier phase retrieval, early methods such as error reduction using iterative optimization [13],

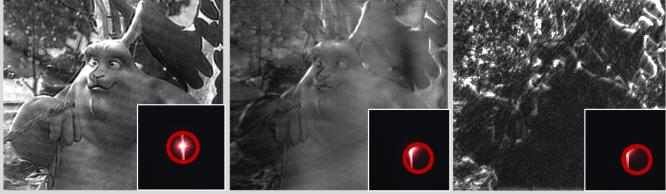


Fig. 3: *The eyebox problem and Smooth-Phase CGH algorithms:* State-of-the-art holograms can achieve high image quality, but their spectral energy is highly concentrated leading to a small, effective eye-box. Here, we show image degradation of smooth-phase holograms on our experimental setup once the eye-pupil (red inset) does not sample the DC-term anymore.

[14] and hybrid input-output (HIO) methods [15], [16] are applicable for both Fourier and Fresnel holograms. Researchers have also explored phase-retrieval methods using first-order nonlinear optimization [17], alternative direction methods for phase retrieval [18], [19], nonconvex optimization [20], and methods overcoming the nonconvex nature of the phase retrieval problem by lifting, i.e., relaxation, to a semidefinite [21] or linear program [22], [23]. Several works [24], [25] have explored optimization approaches using first-order gradient descent methods to solve for holograms with flexible loss functions.

Focal Stack and RGBD Algorithms. Instead of computing the wave propagation for millions of points, a 3D object can be represented as a stack of intensity layers [10], [26], [27], [28]. Wave propagation methods such as the inverse Fresnel transform or angular spectrum propagation are typically used for propagating the waves from several layers of the 3D scene towards the SLM plane, where they are interfered with to produce a complex hologram [3], [20], [24], [29]. Although this approach can be implemented efficiently, it cannot support continuous focus cues and accurate occlusion due to discrete plane sampling. Recent work [30] suggests that holograms based on smooth-phase profiles [31] cannot support accommodation; the ultimate promise of 3d holography. An additional discussion on smooth vs random phase holograms is found in the supplementary.

A further approximation to the layer-based methods is to determine the focal depth of the user, i.e., distance of the object to which the user fixates via an eyetracker and adjusting the focal plane of the 2D holographic projection to match the user focal distance [32]. While emulating a 3D scene by adaptively shifting a 2D holographic projection in space is computationally efficient, operating in a varifocal mode under-utilizes the capabilities of a holographic display. Moreover, achieving natural focus cues and physically accurate occlusion effects still remains a challenge. To compare against SOTA Focal Stack methods, we introduce a novel *Focal Stack supervision algorithm (LF2FS)* that generates Focal Stacks on the fly from Light Fields according to physically accurate pupil states.

Light Field Algorithms. To support occlusion and depth-dependent effects, a Light Field can be encoded into a hologram partitioned spatially into elementary hologram patches, called “hogels” [34], similar to elementary images in a Light Field. These hogels produce local ray distributions that reconstruct multiple (Light Field) views [35], [36], [37].

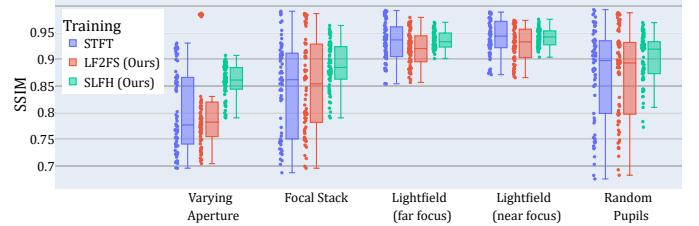


Fig. 4: *Statistical Evaluation:* Statistical performance evaluation of SSIM perceptual similarity metric over the Deep Spaces [33] light field dataset. Holograms are trained using STFT, Focal Stack supervision (LF2FS), and Stochastic Light Field Holography (SLFH). Statistics are plotted for various aperture types: Varying Aperture (focus and position are fixed), Focal Stack (aperture diameter and position are fixed), Light Field (aperture diameter and focus are fixed), and random pupils. Each dot represents the average SSIM of all evaluated pupil-positions for a specific aperture type for one specific light field within the Deep Spaces [33] dataset. Hence, each dot corresponds to the average performance for a specific scene. Our method always produces the best worse case performance, and produces a better mean and variance for random pupils, which means that our algorithm performs better for a wider variety of near-eye viewing conditions.

Such holograms which encode a Light Field are called “holographic stereograms”. Conventional stereograms, where hogels are out of phase with each other, suffer from a lack of focus cues and limited depth of field [35]. To keep the hogels of a holographic stereogram in phase throughout the hologram, researchers have introduced an additional phase factor to calculate what is called a phase-added stereogram (PAS) [36]. However, akin to a microlens array-based Light Field display [38], stereograms suffer from the fundamental spatio-angular resolution trade-off: A larger hogel size leads to a decreased spatial resolution. This fundamental limitation does not allow for high spatial resolution holographic stereogram projections. Recent methods have attempted to overcome this trade-off [39], [40], [41] via Short-Time Fourier Transform (STFT) inversion. However, these methods do not match the image quality achieved for 2D holograms [24], [42] and suffer from artifacts around object discontinuities due to suboptimal inversion of the STFT. More importantly, these algorithms do not incorporate pupil models for viewing the hologram, and therefore overfit to viewing conditions where the eyebox and pupil coincide. In Sec. 3 will further discuss the Wigner function and how it relates to holographic displays. We want to stress that we are not the first to consider the Wigner-function for holographic displays [41], [43], [44]. However, these methods use the Wigner-function to establish a closed-form solution similar to [39], which is different from ours, which uses an iterative approach to recover the displayed phase-modulation.

Learning-Based Holography. Neural networks and deep learning approaches have recently been proposed as tools for optical design and holographic phase retrieval. Holographic microscopy has been tackled by solving phase retrieval problems using neural networks [28], [45]. In a similar fashion, neural networks have been investigated for learning holographic wave propagation from a large train-

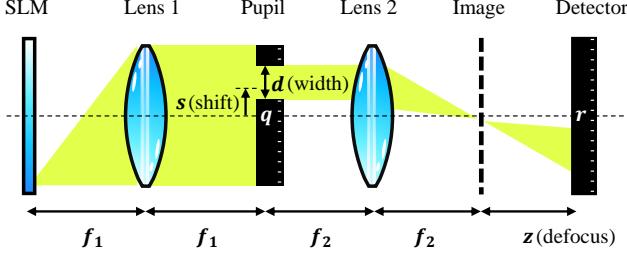


Fig. 5: Simplified diagram of the proposed setup: An SLM (illumination path not shown) diffracts a collimated beam according to an optimized pattern. The modulated beam passes through a 4-f system, with a pupil having programmable shift s and diameter d . The relay system produces an image of the SLM offset from the detector by a defocus distance z .

ing dataset. For example, Horisaki et al. [46] trained a U-net on a pair of SLM phase and intensity patterns, and predicted SLM phase patterns during inference. Recently, Eybposh et al. [28] proposed an unsupervised training strategy and predicted the SLM phase patterns in real time that produced 2D and 3D holographic projections. Peng et al. [24] and Chakravarthula et al. [42] have recently demonstrated camera-in-the-loop (CITL) calibration of hardware using neural networks and high-fidelity holographic images on prototype displays. Shi et al. [47] have demonstrated high-resolution real-time holography with a light weight neural phase-retrieval network that may be suitable for inference on mobile hardware in the future.

