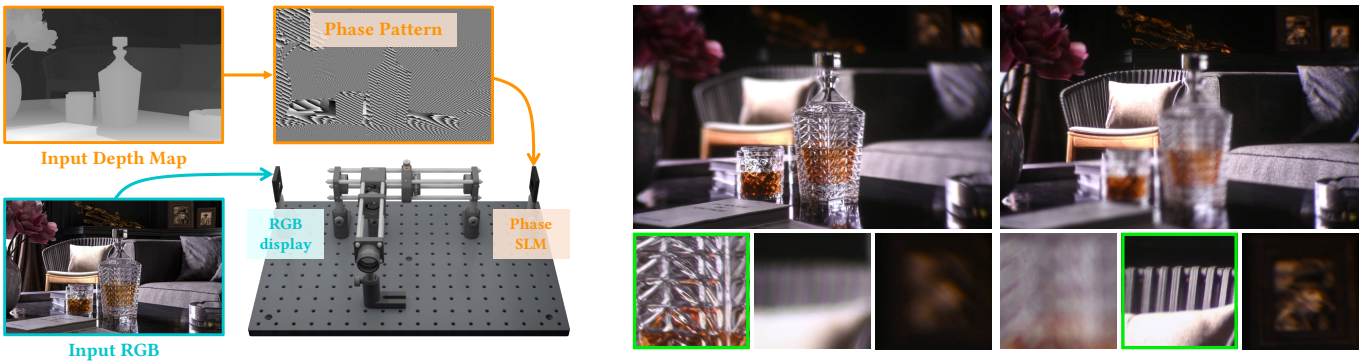
Split-Lohmann Multifocal Displays

YINGSI QIN, Carnegie Mellon University, USA
WEI-YU CHEN, Carnegie Mellon University, USA
MATTHEW O'TOOLE, Carnegie Mellon University, USA
ASWIN C. SANKARANARAYANAN, Carnegie Mellon University, USA



(a) Split-Lohmann Multifocal Display

(b) Photographs captured from lab prototype at near (left) and mid (right) focus

Fig. 1. **Split-Lohmann Multifocal Displays.** This paper designs a near-eye 3D display capable of generating a dense set of focal planes simultaneously using a single exposure. Our key innovation is an optical arrangement, which we refer to as *Split Lohmann*, that enables a high degree of local control in depth selection on a traditional display. Given a scene in the form of an RGBD image, we use a RGB display to project the color image while a phase spatial light modulator (SLM) shows a simple analytical function of the depth map. This produces a 3D multifocal scene *without requiring time multiplexing*. Shown above in (b) are focus stack images acquired from our lab prototype for the RGBD scene in (a) using an observer camera that is not tethered to the display. Split-Lohmann displays enjoy a number of desirable features including the ability to handle scenes with complex depth maps over a large working range, at high spatial and depth resolutions, and a large étendue. The extreme simplicity of our computational pipeline also enables real-time operations for interactive 3D content. (Credits: 3D scene courtesy of "Entity Designer" at Blender Market)

This work provides the design of a multifocal display that can create a dense stack of focal planes in a single shot. We achieve this using a novel computational lens that provides spatial selectivity in its focal length, i.e, the lens appears to have different focal lengths across points on a display behind it. This enables a multifocal display via an appropriate selection of the spatially-varying focal length, thereby avoiding time multiplexing techniques that are associated with traditional focus tunable lenses. The idea central to this design is a modification of a Lohmann lens, a focus tunable lens created with two cubic phase plates that translate relative to each other. Using optical relays and a phase spatial light modulator, we replace the physical translation of the cubic plates with an optical one, while simultaneously allowing for different pixels on the display to undergo different amounts of translations and, consequently, different focal lengths.

We refer to this design as a Split-Lohmann multifocal display. Split-Lohmann displays provide a large étendue as well as high spatial and depth resolutions; the absence of time multiplexing and the extremely light computational footprint for content processing makes it suitable for video and interactive experiences. Using a lab prototype, we show results over a wide range of static, dynamic, and interactive 3D scenes, showcasing high visual quality over a large working range.

CCS Concepts: • **Hardware** → **Displays and imagers**; • **Computing methodologies** → **Virtual reality**.

Additional Key Words and Phrases: Multifocal displays, Computational displays, Vergence-accomodation conflict, Lohmann lenses

ACM Reference Format:
Yingsi Qin, Wei-Yu Chen, Matthew O'Toole, and Aswin C. Sankaranarayanan. 2023. Split-Lohmann Multifocal Displays. *ACM Trans. Graph.* 42, 4, Article 1 (August 2023), 18 pages. https://doi.org/10.1145/3592110

Authors' addresses: Yingsi Qin, yingsiq@andrew.cmu.edu, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA; Wei-Yu Chen, weiyuc@andrew.cmu.edu, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA; Matthew O'Toole, mpotoole@cmu.edu, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA; Aswin C. Sankaranarayanan, saswin@andrew.cmu.edu, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
0730-0301/2023/8-ART1 $15.00
https://doi.org/10.1145/3592110

1 INTRODUCTION

As we gaze at the world around us, parts of it come into focus and defocus. The eye has a lens and, by changing its focal length, it can focus at objects at a specific depth, resolving them at sharp detail, while points at other depths go into defocus. This feature of the eye—referred to as its accommodation—is an important cue that 3D displays need to satisfy.

Satisfying the accommodation cue requires a display that can optically place content at its desired depth. From basic wave optics,

we can infer that the difference between points light sources placed at different depths is the curvature of the wavefronts incident on an observer. Specifically, a point light source produces a spherical wavefront, one that gets flatter as it propagates. Hence, to satisfy the accommodation of the eye, a 3D display needs to induce spherical wavefronts with depth-dependent curvatures onto each of the virtual scene points.

There are numerous approaches to enable accommodation in a display. Perhaps the simplest is to use a focus-tunable lens, as commonly done in multifocal [Akeley 2004; Liu et al. 2008; Love et al. 2009; Rolland et al. 2000] and varifocal [Akşit et al. 2017; Johnson et al. 2016; Padmanaban et al. 2017] displays, where a focus-tunable lens provides control over the curvature of the resulting wave. A display placed behind the lens can be made to appear from a different depth by appropriate selection of the focal length of the tunable lens. However, a focus-tunable lens provides no spatial selectivity, i.e., *all* pixels of the display are mapped to the same field curvature associated with some desired depth. A 3D scene is produced by sequencing through multiple focal planes by rapidly changing the focal length of the lens; the persistence of human vision allows these time-multiplexed focal planes to be perceived as one 3D scene. Time multiplexing presents its own set of challenges due to the need for high bandwidth electronics; for example, the frame rate of the display increases linearly with the number of focal planes that we aim to show. This raises the following question: Is it possible to create a "lens" that can simultaneously place different pixels of the display onto different depths?

This paper presents a unique system that enables focus tunability with spatial selectivity, i.e, a lens that appears to have different focal lengths over different parts of the display. We achieve this with a modified version of the so-called Lohmann lens [Lohmann 1970], which is a focus-tunable lens constructed out of two co-located cubic phase plates. When the two cubic plates are translated with respect to each other, they produce a quadratic phase plate—or a lens—whose focal length depends on the amount of displacement between them. A Lohmann lens does not offer any spatial selectivity; further, its reliance on mechanical motion often makes it an inferior choice to liquid lenses that are electronically controlled. Our contribution is a modification to the Lohmann lens that enables spatial selectivity, while replacing mechanical motion with optical translation. We refer to the resulting system as the *"Split-Lohmann"* multifocal display.

Split-Lohmann displays rely on the following modifications to Lohmann lenses. First, we physically separate the two cubic plates, while optically collocating them using a 4f relay. Second, we introduce a phase SLM in the Fourier/pupil plane of the 4f system. A phase ramp on the SLM, mimicking the effect of a prism or a diffraction grating, produces the same effect as mechanical motion since it is placed in the Fourier plane of the cubic plates. Changing the slope of the phase ramp allows us to *electronically* control the amount of translation between the cubic plates and, consequently, the focal length of the resulting lens. Third, to enable a spatially-varying focal length, we optically collocate a display onto this SLM and change the slope of the phase ramp on the SLM, *locally*. By varying the local slope of the phase ramp, different regions of the display can be associated with different translations between the cubic plates; in turn, each region observes a lens with a different focal length (and is placed at a different depth) thereby achieving the desired ability of spatial-selectivity of focal length.

Figure 1 shows results from our lab prototype. Our prototype produces multifocal scenes that enjoy a high spatial resolution, largely retaining the display's native resolution, at a depth resolution that is greater than 1/8−th of a diopter over a working range of four diopters, and a temporal resolution that is equal to the native frame rate of the display and SLM. It also has a large étendue that is limited only by the choice of optical lenses used.

Contributions. This paper presents the Split-Lohmann multifocal display, a single-shot 3D display capable of providing very high spatial-resolution in depth selectivity.

- *Programmable Split-Lohmann lens.* The centerpiece of our contributions is a novel Lohmann-style focus tunable lens that provides a spatially-varying focal length. We use this system to implement a single-shot near-eye multifocal display with a large working range and dense sampling of focal planes.
- *Design space analysis.* We provide an in-depth analysis of our multifocal display, including a characterization of various features (étendue, depth range and resolution) as well as artifacts.
- *Real-time content generation.* The proposed display has a near-analytical mapping of input RGBD content to the patterns shown on the display and phase SLM. This ultra-light computational footprint allows us to easily handle interactive scenes like 3D games at real time speeds with commodity hardware.
- *Lab prototype.* We design and implement a lab prototype and showcase its performance on virtual and interactive 3D scenes.

This advance in display design allows us to improve upon state-of-the-art in multifocal displays, enabling many of its desirable capabilities without requiring time multiplexing. We have released our code on the project website [Qin et al. 2023] to facilitate reproducibility and follow-up research.

Limitations. The proposed system exhibits artifacts near depth discontinuities due to the way that the display is mapped onto the SLM. Specifically, our system *requires* that one of the cubic phase plates appear in the pupil/Fourier plane of a 4f system mapping the display and the SLM; this cubic plate induces a large blur, in tens of microns in spatial extent in a typical design space. We characterize the effects and appearance of these aberrations in detail; in practical terms, however, they are not particularly distracting as we highlight in results obtained through simulation as well as our lab prototype. Our proposed system also critically relies on properties of the 4f relay, which results in a non-compact system. Achieving the compactness is likely a critical step toward practical adoption of the underlying technique. The results presented in this work only characterize the monocular performance of Split-Lohmann displays, namely, its ability to satisfy focus cues; to characterize its effectiveness in resolving vergence-accommodation conflicts [Hoffman et al. 2008], we would need to perform user studies that evaluate depth perception and immersion with the display. Finally, ours is a multifocal display and we inherent a number of limitations commonly seen in such displays; an example of this is the weakening of occlusion cues and contrast due to leakage of defocus across focal planes.

2 PRIOR WORK

We briefly discuss prior work in 3D displays, concentrating on techniques that share similarities with our work. We also discuss Lohmann lenses in detail, given their central role to this work.

2.1 Multifocal Displays

A multifocal display shows 3D content on multiple virtual or focal planes. There have been numerous implementations of multifocal displays; we introduce the most relevant ones to our work below, deferring to survey articles [Koulieris et al. 2019; Zhan et al. 2020] for a more exhaustive treatment of this topic.

At its core, this paper develops techniques that avoid time multiplexing, which can be achieved in a limited manner using polarization optics. Specifically, polarization-sensitive lenses can simultaneously provide two focal lengths, each encoded in a different polarization state; by controlling the polarization state of each display pixel, we can place content on one of two focal planes [Tan et al. 2018]. Polarization, however, is restricted to two independent modes, which restricts the display to two focal planes.

A different way to avoid time multiplexing is to sacrifice the spatial resolution of the display, instead. The early work of Akeley et al. [2004] use an arrangement of beamsplitters to place different parts of a display at different distances from the eyepiece. Cui and Gao [2017] use a 4f system between the display and the eyepiece, with a diffractive optical element in its Fourier plane. The diffractive optical element is designed to induce specific axial shifts on pre-determined regions of the display while laterally shifting them to the center of the optical axis. All of these designs share the same limitation in the form of reduced spatial resolution; specifically, the number of pixels that can be devoted to each focal plane is inversely proportional to the number of focal planes. Accommodation invariance [Konrad et al. 2017] provides a different approach to tackle ocular focusing; here, a focus-invariant blur is produced by sweeping the display across the working range within the persistence of vision. However, this invariably leads to a loss in the spatial resolution at which the display is perceived.

Focal surface displays [Matsuda et al. 2017] use a phase SLM, placed between the display and the eyepiece, to implement a lens with a spatially-varying focal length. Conceptually, this tackles the same problem as ours. However, the use of the SLM directly as a spatially-varying lens results in certain limitations. Focal surfaces are smooth. In part, this stems from limitations of the SLM and in part, this can be attributed to neighboring pixels on the display illuminating highly overlapping sets of SLM pixels. Extended scenes with depth discontinuities are split into multiple smooth surfaces and displayed using time multiplexing; in contrast, Split-Lohmann displays do not require time multiplexing, and handle extended scenes by quantizing the scene depth (as measured in diopters) at a high resolution. Compared to fixed and adaptive multifocal displays, focal surface displays provide significantly reduced depth errors with fewer time-multiplexed frames. Split-Lohmann pushes this trend further by allowing for even smaller depth error; for example, our lab prototype displays 50 focal planes over a 4 diopters working range with a theoretical max depth error of 0.04 diopters. Finally, finding the SLM phase patterns to realize focal surfaces, single or multiple, requires solving an optimization problem, which adds an additional computational overhead that our design avoids. As a consequence, we can easily port video and even interactive content on our prototype, as we show in Section 5. On the flip side, focal surface displays were implemented in a compact form factor, while ours is an extended lab prototype likely requiring fairly sophisticated optical engineering to be realized as a compact head mountable device. All things considered, focal surface displays provide an interesting alternative to our design, and we compare against it with both simulated as well as real results.

2.2 Lohmann and Alvarez Lenses

Lohmann [1970] and Alvarez [Alvarez 1978; Alvarez and Humphrey 1970] lenses provide an approach for realizing focus-tunable lenses using two translating cubic phase plates. The underlying principle behind both lenses is the interaction of two translating cubic phase plates. Suppose that we have two cubic plates, that are stack together, with height maps $h_1(x) = \frac{x^3}{c_0}$ and $h_2(x) = -\frac{x^3}{c_0}$, where c_0 determines the curvature of the cubic phase plate, with smaller values of c_0 leading to higher curvature. Since they are stacked together, the resulting phase modulation they induce is

$$e^{-j\frac{2\pi}{\lambda}(h_1(x)+h_2(x))(\eta-\eta_{\text{air}})}, \tag{1}$$

where λ is the operating wavelength; η and η_{air} are the refractive indices of the phase plate material and air, respectively. At the first glance, this choice of height maps seems counter-intuitive since $h_1(x) + h_2(x)$ is uniformly zero; that is, the stacked masks do not modulate the phase of the incident wavefront. However, if the two plates are translated in opposing directions, say by $\pm\Delta$, then the resulting phase modulation is given as

$$h_1(x+\Delta) + h_2(x-\Delta) = \frac{(x+\Delta)^3 - (x-\Delta)^3}{c_0} = \frac{6\Delta x^2 + 2\Delta^3}{c_0}. \tag{2}$$

Ignoring the constant term that is independent of x, and substituting Eq. (2) into Eq. (1), we get a phase modulation

$$e^{-j\frac{2\pi}{\lambda}\frac{12\Delta(\eta-\eta_{\text{air}})}{c_0}\frac{x^2}{2}}. \tag{3}$$

The effect of the translated phase plates is to add a quadratic phase onto the incident wavefront, which is identical to the effect of a lens with a focal length equal to

$$\frac{c_0}{12\Delta(\eta-\eta_{\text{air}})}. \tag{4}$$

Lenses of different focal lengths can be obtained simply by changing the amount of translation between the phase plates.

Lohmann and Alvarez lenses both use the principle of translating cubic plates to produce focus tunable lenses and have been used to create multifocal displays [Wilson and Hua 2019]; the key difference between them arises from how they extend to two-dimensional phase plates. Lohmann lenses produce a separable phase function by using a height profile

$$h(x,y) = \frac{x^3+y^3}{c_0}, \tag{5}$$

which requires the translation of the masks to be along the $x = y$ axis or equal amounts of translation along both axes. Alvarez uses a

slightly different arrangement that requires translation along the x-axis instead. The differences between these lenses have been studied extensively in prior work [Barbero 2009]. We use Lohmann lenses in our implementation; yet, to our understanding, our ideas can be implemented with Alvarez lenses with little difference.

2.3 Holographic displays

Holographic displays [Maimone et al. 2017] aim to create the wavefront associated with a scene using a phase SLM. While our proposed work also uses a phase SLM to shape the wavefront, there are some differences here that are worth elaborating.