Pupil and Parallax-aware Holography. Very recently, researchers have investigated the effects of pupil sampling on holographic setups. Chakravarthula et al. [8] work is the most similar work related to the investigation in this paper. Chakravarthula et al. propose a new cost function which enforces a uniform energy distribution over the eye-box by stochastically sampling different pupil positions. However, their approach is limited to 2D-imagery and very preliminary multi-plane images, and, as a result, they did not account for parallax and focus-cues effects. Methods that operate on RGB-D scenes [10], [24], [47] are capable of finding correct accommodation cues by using multi-plane or Focal Stack representations, however, are not able to account for correct parallax cues when the pupil is shifting. Motivated by this, recent work has proposed holographic generation algorithms using Light Field supervision [39], [48], [49], making use of the Short Time Fourier Transform (STFT) and allowing a bidirectional transform between holograms and Light Fields as proposed by [50]. All of these existing methods have in common that they do not allow for pupil sampling and accurate parallax cues at the same time.

3 PHOTOGRAPHIC PROJECTION OPERATORS

In this section, we establish a connection between the coherent (*wave-optics*) and incoherent (*ray-space*) projection operators. We first introduce Light Fields, their coherent equivalent using the Wigner-formalism, and their resemblance to STFT approaches used in current literature.

3.1 Eye Pupil Model

The primary contribution of our work is to introduce a simplified eye/pupil model into the hologram forward model for optimization of photorealistic 3D imagery over a variety of viewing conditions. We approximate a near-eye holographic display system as a 4f relay where an eyepiece images an SLM illuminated by monochromatic light through a pupil, followed by an eye lens which images onto a detector (see Fig. 5). We represent 2D coordinates in the detector plane as $\mathbf{r} = (x, y)$, and pupil plane as $\mathbf{q} = (u, v)$. The eye is assumed to be focused at infinity in the rest state so that converting from pupil to angular coordinates is given by the relation $\mathbf{q} = \tan(\theta)/f$, where f is the focal length of the eye model or objective lens (f_2 in Fig. 5). For simplicity, we parameterize the pupil state p with a set of 4 parameters, that is

$$p = \{\mathbf{s}, z, d\}, \quad (1)$$

corresponding to focus distance (z), 2D-pupil shift $\mathbf{s} = (x, y)$, and aperture diameter (d). We define the pupil aperture function as a circular function that focuses at infinity in the rest state

$$A(\mathbf{q}; p) = \begin{cases} 1, & \text{if } |\mathbf{q} - \mathbf{s}| < d \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

3.2 Light Fields

Light Fields represent the flow of radiance in a scene in every direction and every point in space. This extra information can be used to generate synthetic 2D images from different viewpoints and simulate optical effects like defocus and lens aberrations. The photography operator is used in digital refocusing [12] to render images at different depths from the Light Field under a virtual aperture function. Specifically, the operator \mathbf{P}_{lf} describes the transformation of a 4D Light Field $L(\mathbf{r}, \mathbf{q})$ into a synthetic photograph focused through an aperture function A with a depth defined by α . We parameterize the Light Field such that \mathbf{r} and \mathbf{q} represent the spatial coordinates of the Light Field on detector and pupil planes, respectively. The photographic projection operator can then be expressed as the following projection from 4D to 2D

$$\mathbf{P}_{lf}[L; p](\mathbf{r}) = \int A(\mathbf{q}; p) \cdot L(\mathbf{q} - \frac{\mathbf{q} - \mathbf{r}}{\alpha}, \mathbf{q}) d\mathbf{q}, \quad (3)$$

Equation 3 can be thought of as shearing the 4D space, multiplication with the aperture function and a subsequent projection down to 2D. Note that the effects of changing focus distance z are incorporated into the defocus parameter $\alpha = f/(z + f)$.

3.3 Wigner Function

The projection operator for Light Fields in Eq. 3 works only for incoherent light as it cannot account for interference effects. To overcome these limitations, [51], [52] used the Wigner Distribution-Function (WDF) to establish the connection to Light Fields. While it is computationally intractable to use Wigner functions efficiently for large holograms, it provides a powerful tool for understanding optical systems and their limitations. To illustrate this, let us consider a monochromatic optical field $u(\mathbf{r})$. The correlation

between two points on the field \mathbf{r} and \mathbf{r}' is given by the mutual intensity function, which operates on a 2D field and then produces a 4D quantity as

$$J[u(\mathbf{r})](\mathbf{r}, \mathbf{r}') = \left\langle u\left(\mathbf{r} + \frac{\mathbf{r}'}{2}\right) u^*\left(\mathbf{r} - \frac{\mathbf{r}'}{2}\right) \right\rangle, \quad (4)$$

where $\langle \cdot \rangle$ denotes a time-average. The WDF is then defined as the Fourier transform of the mutual intensity along \mathbf{r}'

$$W[u(\mathbf{r})](\mathbf{r}, \mathbf{q}') = \int J(\mathbf{r}, \mathbf{r}') \exp(-2\pi i \mathbf{r}' \cdot \mathbf{q}') d\mathbf{r}'. \quad (5)$$

The spatial frequency coordinates \mathbf{q}' of the WDF can be converted to spatial pupil coordinates \mathbf{q} via the relation

$$W(\mathbf{r}, \mathbf{q}) = W(\mathbf{r}, \lambda f \mathbf{q}'). \quad (6)$$

Wigner Projection Operator. We next consider the WDF $W(\mathbf{r}, \mathbf{q})$ defined at the plane of a phase-only SLM with displayed phase pattern $\phi(\mathbf{r})$, producing a diffracted field $u(\mathbf{r}) = \exp(j\phi(\mathbf{r}))$. Defining the WDF of the aperture function $A(\mathbf{q}; p)$ to be $W_A(\mathbf{r}, \mathbf{q})$, the WDF that is imaged onto the retina of the eye (assuming unit magnification for simplicity) is

$$W_o[u(\mathbf{r})](\mathbf{r}, \mathbf{q}) = \int W_A(\mathbf{r}, \mathbf{q}_i) W[u(\mathbf{r})](\mathbf{r}, \mathbf{q} - \mathbf{q}_i) d\mathbf{q}_i \quad (7)$$

The effect of a focus shift z in Wigner-space can be represented as a shear [53] given by

$$W_z(\mathbf{r}, \mathbf{q}) = W_o(\mathbf{r} - \lambda z \mathbf{q}', \mathbf{q}) = W_o\left(\mathbf{r} - \frac{z}{f} \mathbf{q}, \mathbf{q}\right), \quad (8)$$

which results in the Wigner projection operator

$$\begin{aligned} \mathbf{P}_{WF}[u(\mathbf{r}); p](\mathbf{r}) &= \\ &\iint W_A(\mathbf{r}, \mathbf{q} - \mathbf{q}_i) \cdot W[u(\mathbf{r})]\left(\mathbf{q} - \frac{\mathbf{q} - \mathbf{r}}{\alpha}, \mathbf{q}\right) d\mathbf{q} d\mathbf{q}_i. \end{aligned} \quad (9)$$

Relationship between Wigner and Light Field Projection. Comparing Eq. 9 with Eq. 3, we observe that the projection geometry is identical while the WDF projection allows us to incorporate interference effects from the eye pupil that naturally arise when viewing a holographic display illuminated with coherent light. Closer inspection reveals that the Wigner projection operator is a generalization of Light Field projection. When the retina detector pixel size Δ is much larger than the diffraction airy disc size $\Delta \gg \lambda \cdot f/d$, then the effect of diffraction from the pupil aperture on the perceived image is negligible. In this case, the WDF for the aperture is approximately independent of spatial coordinates \mathbf{r} so that $W_A(\mathbf{r}, \mathbf{q}) \approx A(\mathbf{q}; p) \cdot W(\mathbf{r}, \mathbf{q})$. Hence, the projection operators for Eq. 9 and Eq. 3 are identical within a scale factor.

Wigner Projection with Wave Optics. The similarity between photographic projection operators for Light Fields and WDFs provides an intuition for how to optimize holograms that produce photorealistic imagery over a variety of viewing conditions. However, while the WDF is an elegant way to describe coherent light transport, it is impractical to compute on large fields [54]. Instead, we implement the projection operator of Eq. 9 using wavefront propagators via stochastic pupil-sampling as described in Sec 4.2.