In a holographic display, the incident light is typically (but not always) *spatially coherent*, in the form of a collimated beam that has zero étendue since its associated solid angle is zero. The SLM expands this solid angle by a small amount as determined by its pitch. This results in the classic limitation of the étendue being restricted by the number of pixels in the SLM. In contrast, the light incident on the SLM in Split-Lohmann is *spatially incoherent*—dense in space and angle with a solid angle determined by the display's cone of light as well as the numerical aperture of the optics. Practically, state-of-the-art results for holographic displays require careful modeling of imperfections in display hardware [Shi et al. 2022] using camera-in-the-loop optimization [Chakravarthula et al. 2020; Peng et al. 2020] and deep neural networks. In contrast, Split-Lohmann is lightweight in both regards, allowing even interactive applications.

Like all 3D displays, both approaches aim to create the wavefront associated with a scene. Holographic displays place the burden of this creation entirely on the SLM, benefitting from the compactness that this ensues. Split-Lohman uses more optical elements in the form of cubic phase plates to create the necessary quadratic wavefronts, with the SLM acting as a selection operator. This avoids a limited étendue, but at the cost of bulky optics.

3 SPLIT-LOHMANN DISPLAY

This section presents the main ideas underlying this work. We begin with a traditional multifocal display using focus tunable lenses, and progressively work our way to the proposed Split-Lohmann display.

3.1 Traditional Multifocal Displays

Figure 2(a) shows a schematic of a multifocal display, consisting of a translating display and an eyepiece. Suppose that the focal length of the eyepiece is f_e, then placing the display at a distance f_e from the eyepiece will result in display pixels being collimated; the net result is that the display is optically placed at infinity. Moving the display closer to the eyepiece results in it being optically placed at a distance closer than infinity. The eye is placed at the *focus plane* of the eyepiece, i.e., a distance f_e from the lens. This choice has certain benefits over placing the eye right against the eyepiece, which we discuss in detail in Section 4.2.

Figure 2(b) provides a way to implement the multifocal display shown in Figure 2(a), using focus tunable lenses that are realized as Lohmann lenses. Here, instead of physically translating the display, we optically do so by using a 4f relay with a focus-tunable lens in the pupil/Fourier plane. When the focus-tunable lens has zero power (or infinite focal length), the 4f lens system produces a copy of the

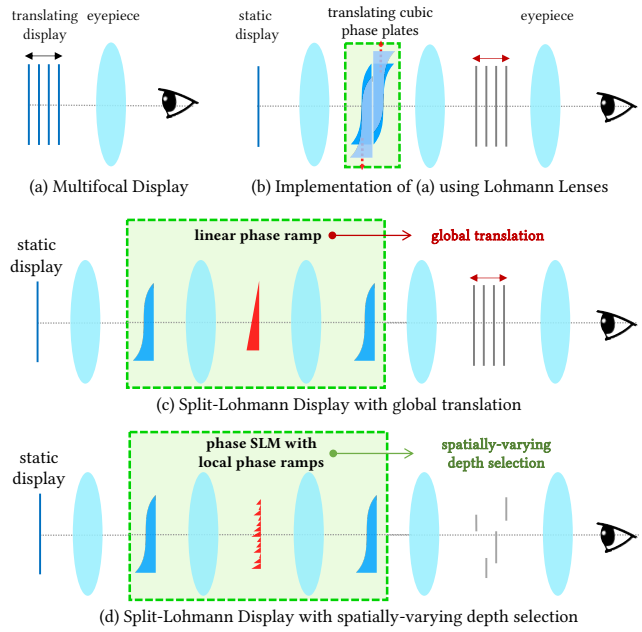


Fig. 2. **Derivation of Split-Lohmann multifocal displays.** We begin with a traditional model for a multifocal display, which involves a translating display behind an eyepiece, as shown in (a). The physical motion of display can be avoided with a 4f relay and a focus tunable lens as shown in (b). Here, we obtain the focus tunablity with a Lohmann lens; relative translation of the two cubic phase plates of the Lohmann lens leads to an optical axial translation of the display. To avoid physical motion of the display, we can split the cubic phase plates using a 4f relay and use a linear phase ramp to induce optical translations instead. This setup is shown in (c); the phase ramp needs to be introduced in the Fourier plane of the cubic plates. Putting it all together in (d), we observe that the display is 4f away from the phase ramp, which we implement using a phase SLM. The 4f system also ensures that the display pixels are now resolved on the SLM, modulo the effect of the cubic phase plate; we can enable local control on the focal length induced on a display pixel. By implementing a phase pattern with piecewise constant slope, we enable spatially-varying axial shifts on the display; we refer to this system as the Split-Lohmann multifocal display.

display at some nominal focus plane. Changing the power of the tunable lens results in an axial translation of the focus plane; it is worth noting that the entire wavefront is axially translated, in that, what was originally at the nominal plane now appears at a different plane. When the tunable lens has positive power—corresponding to a convex lens—the display shifts *away* from the eyepiece, and vice versa for negative powers. The net result is that we can achieve the setup shown in Figure 2(a) with appropriate placement of the eyepiece with respect to the range of virtual displays created by the 4f relay.

3.2 Split-Lohmann Displays

We now present a design that avoids physical motion in a Lohmann lens as well as its reliance on time-multiplexing to create content on multiple focal planes.

3.2.1 Avoiding physical motion. We avoid physical motion using the setup shown in Figure 2(c). Here, we *split* the two cubic phase plates comprising the Lohmann lens, while retaining their collocation by placing them at either end of a 4f relay. Since a 4f relay flips the fields along both axes, we use the same phase profile at both ends of the relay. In their nominal state, the two cubic phase plates cancel themselves out (as is typical to Lohmann lenses).

To introduce translation, without mechanical motion, we rely on Fourier optics. Basic Fourier analysis suggests that translation of a signal can be achieved using a linear phase shift to its frequency domain counterpart, where the slope of the linear phase shift controls the amount of translation. In our setup, this reduces to introducing a phase ramp in the Fourier plane of the cubic phase plates. Such a phase ramp can be implemented using a phase SLM. With this, we can translate the first cubic plate with respect to the second; following notation introduced in Section 2.2, the effective phase modulation is proportional to

$$(x+\Delta)^3 - x^3 = 3\Delta x^2 + 3\Delta^2 x + \Delta^3. \tag{6}$$

Comparing to Eq. (2), and ignoring constants, we can observe the quadratic phase term Δx^2 which enables the focus tunable lens, where Δ is controlled by the slope of the phase ramp on the SLM. We also observe an extra term that is linear in x; this is a consequence of translating just one of the cubic phase plates. The effects of this linear phase term is explored in more detail in Section 4.

3.2.2 Avoiding time multiplexing. We first observe that the display is optically aligned with the SLM via the 4f relay, as illustrated in Figure 2(d). If there were no cubic phase plate in between, the use of the 4f relay ensures that a display pixel is resolved onto the SLM. We can now *replace a global phase ramp on the SLM with a local one*; this implies that depending on the slope of phase ramp shown on the phase SLM locally, we can induce different choices of focal lengths of the tunable lens. In essence, we have enabled spatial selectivity in choice of power of the focus tunable lens, which allows us to avoid time multiplexing all together.

Split-Lohmann displays have many desirable properties. For example, many recent works in multifocal displays have dramatically increased the number of displayed focal planes [Chang et al. 2018; Jo et al. 2019; Rathinavel et al. 2018]; this comes at the cost of requiring a high frame-rate display and high-bandwidth electronics. In contrast, our system provides a similar number of focal planes in a single-shot manner with a standard display. As detailed in Section 4, our system has a large étendue; for the most part, we are limited by the numerical aperture of the optical elements used. Content generation for our display is computationally lightweight since the mapping from the input RGBD content to the phase SLM pattern is near-analytical; in turn, this allows us to effortlessly showcase interactive 3D games in real time.

A limiting feature of this design is that we do not obtain point-to-point mapping between the display and the SLM since the 4f relay connecting the two has a cubic phase plate in its Fourier plane. As a consequence, light from a pixel on the display is spread over multiple pixels on the SLM, denoted as the point spread function (PSF) of the cubic plate. The size and features of this PSF are determined by the parameters of the cubic plate and the focal length of relay lenses; typically, in our simulation and lab prototype, the spread of this PSF can be as large as 50 to 100 µm or tens of SLM pixels. This also provides the limits of the spatial selectivity in depth that our design provides. We discuss these in detail next.

4 DESIGN OF SPLIT-LOHMANN DISPLAYS

This section provides a wave optics-based formation model of Split-Lohmann multifocal display, which is subsequently used for the design and analysis of the system.

4.1 Image Formation in Split-Lohmann Displays

We provide an image formation model on how light propagates through the system. Much of this derivation relies on Fourier transforming properties of lenses as well as spatial filtering of 4f systems. We cover the most important details here, reserving a detailed treatment to the supplementary PDF.

Preliminaries. A 4f system consists of two lenses, which we assume to have the same focal length f_0, that are placed at a distance of f_0 and $3f_0$ from an input plane, which we denote as P_1. The plane that is $2f_0$ from P_1—or the midpoint between the two lenses—is called the Fourier or pupil plane, and is denoted as P_2; finally, the plane $4f_0$ from P_1 is the output plane and is denoted as P_3. We assume that lenses are ideal thin lenses.

The analysis in the paper requires four basic properties.

- *Fourier transforming property of lenses.* When $u_1(x_1, y_1)$ is the incident wave at P_1, the wave after propagation to P_2 is given as
$$u_2(x_2, y_2) = \frac{1}{\lambda f_0} U_1\left(\frac{x_2}{\lambda f_0}, \frac{y_2}{\lambda f_0}\right),$$
where U_1 is the Fourier transform of u_1.
- *Field relay.* If there is no optical element at P_2, then the phasor field at P_3, denoted as $u_3(x_3, y_3)$, is given as,
$$u_3(x_3, y_3) = u_1(-x_3, -y_3). \tag{7}$$
This makes the 4f system an ideal field relay.
- *Linear phase ramp at P_2.* When we have a prism or a linear phase ramp at P_2 of the form
$$e^{-j\frac{2\pi}{\lambda}(x_2 u_0 + y_2 v_0)},$$
then the field at P_3 is a shifted (and flipped) copy of u_1
$$u_3(x_3, y_3) = u_1(-x_3 + u_0 f_0, -y_3 + v_0 f_0). \tag{8}$$
For analytical tractability, we ignore the effects of any limiting apertures here and in the rest of the analysis.
- *A lens at P_2.* Placing a lens of focal length f_t at P_2 results in an axial translation of the field at P_3. Specifically, the wave $u_1(-x_3, -y_3)$, which previously appeared at P_3 or a distance of $4f_0$ from P_1, axially translates to
$$4f_0 + \delta_z = 4f_0 - f_0^2/f_t. \tag{9}$$

We now provide a detailed breakdown of wavefront propagation and image formation in Split-Lohmann displays. This analysis will also touch upon the different sources of non-idealities in the design and provide a mechanism to reason about their effects. Specifically, we derive wavefronts formed at different locations in the system—labeled as the planes $P_1, \ldots, P_5$—as light propagates from the display

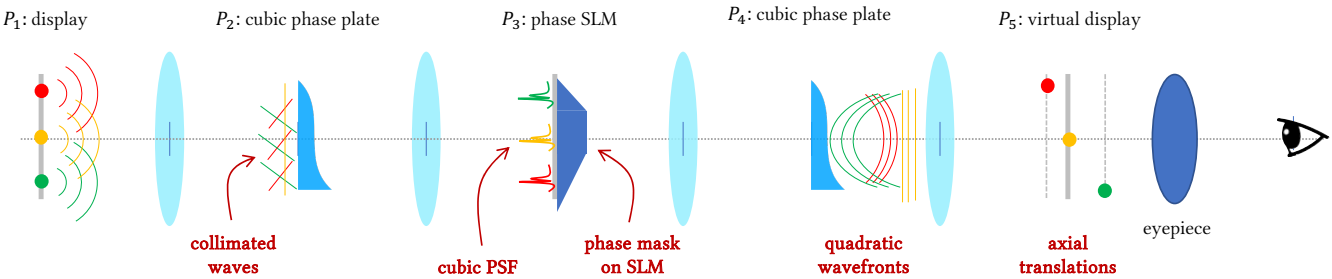


Fig. 3. **Analysis of Split-Lohmann multifocal displays.** The figure illustrates the propagation of waves from three distinct points, color-coded for visualization. All points start at the display as ideal point light sources at P_1, incoherent with each other. A phase plate at P_2 induces a cubic phase on to them. At the SLM placed at P_3, the phase gradient observed by each is different, i.e., the choice of v_0 in Eq. (13) is different: zero for the yellow point, a small positive value for the red, and a larger negative value for the green point. This results in different amounts of shifts between the interfering cubic wavefronts at P_4; the resulting waves after the second cubic plate is planar, converging quadratic and diverging quadratic for yellow, red, and green points, respectively. As a consequence, each point undergoes a different amount of axial shift at the P_5, where the translating virtual display is formed.

and to the eye; each plane is $2f_0$ from its preceding one. A visual overview of this analysis in provided in the Figure 3.

P_1: Display. We assume a display with spatially incoherent pixels, emitting narrow band illumination centered around a wavelength λ Spatial incoherence allows us to analyze the pixels in isolation, since their wavefronts do not interfere. In terms of implementation, an OLED display as well as Liquid Crystal Display with a large area source as a backlight provides us with a good approximation to this model. We now consider a point at a location (x_0, y_0) on the display, and track its wavefront through the system. To help visualize the wavefronts, a few such points are color highlighted in Figure 3.

P_2: The first cubic phase plate. The first lens collimates the light from a point on P_1, to form a *plane wave at* P_2; the wavefront at P_2, prior to the cubic phase plate is given as

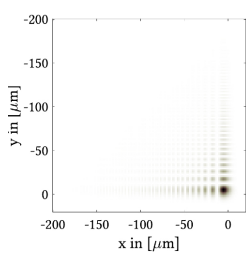
$$u_{2,-}(x_2, y_2) = \frac{1}{\lambda f_0} e^{-j\frac{2\pi}{\lambda}\frac{x_2 x_0 + y_2 y_0}{f_0}} \tag{10}$$

The cubic phase plate changes this wavefront to

$$u_{2,+}(x_2, y_2) = u_{2,-}(x_2, y_2) e^{-j\frac{2\pi}{\lambda}\frac{x_2^3 + y_2^3}{c_0}} \tag{11}$$

Here, c_0 determines the curvature of the phase plate.

P_3: Phase SLM. The wavefront at P_3 is estimated by noting that P_3 is 4f away from P_1, with a cubic phase function at its pupil/Fourier plane at P_2. If we denote $h(\cdot, \cdot)$ as the Fourier transform of the cubic phase plate, then the wavefront prior to the SLM placed at P_3 is given as



$$u_{3,-}(x_3, y_3) = h(x_3 + x_0, y_3 + y_0). \tag{12}$$

That is, we observe a copy of $h(x, y)$ at the location $(-x_0, -y_0)$. This wavefront is phase modulated by the SLM. The magnitude squared of h, which is the PSF of the cubic phase plate, is shown in the inset above for the parameters of our prototype.

As introduced in Section 3, Split-Lohmann displays deploy a phase function on the SLM at P_3 that is piecewise linear. However, for tractability in analysis, we begin with a single phase ramp at P_3. This assumption is accurate for points which lie entirely within a linear phase ramp, delaying a more thorough analysis for piecewise linear phase functions for Section 4.5. With this, the wave after the SLM is given as

$$u_{3,+}(x_3, y_3) = u_{3,-}(x_3, y_3) e^{-j\frac{2\pi}{\lambda}(x_3 + y_3) v_0}. \tag{13}$$

Here, v_0 is the slope parameter that controls the amount of shift we induce on the wavefront at P_4. As discussed in Section 2.2, Lohmann lenses require the same amount of shift along both axes and hence, we use the same slope parameter v_0 for both x and y. The specific choice of v_0 is determined by the depth map we seek to show; we will discuss this in more detail later.

P_4: The second cubic phase plate. Under the assumptions made on the wavefront at P_3 being concentrated into a single linear piece of the SLM, the optical path from $u_{2,+}$ to $u_{4,-}$ is one of a 4f system with a linear phase ramp in its Fourier plane; hence, the wavefront prior to the second cubic phase plate is a translated copy of the one at P_2. Denoting $\Delta = v_0 f_0$, from Eq. (8), we can write this as

$$u_{4,-}(x_4, y_4) = u_{2,-}(-x_4 + \Delta, -y_4 + \Delta) e^{-j\frac{2\pi}{\lambda}\frac{(-x_4+\Delta)^3 + (-y_4+\Delta)^3}{c_0}}. \tag{14}$$

With some simplifications, and ignoring constants, the wavefront after the second cubic phase plate $u_{4,+}(x_4, y_4)$ can be written as

$$\underbrace{u_{2,-}(-x_4 + \Delta, -y_4 + \Delta)}_{\text{(plane wave from display)}} \underbrace{e^{-j\frac{2\pi}{\lambda}\frac{3\Delta(x_4^2 + y_4^2)}{c_0}}}_{\text{(quadratic phase term)}} \underbrace{e^{j\frac{2\pi}{\lambda}\frac{3\Delta^2(x_4 + y_4)}{c_0}}}_{\text{(linear phase term)}}. \tag{15}$$

Since $u_{2,-}$ corresponds to a planar wavefront, its translation only changes it by a constant/scalar.