3.4 Bidirectional Hologram Light Field Transform

As discussed in Sec. 2, a successful branch of LF-CGH algorithms employs STFT-inversion to move between LFs and hologram-domains [50]. The Wigner projections (Sec. 3.3) have striking similarities with the STFT, as we will show in the following. The STFT of a signal $u(\mathbf{r})$ is defined as the convolution with a window function $w(\mathbf{r})$ such that

$$\begin{aligned} \text{STFT}[u(\mathbf{r})](\mathbf{r}, \mathbf{q}) &= \int u(\mathbf{r}') w(\mathbf{r}' - \mathbf{r}) e^{-j2\pi \mathbf{q} \cdot \mathbf{r}'} d\mathbf{r}' \\ &= \mathcal{F}^{-1}\{U(\mathbf{q}') \cdot W(\mathbf{q}' - \mathbf{q})\}, \end{aligned} \quad (10)$$

where $U(\mathbf{q}')$ and $W(\mathbf{q}')$ are Fourier transforms of the field $u(\mathbf{r})$ and window function $w(\mathbf{r})$, respectively. The relationship between the STFT of the field $u(\mathbf{r})$ and the target lightfield is then

$$L(\mathbf{r}, \mathbf{q}) = |\text{STFT}[u(\mathbf{r})](\mathbf{r}, \mathbf{q})|^2. \quad (11)$$

Together, Eqns. 10 and 11 express the relationship between STFT-based CGH algorithms and the Wigner projection operator implemented using wave optics as detailed in Sec. 4.2. Implicitly, the STFT computes an image of the hologram as viewed with a pupil function $W(\mathbf{q}')$, which is the Fourier Transform of the STFT window function $w(\mathbf{r})$. Typically a Hanning window or similar function is used to suppress ringing, and a grid of pupil positions is computed by applying a 2D FFT to the windowed field $u(\mathbf{r}') w(\mathbf{r}' - \mathbf{r})$.

The method we propose in this work can be thought of as a generalization of STFT-based methods, with a few important differences. First, while the STFT method uses fixed pupils for hologram generation that provide photo-consistency with lightfield views, our method ensures that arbitrary pupil functions applied to the holographic display are consistent with Light Field projection using the same pupil. Second, our method relies on the relationship between spatial frequency coordinates and angular frequency coordinates of the field, which is necessary to produce accurate color reconstructions with multiple wavelengths. Lastly, the proposed method decouples the number of pupil positions from the forward model so that less memory can be used to compute larger holograms.

3.5 Eyebox, Pupils and Speckle

This paper focuses on the problem of developing optimization algorithms for holograms that produce photorealistic 3D imagery over a variety of random pupil states. A necessary requirement for these algorithms is that they produce an eyebox with a relatively smooth distribution of intensity so that there is not a drastic change in perceived brightness as the pupil state changes during hologram viewing. In order for a hologram to produce a uniform eyebox, it must also produce a diffracted field with a relatively even distribution of angular frequencies, which means that the CGH will typically have highly randomized phase. This has implications on the type of CGH algorithm used, as well as the quality of imagery displayed since random phase holograms naturally introduce speckle.

Many recent studies on holography do not address the issue of eyebox uniformity, and it is often assumed that the eye box is solely determined by the system etendue, which defines the maximum diffraction angle given by the SLM-pixel pitch. However, for most near-field hologram algo-

rithms that have recently reported great image quality [24], [47], the light distribution in the eye-box is heavily concentrated around the DC term. This concentration of spectral energy leads to a significant reduction in the "effective eye-box size". This problem arises naturally in many existing holographic generation algorithms. Double Phase (DPAC) encoding provides a direct method for encoding a complex field onto a phase only SLM, but only works well with smooth phase holograms that contain a Fourier spectrum highly concentrated around the optical axis (DC angular frequency). Likewise, similar characteristics are observed when using holographic generation based on gradient-descent style algorithms that are initialized with smooth phase (see Fig. 3).

Speckles are interference effects that occur when a diffracted field is highly random compared to its effective aperture. The speckle size on the retina of a pupil sampled hologram is inversely proportional to the pupil diameter d [55]. In this work, we focus on CGHs which are viewed with pupil diameters much smaller than the eyebox ($d \in [w_{\text{eyebox}}/12, w_{\text{eyebox}}]$), which can introduce significant speckle noise in simulated and captured imagery relative to sampling the full eyebox. To improve image quality and reduce the effect of speckle we introduce incoherent averaging in the form of temporal averaging of 8 frames and spatial averaging of 2×2 blocks of pixels. For fair comparisons, we apply the same incoherent averaging to all algorithms used in this paper.

4 STOCHASTIC LIGHT FIELD HOLOGRAPHY

Building on the projection operators from the previous section, we now introduce the proposed method. We first introduce the forward model that describes the mapping from a displayed phase-modulation ϕ on the SLM to the complex wavefront u , assuming that the image is captured under an arbitrary pupil state. Then, we will introduce the proposed novel CGH-algorithm using Light Field supervision (see Fig. 1).

4.1 Enforcing Photo-Consistency

The proposed method enforces photo-consistency between image formation of a known Light Field $L(\mathbf{r}, \mathbf{q})$ and a unknown Wigner distribution $W[u(\mathbf{r})](\mathbf{r}, \mathbf{q})$, where $u(\mathbf{r}) = \exp(j\phi(\mathbf{r}))$ is the complex field formed by the holographic display with phase-only slm pattern $\phi(\mathbf{r})$. We implement projection operators for both Light Field $P_{lf}[\cdot]$ (see Eq. 3) and the Wigner Distribution $P_{wd}[\cdot]$ (see Eq. 5) such that they match the geometry of our holographic display. The projections of the known Light Field produce a set of target photographs given the viewing parameters

$$T_i(\mathbf{r}) = P_{lf}[L, p_i]. \quad (12)$$

We then generate the corresponding Wigner projections corresponding to observations of the diffracted optical field with the same viewing parameters

$$\hat{T}_i(\mathbf{r}, \phi(\mathbf{r})) = P_{wd}[\exp(j\phi(\mathbf{r})), p_i]. \quad (13)$$

The optimization objective is then to find the SLM pattern $\phi(\mathbf{r})$ that solves

$$\phi(x)^* = \operatorname{argmin}_{\phi} \sum_i \mathcal{L}(T_i(\mathbf{r}), \hat{T}_i(\mathbf{r}, \phi(\mathbf{r}))) \quad (14)$$

$$\text{given } p_i = \begin{cases} x_i \sim \mathcal{U}[-r_{\max}, r_{\max}] & z_i \sim \mathcal{U}[z_{\min}, z_{\max}] \\ y_i \sim \mathcal{U}[-r_{\max}, r_{\max}] & d_i \sim \mathcal{U}[r_{\min}, r_{\max}] \end{cases} \quad (15)$$

where \mathcal{U} is the uniform distribution and \mathcal{L} is an arbitrary loss-function (L_2 in our case).

4.2 Implementation of Projection Operators

Wigner-Projection using Wave Optics. While the Wigner projection operator of Eq. 5 is useful to build an intuition for how to optimize a hologram that is radiometrically consistent with a target lightfield, *computing the full WDF is impractical* because the WDF is a 4D function which results in quartic memory growth. Instead, we use a Fourier optics forward model to stochastically sample pupil positions, resulting in the simplified 2D projection operator that does not require storing a 4D WDF, that is

$$P_{wd}[u(\mathbf{r}), p](\mathbf{r}) = \int U(\mathbf{q}) K(\mathbf{q}, p) e^{-j2\pi \mathbf{q} \cdot \mathbf{r}} d\mathbf{r}. \quad (16)$$

Here, U is the frequency representation of the phase-only SLM-modulation

$$U(\mathbf{q}) = \mathcal{F}\{u(\mathbf{r})\} \quad (17)$$

and the aperture function K incorporates the effects of the pupil parameters p through the the aperture function A , and the propagation kernel \mathcal{H} as

$$K(\mathbf{q}, p) = A\left(\mathbf{q} - \frac{\mathbf{s}}{\lambda f}, \frac{d}{\lambda f}\right) \mathcal{H}(\mathbf{q}; z). \quad (18)$$

The angular spectrum propagation kernel \mathcal{H} defined in the frequency domain is given by

$$\mathcal{H}(\mathbf{q}; z) = \begin{cases} e^{i\frac{2\pi}{\lambda} \sqrt{1 - ||\lambda\mathbf{q}||^2} z}, & \text{if } \sqrt{||\mathbf{q}||^2} < \frac{1}{\lambda}, \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

Relationship to Focal Stack and STFT Supervision. The propagation kernel expressed by Eqn. 16 is a generalization of both Focal Stack [10], [11] and STFT [48] supervision applied to hologram optimization. For focal stack supervision, the pupil parameters $p = (\mathbf{s}, d, z)$ are fixed so that pupil shift $\mathbf{s} = 0$, the aperture diameter d is equal to the eyebox size, and the depth is discretized into N -layers over $z \in [z_{\min}, z_{\max}]$. As discussed in Sec. 3.3, STFT supervision closely resembles Eq. 16, but the kernel $K(\mathbf{q}, p)$ is equal to the Fourier Transform of the STFT window $W(\mathbf{q})$, and pupil shift positions are computed over a grid of positions $\mathbf{s} \in [d_{\min}, d_{\max}]$ that is determined by the window size and FFT algorithm. However, this analysis reveals that a correct implementation of STFT supervision would require the window $W(\mathbf{q})$ to also be used for generating projections of the Light Field following Eq. 3. As a result, Eq. 11 is not strictly valid unless the window function is a jinc function, the 2D polar analog of the sinc function, whose width depends on the wavelength:

$$w(\mathbf{r}) = \frac{d}{\lambda f} \text{jinc}\left(\frac{d\mathbf{r}}{\lambda f}\right), \quad (20)$$

so that its Fourier transform is a circular aperture:

$$W(\mathbf{q}) = A\left(\mathbf{q}, \frac{d}{\lambda f}\right). \quad (21)$$

Focal stack and STFT supervision are, therefore, specific cases of supervising with random pupil positions, and, as a result, they overfit to specific pupil conditions, which we show does not generalize well to arbitrary viewing

conditions, see in Figs. 4 and 7. Furthermore, because Focal Stack supervision is a special case of supervising with random pupil states, we introduce a new LF2FS algorithm that computes Focal Stacks on the fly for supervision by varying only the defocus parameter of pupil states and holding the rest constant.

4.3 Implementation

Experimental Setup. We employ a conventional setup for Near-Eye Hologram as used in [24]. The only addition is a 2D translation and a motorized iris placed in the Fourier-plane of the 4f system to measure parallax and depth-of-field effects. For exact setup details, we refer to the supplementary.

Implementation details. Optimization is implemented using ADAM [56] using automatic-differentiation following Wirtinger Holography [25] with the standard L_2 -loss used in all experiments. Further implementation details on our and the comparison method STFT [48] are found in the supplementary.

5 RESULTS

The proposed method aims to find an optimal hologram that is optimized for the best average image quality over a four-dimensional pupil state-space, which includes diameter, location, and focus. For all the simulations and experimental results in the paper, the pupil parameters are chosen such that the defocus range is $z_{min} = 0\text{mm}$, $z_{max} = 15\text{mm}$ and the aperture range is $d_{min} = 2\text{mm}$, $d_{max} = 20\text{mm}$. This is slightly smaller than the smallest eye-box size (22mm) that is produced by the blue channel (440nm) under the 400mm focal length that was employed in the setup with an 8um SLM pitch.

Comparison to Smooth-Phase Holograms. Our paper focuses on algorithms that produce random-phase holograms such that the maximal possible eyebox is filled. Such holograms inherently exhibit speckle and are further harder to control due to physical effects such as field-fringing [57], [58]. We expect that future research on how to combine [24], with random phase holograms. Recent holographic literature such as algorithms [10], [32], [47] show best-in-class image quality, but rely heavily on smooth-phase to achieve speckle-reduction, which leads to a localized eye-box. As these holograms are unlikely to achieve real 3D holography [30], we do not compare against them in our main manuscript. However, we have added further justification and discussion of this in the supplementary materials, including a simulation using our implementation of the method from [10] that demonstrates how significant pupil variations are produced by methods that optimize for smooth-phase holograms.

Quantitative Validation (Simulation). To obtain quantitative results, we optimize our method on 85 different Light Fields from the DeepFocus [33] dataset. This dataset includes a variety of scenes with randomly sampled objects and textures. Figure 4 presents a comparison of our method with Focal Stack and STFT-supervision [48]. Our LF2FS Focal Stack supervision method is similar to [10], [11], however there is one significant difference: We compute our focal stack directly from the Light Field to ensure physically

correct defocus cues. Layered approaches with inaccurate synthetic defocus, create a mismatch between projection operators which will introduce incorrect depth cues. Our findings validate that the proposed algorithm fares favorably for random pupil states other than those for which STFT-supervision and LF2FS are overfitted for. Our SLFH method produces the *best-worse case performance for all pupil states*, as the best average case performance for random pupil states. Furthermore, the variance is significantly reduced for our approaches. This verifies that our algorithm works better for a large range of viewing conditions.

Experimental Results (Deep Focus). Experimental results from our prototype experimental holography setup are shown for the Deep Focus dataset in Fig 2. The images are captured with a pupil diameter of 8mm at positions 0mm (*back-focus*) and 12mm (*front-focus*). In addition we extracted epipolar images by capturing a 1D-trajectory from left to right using again an 8mm-sized pupil for both front and back focus. Additionally, we encourage the reader to view the videos found in the Supplementary Material as differences between the result are best visible in animation. Our SLFH method produces fewer color artifacts than either STFT or Focal Stack supervision.

Experimental and Simulation Results (Robot Scene). We further evaluate our method on a blender scene in both simulation and experiments. For experimental results, we evaluated the hologram with a focal stack trajectory as shown in Fig. 6. For this, we use five centered focus positions with a large aperture (80% of eyebox) and sample the volume in equidistant steps from 0mm to 12mm. We further evaluate the hologram in *simulation* in Fig. 7 for five randomly chosen pupil positions. The results validate that our stochastic pupil sampling algorithm can optimize holograms with correct parallax information for large variety of pupil states.

Artifacts at the edge of the eye-box In the supplementary video, we show the parallax that is created by one hologram when we sample the eyebox from left to right. In those, one can see that STFT is doing particularly poor at the edges of the eyebox. This is because in order to make small propagation distances work, we had to introduce an additional bandpass filter for STFT/LF2FS, which allow some energy to diffract into no-care-areas. Without this adaption, STFT cannot form meaningful images as shown in 10. Our proposed SLFH algorithm doesn't suffer from this and the complete eyebox can be used during optimization.

Large SBP (16k x 16k) Holograms in Far-Field Configuration. Current state-of-the-art displays are constrained by the etendue, which is determined by the product of the number of pixels and their pixel-pitch, also known as the Space Bandwidth Product (SBP). As a result, there is a trade-off between parallax over a larger eye box and resolution/defocus. This is a practical limitation for our prototype constrained by current technology, but it is not a fundamental limitation to our proposed algorithm. With advances in technology, we can expect improvements in display technology as well as system design [3], [5] As the system etendue increases, this results in need to perform large matrix multiplications (FFTs), which can quickly exhaust available GPU memory. This problem can be mitigated by optimizing for Far-Field Holograms where the angular spectrum corresponds to the field created by the SLM. With a small change

in the forward model and the same loss-function, we can directly supervise a large SBP hologram by stochastically sampling pupils over the Fourier-domain. Our pupil-aware Far-Field approach is an extension to [5], [7] who discuss iterative algorithms in Far-Field configuration with either multi-layer or a light field loss similar to STFT [48] with a fixed pupil grid. These works limit their discussion to only a fixed set of pupils, while we outline how Far-fields holograms with photometric consistency over the whole pupil space. We show an example of such a simulation for a $16k \times 16k$ hologram in Fig. 9.