In simpler words, the wavefront after the second cubic phase plate is the same as the one *before* the first cubic plate, multiplied by two phase terms: a quadratic term corresponding to a lens with focal length $f_t = c_0/(6\Delta) = c_0/(6v_0 f_0)$, and a linear term. As we see next, the quadratic term results in a desired axial shift of the point and the linear term results in an unwanted lateral shift.

P_5: Virtual display plane. If we ignore the quadratic and linear phase terms in Eq. (15), then the wavefront at P_5 is the same as the one in P_1—a delta function centered at (x_0, y_0). The quadratic phase term induces an axial shift $\delta_z = -f_0^2/f_t$, from Eq. (9), and so this delta function appears at a location of

$$\delta_z = -\frac{6v_0 f_0^3}{c_0}, \tag{16}$$

with respect to P_5 plane, with positive values indicating displacements towards the eyepiece. Since we control v_0 via the phase patterns shown on the SLM, we can place a display pixel at a specific axial distance from the eyepiece, which in turn allows the eye to perceive it as arising from a desired depth.

The linear phase term provides a lateral shift so that the delta function appears at

$$\left(x_0 - \frac{3v_0^2 f_0^3}{c_0}, y_0 - \frac{3v_0^2 f_0^3}{c_0}\right). \tag{17}$$

This lateral shift is undesired, but is often very small due to its dependence on v_0^2; we discuss it in more detail later in this section.

4.2 Determining System Parameters

The Split-Lohmann multifocal display described thus far and illustrated in Figures 2 and 3 have a number of free parameters, which include: f_e, the focal length of the eyepiece, f_0, the focal length of the relay lenses, and c_0 the parameter determining the curvature of the cubic phase plate. Each of these parameters controls salient attributes of the display, such as its field of view, eyebox, and its depth range, and depth resolution.

Eye relief. Our system assumes that the eye is at a distance f_e from the eyepiece. This placement ensures that the axial shift of a display pixel does not change its viewing direction; the associated ray diagram is provided in Figure 4. This allows us to decouple depth placement and viewing direction, which is extremely helpful when porting content onto the device. This positioning of the eye has another advantage. If the eye were in a different position, the coupling between viewing direction and depth placement could result in requiring the same display pixel, under different axial shifts, for multiple scene points. This would significantly complicate our ability to faithfully reproduce a 3D scene on our system.

Eyepiece focal length f_e. A primary criteria in the selection of the eyepiece is the field of view of the display. Suppose that ℓ_{disp} is the smaller of the display and SLM width/height, the field of view of the multifocal display is given as

$$\text{field of view } = 2\tan^{-1}\left(\frac{\ell_{\text{disp}}}{2f_e}\right). \tag{18}$$

All things considered a smaller focal length is desirable; however, the increase in field of view comes with a commensurate reduction in the eyebox of the display since refractive optics preserve the étendue of the system.

Another factor that needs to be kept in mind is the working range of the display in terms of the depth it can produce. Suppose we seek to have a working range of W diopters (D), say from $[0, W]$ D. The focal length of the eyepiece allows us to characterize the amount of

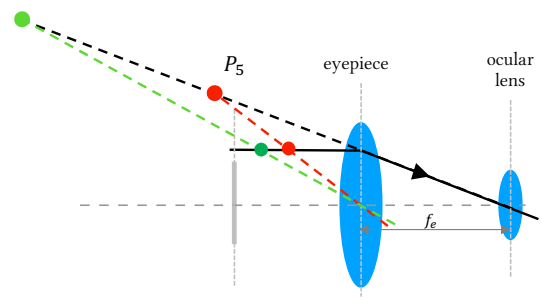


Fig. 4. **Position of the eye.** We require that the eye be placed at the focal plane of the eyepiece, i.e, a distance of f_e from the eyepiece. This placement is important as it ensures that an axially translating point on the display remains at the same viewing angle at the eye.

axial displacement of the virtual displays. Using basic ray tracing, we can show that the total axial displacement Z that we need in the plane P_5 is given as

$$Z = W f_e^2. \tag{19}$$

The total axial displacement is usually in the order of a few millimeters; for example, our prototype is built with $f_e = 40$ mm and for a working range of $W = 4$ D which results in $Z = 40^2 \times 4/1000 = 6.4$ mm. A smaller eyepiece focal length results in requiring a smaller range of axial displacement of the virtual displays, which is beneficial; however, this invariably comes at the cost of having fewer focal planes over the working range, due to inherent limitations in axial resolution of virtual displays. It is worth noting though that a number of prior works have tried to characterize the number of focal planes needed in multifocal displays; this number can be as low as five focal planes—when the evaluation metric is a satisfactory level of accommodation [Akeley 2004]—to as large as 40—when the evaluation metric is visual acuity [Chang et al. 2018; Rolland et al. 2000]. Our implementation aims for 50 focal planes for reasons we enumerate shortly.

Curvature of the cubic phase plate c_0 and tilt range of the SLM. The curvature of the cubic plates plays a direct role in the axial displacement as seen in Eq. (16). A higher curvature, corresponding to a smaller c_0, allows for a smaller value of SLM tilt v_0 to produce the necessary axial translation; however, this results in a larger PSF on P_3 that reduces the spatial resolution of depth selection.

To understand these dueling concerns better, we need to first characterize the limits on v_0, the slope of the phase modulation performed by the SLM. Given an SLM with a pixel pitch of δ_{slm}, the phase gradient cannot exceed π/δ_{slm} in order to avoid aliasing effects. This constrains

$$|v_0| \leq \frac{\lambda}{2\sqrt{2}\delta_{\text{slm}}} = v_{\max}. \tag{20}$$

Combining this with Eq. (16), we can bound the working range Z as

$$Z \leq \frac{12 v_{\max} f_0^3}{c_0}, \tag{21}$$

and using Eq. (19), this gives us

$$c_0 \leq \frac{12 v_{\max} f_0^3}{W f_e^2}. \tag{22}$$

Generally, large values of c_0 is preferred as the resulting surface with its smaller curvature is more easily fabricated.

The value of c_0 is also crucial in determining the spread of the wavefront on the SLM, placed at P_3, for light arising from a single pixel on the display at P_1. As derived in Eq. (12), this wavefront is shaped by the Fourier transform of the cubic phase function. The scaling property of Fourier pairs can be used to show that doubling c_0 reduces the size of the PSF h by a factor of two in each dimension. Thus using the largest value of c_0 will produce a wavefront at the SLM whose energy is most concentrated, spatially. Reducing the spatial spread of this wavefront in turn reduces the amount of artifacts generated near depth discontinuities and characterizes the spatial resolution at which we can obtain depth selectivity.

Selecting relay lens focal length f_0. Given its central role of the relay lenses in the overall design, the choice of their focal length f_0 affects many aspects of the display.

- *Size.* By its construction, the length of the light path scales linearly with f_0; this suggests that smaller values of f_0 are preferred.
- *Diameter.* All of the analysis of wavefront propagation presented so far are with ideal lenses, which assumes little of the wavefront is clipped by the relay lenses; in turn, this requires that the relay lenses have sufficiently large diameters. Practically, a rule of thumb that arises from ray tracing is that the diameter of relay lenses should be greater than the sum of the image and Fourier plane aperture widths. Since the focal length of a spherical lens cannot be smaller than its diameter, this forces us to avoid lenses with small f_0.
- *The cubic phase plate.* Since the upper bound on the c_0 derived in Eq. (22) is *cubic* in f_0, decreasing f_0 forces us to reduce c_0 rather significantly, which has the negative consequences discussed above, suggesting the use of larger f_0.

Given these constraints—further amplified by the lack of quantitative metrics—the choice of f_0 is perhaps the trickiest.

Placement of the eyepiece. The placement of the eyepiece with respect to P_5 determines the working range of the display. From Eq. (16), we observe that pixels on the display can be virtually translated, with respect to P_5, to axial locations in the range

$$\left[-\frac{6 v_{\max} f_0^3}{c_0}, +\frac{6 v_{\max} f_0^3}{c_0} \right].$$

To use the entire working range, the eyepiece should be placed such that this entire range lies closer the lens than its focal plane; further, aligning the furthest axial shift with the focal plane of the eyepiece will place that plane at infinity to the eye. As seen in Figure 5, this requires the eyepiece to be placed at a distance z_e from P_5 such that

$$z_e + \frac{6 v_{\max} f_0^3}{c_0} = f_e. \tag{23}$$

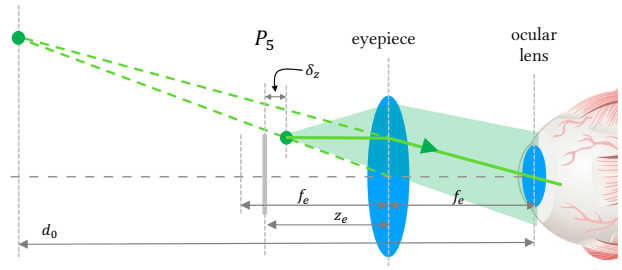


Fig. 5. **Placement of 3D content.** To place a point at a depth d_0 from the eye, we need to induce an axial shift to the display pixel corresponding to that point by an amount δ_z from P_5. Using basic ray tracing, we can derive an expression relating d_0 to δ_z; see Eq. (25). Using Eq. (26), we can compute the phase gradient needed on the SLM to achieve this axial shift.