6 DISCUSSION

We propose investigating the issue of CGH-optimization from a machine-learning perspective, specifically with regard to overfitting. Given that the capacity of a hologram is constrained by the available SBP, there is no single, optimal solution. In the context of our proposed approach of stochastic light-field holography optimization, the objective is to identify a hologram that yields the highest average image quality across the entire possible space of pupil states. Our findings indicate that, while overall correct vision cues are maintained across the entire state-space, performance is inferior when evaluated specifically at the pupil states for which other algorithms were optimized. For instance, a focal-stock loss approach may achieve near-perfect reconstruction *at the depths for which it was optimized for*, but fails to accurately reproduce parallax and depth-of-field effects. Similarly, the STFT-method demonstrates superior performance for given pupil positions around its stand-off distance, but experiences a significant decline in image quality when evaluated at larger propagation distances.

Relation to Ptychography. We would like to note that our method has a striking resemblance with a different technique known from Scientific Imaging called X-Ray Ptychography [59]. There, the forward model and optimization are quite similar as a complex waveform is reconstructed from intensity-only projections sampled at different, overlapping pupil positions. One can think of X-Ray Ptychography as a dual problem in imaging compared to the holographic display problem. Furthermore, X-Ray Ptychography often deals with very high resolutions, which requires highly parallelized, distributed optimization frameworks [60] that significantly accelerate computation. Such an increase in performance could potentially enable the development of human-sized, holographic displays.

7 FUTURE WORK

Light-Field Representation. In this study, we employed a traditional representation of a Light Field and applied the shift-and-add algorithm to synthesize new views. However, the accuracy of the Light Field representation is constrained by its angular and spatial resolution. Due to limitations in terms of memory, we encountered challenges such as aliasing during refocusing (details in Supplementary). There are approaches in recent Graphics research to overcome these limitations [33], [61], [62]. Among those are removing angular aliasing [61], directly synthesize refocused images using neural-networks [33], super-resolve in angular resolution using prior information [63] or fast neural-radiance fields [62] storing a highly compressed light-field.

Speckle Reduction. When compared to our simulation, the image quality of our experimental results is significantly lower. This is due to unmodeled physical effects such as cross-talk in the LCOS-SLMs. Although camera-in-the-loop methods have been proposed, physically accurate calibration with full-random phase holograms is a novel topic [64] and further research is needed.

Notably, speckle remain a fundamental challenge to holographic Light Field displays. We use time averaging to reduce speckle artifacts, and implemented our algorithms at video rates with new high speed SLMs is a promising direction for future work. Likewise, exploring other methods of speckle reduction [65] via incoherent averaging using partially coherent illumination [65] or subjective speckle optimization [66] is also a promising avenue of exploration.

Temporal Multiplexing. Recent papers propose to use temporal multiplexing to either do simultaneous color [67] or reduce the overall required framerate [68] using joint optimization techniques. future work could investigate how such approaches can be applied to our SLFH-algorithm to further reduce the number of temporal frames required to suppress speckle noise.

Etendue Expanders. Etendue for holographic displays remains impractically small due to the limited pixel count of commercially available SLMs. A further promising direction for future work is to explore our stochastic pupil-sampling cgh algorithms together with etendue expanders [4] that can help meet AR/VR requirements for combined eyebox size and FoV [69].

8 CONCLUSION

We introduce a new optimization method for generating light-field holograms using stochastic pupil-sampling within a Gradient-Descent type optimization. Simulation results demonstrating the effectiveness of the proposed method in near-eye display and large etendue setups in far-field configurations were presented. Experimental results validate that our algorithm is able to provide accurate depth cues such as parallax and defocus under a large variety of different pupil states. All our findings confirm that the proposed method produces the best-worse case performance for all pupil states, as well as the best average case performance for random pupil states – corresponding to average viewing conditions.

REFERENCES

- [1] M. Walton, P. Hao, M. Vermeulen, F. Willomitzer, and O. Cossairt, “Characterizing the immaterial. noninvasive imaging and analysis of stephen benton’s hologram engine no. 9,” *arXiv*, 2021.
- [2] F. Zhong, A. Jindal, Ö. Yöntem, P. Hanji, S. Watt, and R. Mantiu, “Reproducing reality with a high-dynamic-range multi-focal stereo display,” *ACM Transactions on Graphics*, vol. 40, no. 6, 2021.
- [3] G. Kuo, L. Waller, R. Ng, and A. Maimone, “High resolution étendue expansion for holographic displays,” *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, p. 66, 2020.
- [4] S.-H. Baek, E. Tseng, A. Maimone, N. Matsuda, G. Kuo, Q. Fu, W. Heidrich, D. Lanman, and F. Heide, “Neural etendue expander for ultra-wide-angle high-fidelity holographic display,” *arXiv*, 2021.
- [5] S. Monin, A. C. Sankaranarayanan, and A. Levin, “Exponentially-wide étendue displays using a tilting cascade,” in *ICCP*, 2022.

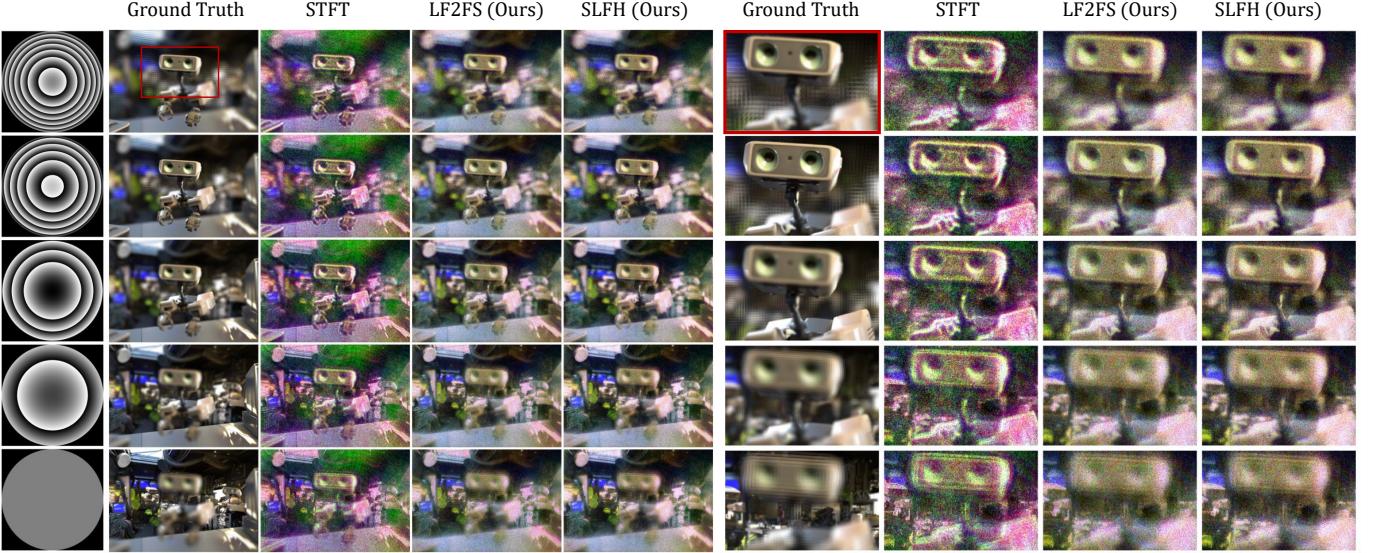


Fig. 6: Experiment: Focal-stack evaluation: We show experimental results for SOTA algorithms evaluated with a wide open aperture and defocus varying from $z \in [0\text{mm}, 13\text{mm}]$. The STFT training method produces significant color artifacts, while our new SLFH method and our implementation of Focal Stack supervision (LF2FS) produce high quality results with similar performance.