4.3 Simulating the Display

We use a wave optics model to simulate Split-Lohmann displays, or specifically, to propagate wavefronts from P_1 all the way to the eye. Since P_1 to P_5 form consecutive Fourier planes, we can use the Fast Fourier Transform to propagate the wavefront with careful discretization of the wavefronts at each plane. To propagate from P_5 to the eye, we can use the Fourier transforming property again; but since P_5 is closer to the eyepiece than its focal plane, we need to post-multiply the propagated wavefront with a quadratic phase term (see Goodman [2005]; Chapter 4). The lens in the eye is allowed to have different focal lengths depending on its accommodation state; we assume that the retina is planar and is at a distance 25 mm from the ocular lens. We use Fresnel propagation for this last propagation. In this propagation, all lenses—except the one in the eye—are assumed to be ideal and thin, with infinite diameter; the lens in the eye is assumed to have a diameter of 5 mm. The cubic phase plates are assumed to be thin. We refer interested readers to Voelz [2011] for details on the numerical implementation of wave propagation.

The wave propagation described above assumes temporal and spatial coherence of the underlying wave. While temporal coherence is not a concern, our display requires spatial incoherence of the light emitted at the display. To incorporate this incoherence, we propagate the wave multiple times, with a random initial phase added to the wave at P_1 in each trial; the image at the eye is computed as the average *intensity* of the wave across the trials. This process of adding random phase and averaging the resulting intensity field simulates the physical process underlying spatial incoherence, namely, a random perturbation across points that are not spatially coherent. Our simulation codebase can be found on the project website [Qin et al. 2023].

Separate from the wave model, we also built a ray tracing-based simulator of our device in Blender. This simulator uses realistic optics—in fact, the exact set of lenses used in the prototype—and captures the effects of the thickness of various optical components including the cubic phase plate; however, phase SLMs, which are hard to model using ray optics, is implemented as a micro-mirror array. The supplemental material has this Blender model, pre-configured with a scene, for the interested reader.

4.4 Content Fitting

To display a 3D scene, our system takes in as input an intensity image $t(\alpha, \beta)$ and a depth image $d(\alpha, \beta)$, where α and β are in the tangent of the angle subtended at the eye.[1] Therefore, the intensity $t(\alpha_0, \beta_0)$ at a pixel (α_0, β_0) corresponds to a 3D scene point $d(\alpha_0, \beta_0)[\alpha_0, \beta_0, 1]^\top$.

Figure 5 provides a detailed view on the placement of the eyepiece and eye with respect to the plane P_5, and ray diagrams on how to compute the virtual display location to obtain a specific depth. The eyepiece is placed at a distance z_e from P_5, as given in Eq. (23). As noted in Figure 4, the eye is placed at a distance f_e from the eyepiece which ensures that the axial shift of a display pixel does not change its viewing direction. Hence, we need to ensure that the intensity $t(\alpha_0, \beta_0)$ is displayed at a location $(f_e\alpha_0, f_e\beta_0)$ on the display at P_1; note that this calculation ignores the effect of the unwanted shift which, as we will show shortly, can be made extremely small.

To place the same point at a depth $d_0 = d(\alpha_0, \beta_0)$ from the eye, we need to induce a displacement δ_z from the nominal plane P_5 such that

$$\frac{1}{z_e - \delta_z} - \frac{1}{d_0 - f_e} = \frac{1}{f_e} \tag{24}$$

This expression can be derived from the ray diagram shown in Figure 5, and simplifies to

$$\delta_z = z_e - \frac{(d_0 - f_e)f_e}{d_0} \tag{25}$$

To induce this axial shift, we can use Eq. (16) to calculate that the SLM slope parameter v_0 corresponding to this pixel:

$$v_0 = -\frac{\delta_z c_0}{6 f_0^3} \tag{26}$$

Consequently, accounting for the xy flip on P_3, we need to ensure that the pattern on the SLM has a slope of v_0 as in Eq. (26) at the location $(-f_e\alpha_0, -f_e\beta_0)$.

Need for quantizing the depth map. Given the input depth map $d(\alpha, \beta)$, it is tempting to calculate an axial shift map $\delta_z(\alpha, \beta)$ by replacing d_0 in Eq. (25) with $d(\alpha, \beta)$, and subsequently a phase SLM slope map $v_0(\alpha, \beta)$ using Eq. (25). We can now create a phase image $\phi(x_3, y_3)$ for the SLM, using Eq. (13),

$$\phi(x_3, y_3) = \frac{2\pi}{\lambda}(x_3 + y_3)\, v_0\left(-\frac{x_3}{f_e}, -\frac{y_3}{f_e}\right).$$

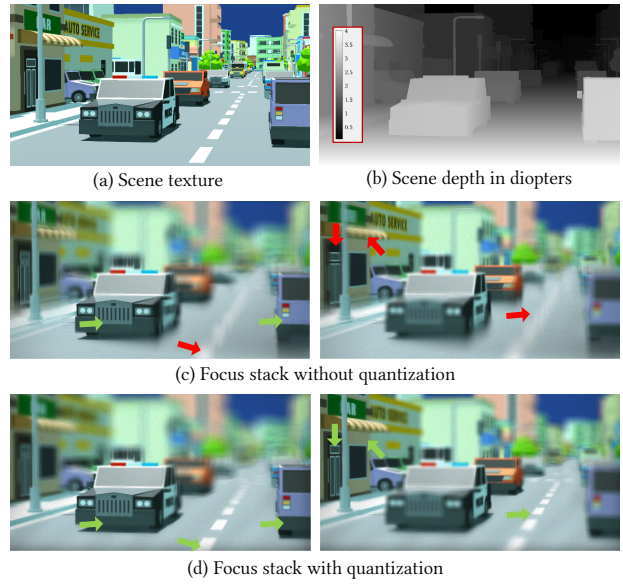
This phase pattern, however, will not work as its gradient is not the desired one; specifically,

$$\frac{\partial \phi}{\partial x_3} = \frac{2\pi}{\lambda} v_0 - \frac{2\pi}{\lambda}(x_3 + y_3)\frac{1}{f_e}\frac{\partial v_0}{\partial x_3}$$

The second term can produce grossly incorrect phase gradients, which leads to incorrect 2D translation of the cubic phase plates.

To fix this, we use a simple heuristic: we quantize the depth map $d(\alpha, \beta)$ in the diopter space. Given that natural scenes has depth maps that are piecewise smooth, quantizing it makes it piecewise constant. A piecewise constant image has zero-valued gradients, except at discontinuities; this simple fix ensures that the second term of the phase gradient is zero at most locations. For the results in the paper, we quantize the scene with 50 uniformly spaced values in the working range of the display, which is typically 25 cm to ∞ or $[0, 4]$ diopters ; this results in a quantization error of 0.04 diopters. Our choice of 50 quantization levels is based on prior work [Chang et al. 2018] that characterizes the number of focal planes required to maintain visual acuity when a point is in focus at the eye.

[1]Tangent of angle is the space corresponding to the locations on the image plane of a perspective camera; so this is the natural choice when describing images.



(a) Scene texture (b) Scene depth in diopters

(c) Focus stack without quantization

(d) Focus stack with quantization

Fig. 6. **Quantizing the depth map.** For the scene shown in (a, b), we present simulation results with (d) and without (c) quantizing the depth map prior to deriving the SLM phase function. The columns in (c, d) corresponds to the eye in focus at 4D (left) and 3D (right). As seen in (d), using a quantized results in realistic focus and defocus cues. In contrast, in the absence of quantization, we only observe correct focus cues for objects with locally-constant depth maps like the front of the vehicles. Even defocus cues, for example the text on the building profiles are incorrect in (c). This is a consequence of the SLM inducing unequal shifts of the cubic plates along the two axes; this results in elliptical phase delays being induced on points that makes it impossible for the eye to focus on them. The arrows overlaid on the images show interest points, color coded in green for correct and red for incorrect (de)focus cues. (Credits: "SimplePoly City" 3D scene courtesy of "VenCreations" at Unity Asset Store)

Figure 6 shows simulations results with and without the quantization of the depth map. In the absence of quantization, the eye is only able to focus on regions that have a locally-constant depth map, for example, the front of the car and truck. For parts of the scene with smoothly-varying depth, like the road and the building facade, the shift induced by the SLM is no longer equal along x and y axis; the resulting modulation is no longer quadratic and hence, the eye is unable to focus on it.