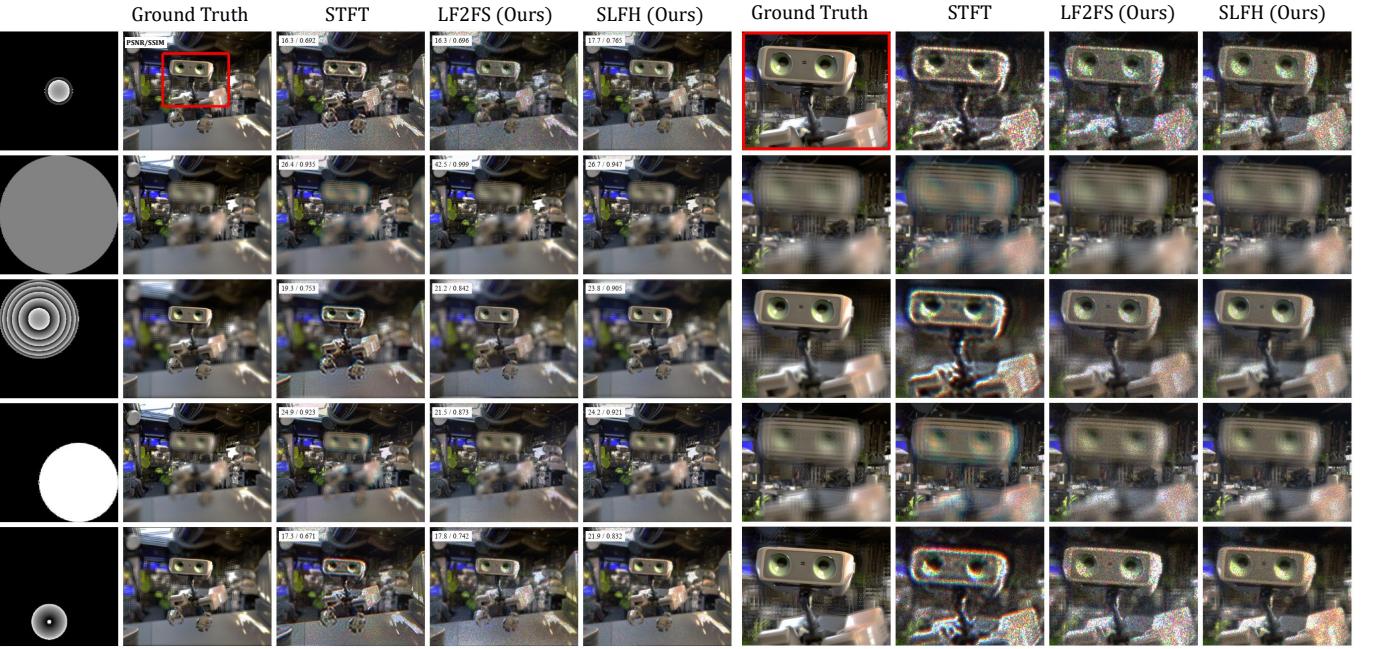


Fig. 7: Simulation: Comparison with state-of-the art: We compare our proposed method with STFT-lightfield [48] and our novel implementation of Focal Stack supervision (LF2FS). SOTA algorithms perform close-to-perfect for the pupil-states they were over-fitted for, but fail to produce accurate results for samples they have not seen during optimization. As our proposed algorithm stochastically supervises with small to medium-sized pupils (diameter was 10% to 40% of the eyebox), we achieve overall superior image fidelity for arbitrary pupil-states. We evaluate the performance by testing with a set of random pupil states shown in the first column. Our SLFH algorithm produces significantly less noise than both LF2FS and STFT algorithms, as can be seen by PSNR/SSIM numbers in the left hand corner as well as visual comparisons.

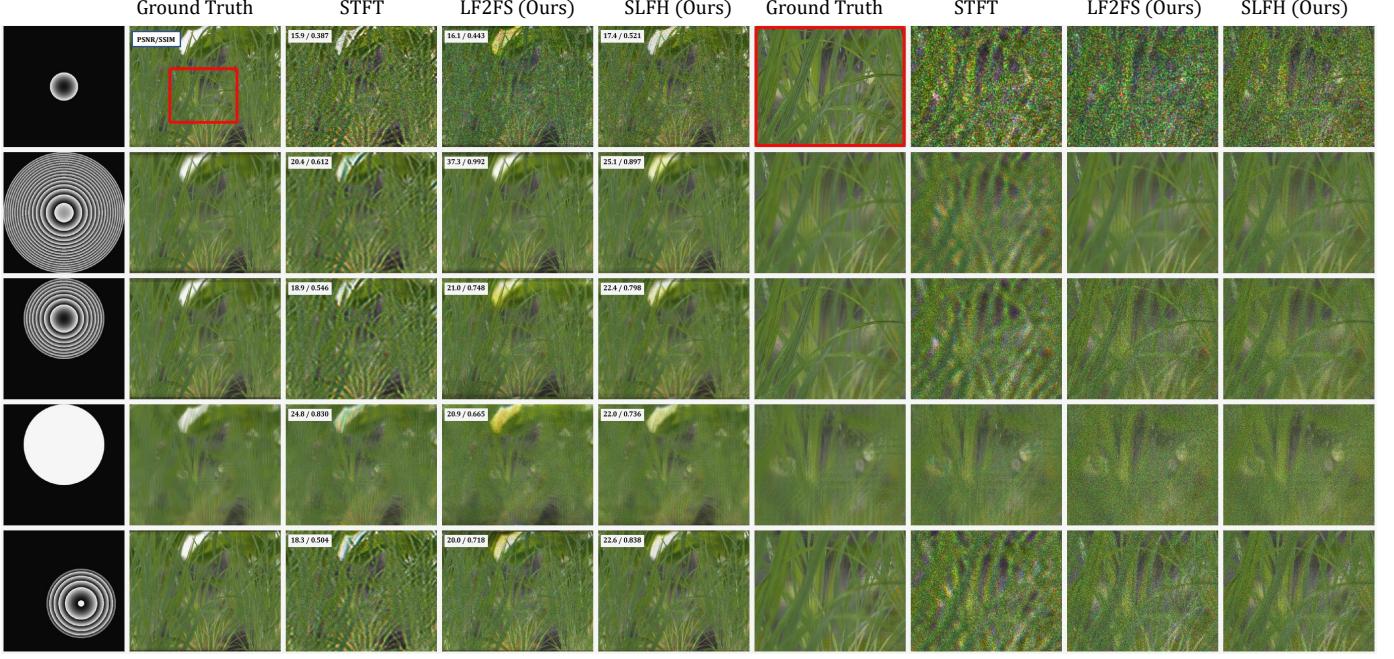


Fig. 8: **Simulation: Dense occlusions:** We evaluate the performance of SOTA CGH algorithms on a scene with dense occluders consisting of a cat hiding behind dense blades of grass. We evaluate the performance by testing with a set of random pupil states shown in the first column. Our algorithm produces significantly less noise than both LF2FS and STFT algorithms, as can be seen by PSNR/SSIM numbers in the left hand corner as well as visual comparisons.