A pseudocode for this process is provided in Algorithm 1. This process is near-analytical, having a light computational footprint. As a consequence, we can easily enable interactive applications without requiring sophisticated compute platforms or even optimized implementations. Section 5 and the supplemental video showcases such interactive scenes.

ALGORITHM 1: Pseudocode for estimating the SLM phase function given scene depth map

function SLMPhaseFunction $(d(x, y))$;
Input : depth image $d(x, y)$
Output: SLM phase pattern $\phi(x_3, y_3)$
$D(x, y) \leftarrow 1/d(x, y)$ % compute diopter map %
$D(x, y) \leftarrow \text{Quantize}(D(x, y), 50)$ % Quantize in diopter space %
$d_0(x, y) \leftarrow 1/D(x, y)$ %back to depth map%
$\delta_z(x, y) \leftarrow z_e - \frac{(d_0(x,y)-f_e)f_e}{d_0(x,y)}$ % axial displacement via Eq. (25)%
$v_0(x, y) \leftarrow -\frac{\delta_z(x,y)c_0}{6f_0^3}$ %phase gradient map via Eq. (26)%
$\phi(x_3, y_3) = 2\pi v_0(-x_3, -y_3)(x_3 + y_3)$ %SLM phase pattern%

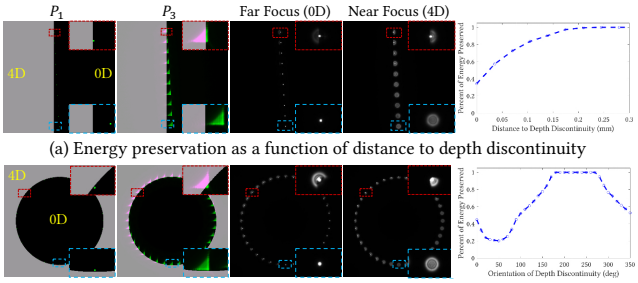
4.5 Effect of Discontinuities on Phase SLM

We address a question that is central to this work: what are the capabilities of our system in terms of the spatial resolution in depth placement? In the analysis presented so far, there were two observations made that limit depth selectivity. The first is in the assumption that most of the wavefront associated with a pixel on the display is associated with a single slope on the SLM (or specifically, not split across multiple slopes); as stated earlier, this is often not the case, due to the spreading of light by the cubic phase plate at P_2. The second is in making the depth map piecewise constant to avoid unwanted gradients when designing the phase function; however, this is not a stringent constraint as long as we have a sufficient number of focal planes in the scene [Chang et al. 2018]. We now present a detailed study of the artifacts around depth discontinuities.

Figure 7 considers, in simulation, a set of points that are at different distances from a depth edge. For the point that are sufficiently far enough from the edge, so that they are largely modulated by a single phase ramp, the conclusions of the theory holds; a nicely resolved point when the eye focuses at the chosen depth and defocus blur when out of focus. However, when the cubic PSF is split across the depth edge, we observe the following:

- *Focus at both depths.* The point comes into focus at both depths that its wavefront interacts with the relative intensity of the focused spots depending on the amount of overlap. Specifically, the energy in leakage is equal to the energy in the desired depth when a point is ~25 µm from the edge; the leaked energy is 10 times smaller when the point is is ~150 µm from the edge.
- *Distorted bokeh.* The defocus blur kernel takes on a distorted shape, likely due to the interference between the two focus spots we have now created.
- *Dependence on the orientation of the discontinuity.* The defocus blur of the cubic plate is anisotropic; as a result, the orientation of the depth discontinuity plays an important role in determining how much leakage occurs. Figure 7(b) shows an example of this. In the most favorable configuration, there is no leakage even if the point is immediately adjacent to the discontinuity.

While these artifacts can be distracting for a point light source, they are less dramatic for scenes with dense textures. Across dozens of scenes that we tested on, we found the artifacts to be innocuous and mildly distracting at best. However, in some cases where multiple copies of a texture can appear, the focusing artifact can be distracting. Figure 8 provides an example.



(a) Energy preservation as a function of distance to depth discontinuity

(b) Energy preservation as a function of orientation of depth discontinuity

Fig. 7. **Analysis of artifacts at depth discontinuities.** We analyze the effect of light leakage across depth discontinuities using scenes consisting of point light sources. (From left to right) The scene depth map with the location of points overlaid on top, with the black portion corresponding to a depth of 0 D and gray corresponding to 4 D; the image observed on P_3 overlaid over the depth map with pink color denoting light that has leaked across depth discontinuity; image of the scene when the eye focuses on 0 D; image for focus at 4 D. In (a), we look at how distance to the discontinuity affects leakage, and (b) looks at the effects of the orientation of the depth discontinuity. Where there is significant leakage across the discontinuity, we observe the following: (i) points start appearing to be in focus at both depths when there is significant leakage; (ii) the defocus blur is no longer circular; and (iii) the effects of leakage are highly anisotropic. The plots on the right shows the fraction of light that is not leaked as a function of distance (a) and orientation (b). We observe that less than 50% of the light leaks when a point is ~ 25 µm away from a discontinuity.

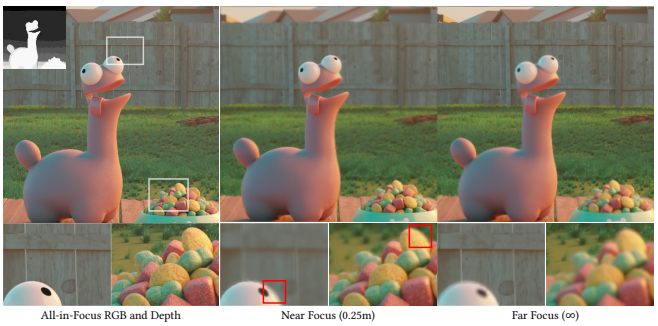


Fig. 8. **Artifacts at large depth discontinuities.** We present simulation results of a scene with a large depth discontinuity near the face of the toy. The red boxes indicate regions of large depth discontinuities where artifacts are visible. (Credits: "The Daily Dweebs" 3D scene courtesy of Blender Studio and Hjalti Hjalmarsson)

4.6 Effect of the Undesired Lateral Shift

Split-Lohmann displays also induce an unwanted lateral shift, as previously discussed and characterized in Eq. (17). It is worth getting a sense of the scale of the lateral shift before we discuss potential approaches to reduce it. For example, for the GAEA SLM, the pixel pitch is $\delta_{\text{SLM}} = 3.74\,\mu\text{m}$, and so $v_{\max}$ in Eq. (20) is equal to 0.05 for $\lambda = 532\,\text{nm}$. Assuming focal lengths $f_0 = 100\,\text{mm}$ and $f_e = 40\,\text{mm}$,

a working range of $W = 4$ diopters results in $c_0 = 0.093\,75\,\mathrm{m}^2$ from Eq. (22). With this, we can calculate the maximum lateral shift using Eq. (17) to be 80 µm or about 10 pixels of the display. Further, the quadratic dependence of the lateral shift on v_0 also means that most depths will not see such a large shift; that is, while content at 0D and 4D, which are obtained using the largest slope on the SLM, so observe such a large shift, content between 1D and 3D for example would only see a shift smaller than 20 µm or 2.5 pixels.

It is possible to reduce the lateral shift further by sacrificing the number of depth planes we can select. Since depth selectivity comes from the slope on the SLM, an 8-bit SLM provides up to 256 unique choices in the SLM slope. This is overkill, as we seldom need more than 40 or so focal planes to maintain sharpness at focus as well as satisfying accommodation cues. So, we can use a cubic phase plate with a smaller c_0 which increases the working range; for example, choosing half the value of c_0 to $0.047\,\mathrm{m}^2$ doubles the working range W to 8 D. Since we only need the central 4D of this working range, we can choose to only use half the tilting power of the SLM, which reduces the maximum lateral shift by a factor of two once we account for the smaller c_0, and so the lateral shift becomes 40 µm in each axes. This does come at a cost of a broader blur from the cubic phase plate, reducing spatial selectivity of depth. The results in the paper, both simulation and real, were obtained with $c_0 = 0.0193\ m^2$; this reduces the maximum shift to $\sim$ 10 µm which is imperceptible to the human eye and the number of depth planes to 52.

4.7 Attributes of the Split-Lohmann Display

We now briefly discuss the properties of Split-Lohmann displays when it comes to metrics used to characterize near-eye displays.