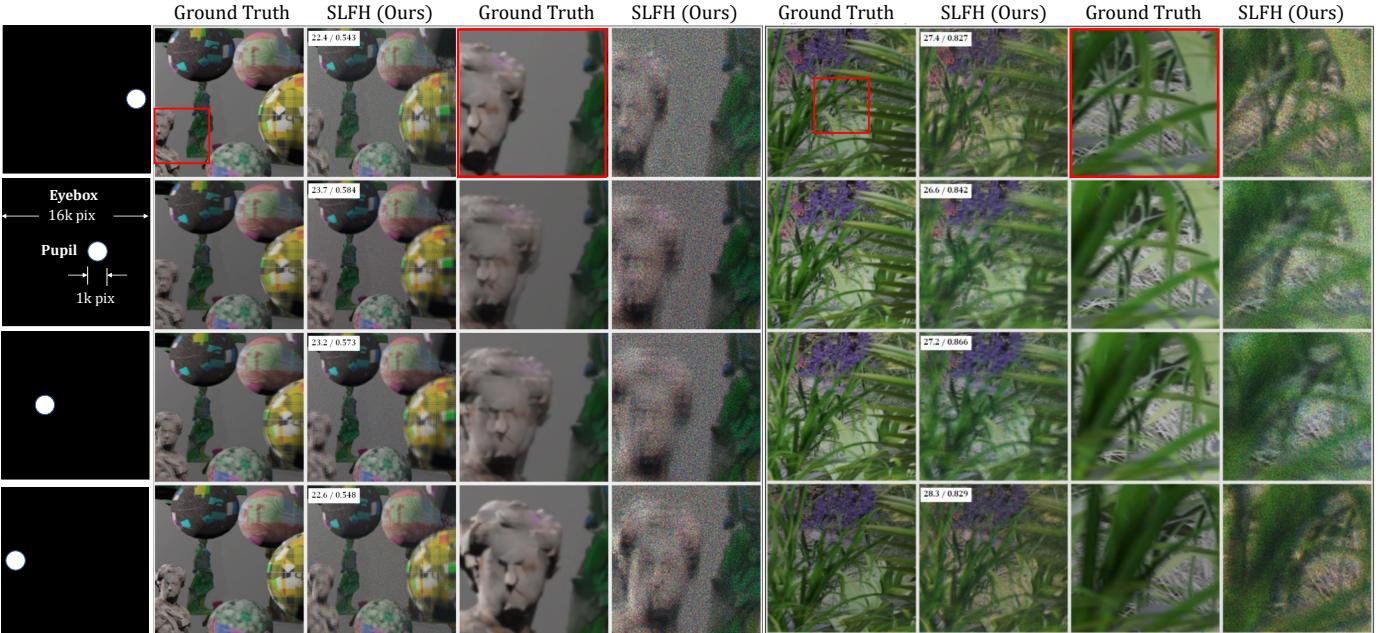


Fig. 9: **Simulation: Large eyebox results:** Here we show the benefit of our SLFH approach for optimizing over a very large eyebox in a Fourier CGH configuration. The hologram is placed in the pupil of the imaging system, and is 16k x 16k pixels, which is too large to optimize for using conventional Fourier CGH algorithms. However, our method can efficiently optimize holograms with large Space-Bandwidth Product (SBP). The figure demonstrates that larger SBP produces high quality results with significant parallax.



Fig. 10: Artefacts in competing methods with and without no-care-areas at the edge of the eye-box (Experiment). Plain STFT/Focal-Stack cannot produce holograms at close propagation distances (left). Only by introducing a band-pass filter during propagation (right), we can compute holograms for those. Even with correction, at the edge of the eye-box, STFT and FS strong artifacts streak artifacts or noisy images, respectively.

- [6] Y. Jo, D. Yoo, D. Lee, M. Kim, and B. Lee, "Multi-illumination 3d holographic display using a binary mask," *Optics Letters*, vol. 47, no. 10, pp. 2482–2485, 2022.
- [7] 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.
- [8] P. Chakravarthula, S.-H. Baek, E. Tseng, A. Maimone, G. Kuo, F. Schiffrers, N. Matsuda, O. Cossairt, D. Lanman, and F. Heide, "Pupil-aware holography," *arXiv preprint arXiv:2203.14939*, 2022.
- [9] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Ph.D. dissertation, Stanford University, 2005.
- [10] S. Choi, M. Gopakumar, Y. Peng, J. Kim, and G. Wetzstein, "Neural 3d holography: Learning accurate wave propagation models for 3d holographic virtual and augmented reality displays," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, pp. 1–12, 2021.
- [11] K. Kavaklı, Y. Itoh, H. Urey, and K. Aksit, "Realistic defocus blur for multiplane computer-generated holography," *arXiv preprint arXiv:2205.07030*, 2022.
- [12] R. Ng, "Fourier slice photography," *Arxiv*, pp. 735–744, 2005.
- [13] L. Lesem, P. Hirsch, and J. Jordan, "The kinoform: a new wavefront reconstruction device," *IBM Journal of Research and Development*, vol. 13, no. 2, pp. 150–155, 1969.
- [14] R. W. Gerchberg, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.
- [15] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied optics*, vol. 21, no. 15, pp. 2758–2769, 1982.
- [16] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Hybrid projection-reflection method for phase retrieval," *JOSA A*, 2003.
- [17] R. Lane, "Phase retrieval using conjugate gradient minimization," *Journal of Modern Optics*, vol. 38, no. 9, pp. 1797–1813, 1991.
- [18] Z. Wen, C. Yang, X. Liu, and S. Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," *Inverse Problems*, vol. 28, no. 11, p. 115010, 2012.
- [19] S. Marchesini, Y.-C. Tu, and H.-t. Wu, "Alternating projection, ptychographic imaging and phase synchronization," *Applied and Computational Harmonic Analysis*, vol. 41, no. 3, pp. 815–851, 2016.
- [20] J. Zhang, N. Pégarde, J. Zhong, H. Adesnik, and L. Waller, "3d computer-generated holography by non-convex optimization," *Optica*, vol. 4, no. 10, pp. 1306–1313, 2017.
- [21] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.
- [22] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," *IEEE Transactions on Information Theory*, 2018.
- [23] S. Bahmani and J. Romberg, "Phase retrieval meets statistical learning theory: A flexible convex relaxation," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 252–260.
- [24] 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.
- [25] 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.
- [26] H. Zhang, Y. Zhao, L. Cao, and G. Jin, "Layered holographic stereogram based on inverse fresnel diffraction," *Applied optics*, vol. 55, no. 3, pp. A154–A159, 2016.
- [27] Y. Zhao, L. Cao, H. Zhang, D. Kong, and G. Jin, "Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method," *Optics express*, vol. 23, no. 20, pp. 25440–25449, 2015.
- [28] M. H. Ebyposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pégarde, "Deepcgh: 3d computer-generated holography using deep learning," *Optics Express*, vol. 28, no. 18, 2020.
- [29] M. Makowski, M. Sytek, A. Kolodziejczyk, G. Mikula, and J. Suszek, "Iterative design of multiplane holograms: experiments and applications," *Optical Engineering*, vol. 46, no. 4, pp. 1–6, 2007.
- [30] D. Kim, S.-W. Nam, B. Lee, J.-M. Seo, and B. Lee, "Accommodative holography: improving accommodation response for perceptually realistic holographic displays," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.
- [31] D. Yoo, Y. Jo, S.-W. Nam, C. Chen, and B. Lee, "Optimization of computer-generated holograms featuring phase randomness control," *Optics Letters*, vol. 46, no. 19, pp. 4769–4772, 2021.
- [32] 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.
- [33] L. Xiao, A. Kaplanyan, A. Fix, M. Chapman, and D. Lanman, "Deepfocus: Learned image synthesis for computational display," in *ACM SIGGRAPH 2018 Talks*, 2018, pp. 1–2.
- [34] H. Zhang, Y. Zhao, L. Cao, and G. Jin, "Fully computed holographic stereogram based algorithm for computer-generated holograms with accurate depth cues," *Optics express*, vol. 23, 2015.
- [35] M. Lucente and T. A. Galvean, "Rendering interactive holographic images," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 1995, pp. 387–394.
- [36] M. Yamaguchi, H. Hoshino, T. Honda, and N. Ohyama, "Phase-added stereogram: calculation of hologram using computer graphics technique," in *Proc. SPIE*, vol. 1914, 1993, pp. 25–31.
- [37] Q. Y. Smithwick, J. Barabas, D. E. Smalley, and V. M. Bove, "Interactive holographic stereograms with accommodation cues," in *Practical Holography XXIV: Materials and Applications*, vol. 7619. International Society for Optics and Photonics, 2010, p. 761903.
- [38] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 220, 2013.
- [39] N. Padmanaban, Y. Peng, and G. Wetzstein, "Holographic near-eye displays based on overlap-add stereograms," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, p. 214, 2019.
- [40] D. Blinder and P. Schelkens, "Accelerated computer generated holography using sparse bases in the stft domain," *Optics express*, vol. 26, no. 2, pp. 1461–1473, 2018.
- [41] J.-H. Park and M. Askari, "Non-hologram-based computer generated hologram from light field using complex field recovery technique from wigner distribution function," *Optics express*, vol. 27, 2019.
- [42] P. Chakravarthula, E. Tseng, T. Srivastava, H. Fuchs, and F. Heide, "Learned hardware-in-the-loop phase retrieval for holographic near-eye displays," *ACM TOG*, vol. 39, no. 6, 2020.
- [43] L. Shi, F.-C. Huang, W. Lopes, W. Matusik, and D. Luebke, "Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3d computer graphics," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–17, 2017.
- [44] K. Min, D. Min, and J.-H. Park, "Wigner inverse transform based computer generated hologram for large object at far field from its

- perspective light field," *Optics Communications*, vol. 532, p. 129229, 2023.
- [45] Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light: Science & Applications*, vol. 7, no. 2, p. 17141, 2018.
- [46] R. Horisaki, R. Takagi, and J. Tanida, "Deep-learning-generated holography," *Applied optics*, vol. 57, no. 14, pp. 3859–3863, 2018.
- [47] 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.
- [48] 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.
- [49] P. Chakravarthula, E. Tseng, H. Fuchs, and F. Heide, "Holog-free holography," *ACM Transactions on Graphics (TOG)*, 2022.
- [50] R. Ziegler, S. Bucheli, L. Ahrenberg, M. Magnor, and M. Gross, "A bidirectional light field-hologram transform," in *Computer Graphics Forum*, vol. 26. Wiley Online Library, 2007, pp. 435–446.
- [51] Z. Zhang and M. Levoy, "Wigner distributions and how they relate to the light field," in *2009 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2009, pp. 1–10.
- [52] T. Cuypers, R. Horstmeyer, S. B. Oh, P. Bekaert, and R. Raskar, "Validity of wigner distribution function for ray-based imaging," in *IEEE ICCP*. IEEE, 2011, pp. 1–9.
- [53] M. J. Bastiaans *et al.*, "Wigner distribution in optics," 2009.
- [54] S. Hamann, L. Shi, O. Solgaard, and G. Wetzstein, "Time-multiplexed light field synthesis via factored wigner distribution function," *Optics letters*, vol. 43, no. 3, pp. 599–602, 2018.
- [55] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.
- [56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [57] S. Moser, M. Ritsch-Marte, and G. Thalhammer, "Model-based compensation of pixel crosstalk in liquid crystal spatial light modulators," *Optics express*, vol. 27, no. 18, pp. 25 046–25 063, 2019.
- [58] P. Schroff, A. La Rooij, E. Haller, and S. Kuhr, "Accurate holographic light potentials using pixel crosstalk modelling," *Scientific Reports*, vol. 13, no. 1, p. 3252, 2023.
- [59] P. Sheldigeri, F. Schiffers, S. Barutcu, P. Ruiz, A. K. Katsaggelos, and O. Cossairt, "Improving acquisition speed of x-ray ptychography through spatial undersampling and regularization," in *ICIP*. IEEE, 2021, pp. 2968–2972.
- [60] Y. S. Nashed, T. Peterka, J. Deng, and C. Jacobsen, "Distributed automatic differentiation for ptychography," *Procedia Computer Science*, vol. 108, pp. 404–414, 2017.
- [61] Z. Xiao, Q. Wang, G. Zhou, and J. Yu, "Aliasing detection and reduction in plenoptic imaging," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3326–3333.
- [62] A. Yu, R. Li, M. Tancik, H. Li, R. Ng, and A. Kanazawa, "Plenoc-trees for real-time rendering of neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5752–5761.
- [63] P. Sheldigeri, F. Schiffers, S. Ghosh, O. Cossairt, and K. Mitra, "Selfvi: Self-supervised light-field video reconstruction from stereo video," in *ICCV*, 2021, pp. 2491–2501.
- [64] C. Jang, K. Bang, M. Chae, B. Lee, and D. Lanman, "Waveguide holography: Towards true 3d holographic glasses," *arXiv preprint arXiv:2211.02784*, 2022.
- [65] Y. Peng, S. Choi, J. Kim, and G. Wetzstein, "Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration," *Science advances*, vol. 7, no. 46, p. eabg5040, 2021.
- [66] P. Chakravarthula, Z. Zhang, O. Turşun, P. Didyk, Q. Sun, and H. Fuchs, "Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 11, 2021.
- [67] E. Markley, N. Matsuda, F. Schiffers, O. Coissart, and G. Kuo, "Simultaneous color holography," *arXiv:2303.11287*, 2023.
- [68] K. Kavaklı, L. Shi, H. Ürey, W. Matusik, and K. Akşit, "Holohdr: Multi-color holograms improve dynamic range," *arXiv preprint arXiv:2301.09950*, 2023.
- [69] X. Xia, Y. Guan, A. State, P. Chakravarthula, T.-J. Cham, and H. Fuchs, "Towards eyeglass-style holographic near-eye displays with statically expanded eyebox," in *ISMAR*. IEEE, 2020.

Florian Schiffers is a Ph.D. candidate at Northwestern University and has undertaken the research of this paper during a research internship at Reality Labs Research, Meta. He received dual M.Sc. degrees, in Physics and Advanced Optical Technologies, from FAU Erlangen in Germany in 2017. Prior to his ongoing doctoral study, Schiffers held a research assistant role at Siemens Healthineers.

Praneeth Chakravarthula is a research scholar at Princeton University in the Princeton Computational Imaging Lab. He obtained his Ph.D. from UNC Chapel Hill on novel near-eye displays for virtual and augmented reality. His research interests lie at the intersection of optics, perception, and graphics. Prior to joining UNC, he obtained my B.Tech and M.Tech degrees in Electrical Engineering from IIT Madras.

Nathan Matsuda is a research scientist at Meta developing computational camera and display systems that support immersive visual experiences in VR.

Grace Kuo is a research scientist at Reality Labs Research, Meta working on the joint design of optical hardware and algorithms for imaging and display systems. She earned her PhD from the Department of Electrical Engineering and Computer Sciences at UC Berkeley, advised by Dr. Laura Waller and Dr. Ren Ng.

Ethan Tseng is currently a Ph.D. candidate in Computer Science at Princeton University, supervised by Prof. Felix Heide. He received his B.S. in Electrical and Computer Engineering from Carnegie Mellon University.

Douglas Lanman is the Senior Director of Display Systems Research at Reality Labs Research, Meta, where he leads investigations into advanced display and imaging technologies. His prior research focuses on head-mounted displays, glasses-free 3D displays, light field cameras, and active illumination for 3D reconstruction and interaction. Douglas holds a B.S. in Applied Physics from Caltech, and M.S. and Ph.D. degrees in Electrical Engineering from Brown University.

Felix Heide is an assistant professor at Princeton University, where he leads the Computational Imaging Lab. He obtained his BS and MS from the University of Siegen and his PhD from the University of British Columbia, where his doctoral dissertation won the Alain Fournier PhD Dissertation Award and the SIGGRAPH Outstanding Doctoral Dissertation Award. He completed his postdoc at Stanford University and has been at Princeton since 2020.

Oliver S. Cossairt is a research scientist at Reality Labs Research, Meta and an Associate Professor in the Electrical Engineering and Computer Science Department at Northwestern University. He holds an M.S. from the MIT Media Lab, and a Ph.D. in Computer Science from Columbia University. With a background as an Optical and Software Engineer at Actuality Systems, Oliver has received several accolades, including the NSF Graduate Research Fellowship, ICCP Best Paper and Honorable Mention, and an NSF CAREER Award.