Field of view and eyebox. Field of view is determined by the OLED width and the eyepiece focal length as described in Eq. 18. For our lab prototype, this comes to a diagonal field of view of 25°.

Like other multifocal displays, the size of the eyebox is largely determined by the numerical aperture of lenses used and any limiting apertures; due to the use of 4f relays, it is the latter that limits our prototype. The limiting aperture is reduced slightly due to the shift induced by the SLM, and this shift is depth dependent; points that are axially resolved at P_5 do not see any reduction and the eyebox gradually reduces as we shift away from P_5. Ignoring the effects of this shift, which we expect to be small as we do not use the SLM's full bending power, we can calculate the size of the eyebox as

$$\text{diameter of 4f aperture} \times \frac{f_e}{f_0},$$

where the factor f_e/f_0 reflects the scaling induced by the eyepiece. For our lab prototype, this comes to 5 mm ($f_e = 40$ mm, $f_0 = 100$ mm, aperture diameter = 12.5 mm).

It is worth noting that field of view can be easily increased by using an eyepiece of a smaller focal length. However, due to étendue invariance of refractive optics, this expansion of the field of view will lead to a commensurate reduction in the eyebox. We can counteract this by using higher numerical aperture optics in the 4f relay; for example, using 50 mm or 30 mm lenses, which would double or triple the eyebox, respectively, along each dimension.

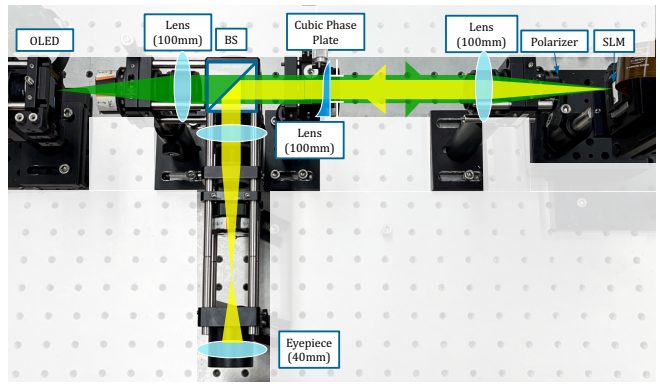


Fig. 9. **Schematic of our lab prototype.** See Figure S4 in supplemental material for a more detailed schematic with a list of components.

Working range and depth resolution. Working range of the display is one of the primary design targets, and our display can easily achieve the typical 4 D of range that multifocal displays are normally configured for. Generally, a large working range would require a large axial motion or a smaller eyepiece focal length. From Eq. 21, we see that a larger working range can be achieved with a more powerful SLM (larger $v_{\max}$), larger f_0 or a more powerful cubic phase plate (smaller c_0). Hence, in principle, we can achieve the working range of 20 D that is associated with children [Chen et al. 2000]; in fact, our prototype is capable of this if we use the entire tilting range of the SLM.

Depth resolution is controlled by the precision at which we can implement phase ramps on the SLM. As a coarse approximation, this is bounded by the number of quantization levels available in addressing the SLM, which is typically 8 bits; this sets a maximum of 255 depth planes that we can display. Most applications do not need such a large selection; however, as mentioned in Section 4.6, we can trade-off depth resolution to reduce lateral shifts (and at the cost of increased leakage from the cubic phase plate).

5 EVALUATION

We present details of the lab implementation of a Split-Lohmann display, and present results detailing the performance of the system on a variety of scenes as well as comparisons to related prior work.

5.1 Hardware Implementation

We experimentally evaluate Split-Lohmann Display by building a proof-of-concept hardware prototype whose schematic is shown in Figure 9. A detailed component list can be found in the Supplemental Section S2.

Hardware details. The OLED used is a 1.03"-diagonal Micro-OLED having a 2560 × 2560 RGB resolution with a pixel pitch of $\delta_{\mathrm{OLED}} = 7$ µm. The SLM we use is a Holoeye GAEA-2, which is a 0.7"-diagonal phase-only reflective SLM having a 2464 × 4000 resolution with a pixel pitch of $\delta_{\mathrm{SLM}} = 3.74$ µm. The smaller area of the SLM makes it the limiting factor in terms of displaying content and hence, the field of view. We use a 40 mm eyepiece, which provides a diagonal field of view of approximately 25°.

In terms of optics, the relay lenses were achromatic doublets with focal length $f_0 = 100$ mm. The cubic phase plate was a custom component, fabricated using subtractive manufacturing, i.e., by using laser to sculpt the desired shape. It was fabricated over a diameter of 12 mm with a thickness parameter of $c_0 = 0.0193\,\mathrm{m}^2$ using a material with a refractive index of 1.46. Under these settings, the maximum sag in the phase plate was 50 µm, which was also the limit for the manufacturing process.

Working range and depth quantization. The maximum tilt implementable with our SLM allows a maximum shift of 6.952 mm between the cubic phase plates; using Eqs. (19), (20), and (21), we estimate a maximum working range of $W = 27.535$ diopters. However, as discussed in Section 4.6, we use only a working range of 4 diopters, which allows us to reduce the maximum unwanted shift to ~ 10.293 µm. This shift is imperceptible given the OLED pixel pitch of 7 µm. Finally, as suggested in Algorithm 1, all results were obtained with the depth map, measured in diopters, being quantized uniformly with 50 bins.

Handling color. Using a spectrometer, we measured the OLED color channels to have spectral peaks at 619 nm, 530 nm, and 461 nm for red, green, and blue channels, respectively; the full-width half max of the respective channels were 28 nm, 31 nm, and 17 nm. With the exception of color sequencing results, all results were captured with the SLM's gamma curve optimized for the green channel; this choice was motivated by the predominance of green in natural imagery; Under this operation that we refer to as "field simultaneous color", all three color channels were simultaneously illuminated at the OLED with a single SLM phase mask optimized for the peak green wavelength. A more precise evaluation of color handling is presented later in this section.

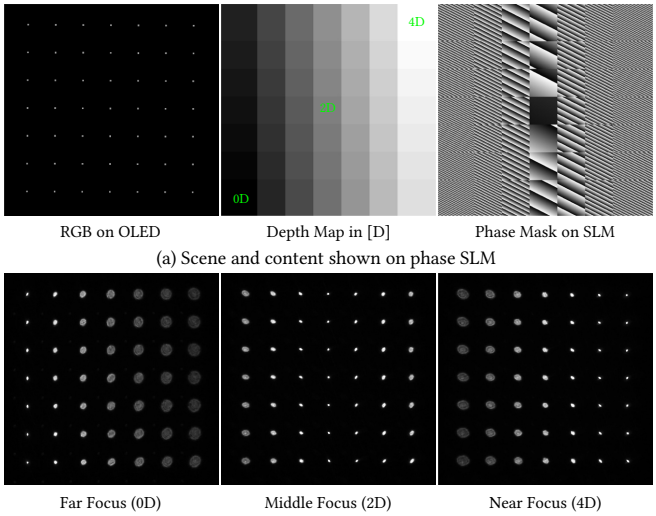
Capturing results with an observer camera. The results presented here were acquired using a Nikon Z5 Camera with a 18 - 55 mm lens operated at focal length of 55 mm and aperture of $F/5.6$. For focus stack results, we used an ISO of 100 and remotely controlled the focus setting of the lens, typically with a exposure time of 1 to 3 sec. For the video results in the supplemental material, we used the video mode of the camera, typically with ISO ranging around 3200, at a frame rate of 24 fps. In all of these, the camera was completely untethered to the display without any synchronization.

5.2 Static VR scenes

We first present results on static virtual reality (VR) scenes.

Bokeh. We experimentally demonstrate the bokeh of white dots at different depths in Figure 10. We use a 3D scene of 49 dots at 49 depths, uniformly sampled in diopters over the $[0, 4]$ D range. We can observe that (a) the blur kernels of an out-of-focus dot remains naturally circular while the in-focus dot remains sharp, and (b) this the size of the circular blur grows as the content depth is further away from the focused distance.

Spatial and depth resolution. To characterize the spatial and depth resolution of our prototype, we display a resolution chart on the device and report the results in Figure 11. We fix the camera focus to 0 D, and place content at different depths, characterizing the defocus in resulting images using their modulation transfer function (MTF). We can clearly observes difference when the scene is placed at 0.125 D, and the MTFs are different even at a spacing of 0.0625 D. Given the working range of our system is 4 D, this suggests that our display is capable of displaying at least 32 to 64 focal planes. The MTF when the content is at 0 D characterizes the spatial resolution of the device, since the camera is at focus on the content. We observe a value of MTF30 at a spatial frequency of 35 lp/mm on the OLED display plane, corresponding to a spatial period of 4 pixels, which is only 2× worse than the display's native resolution.

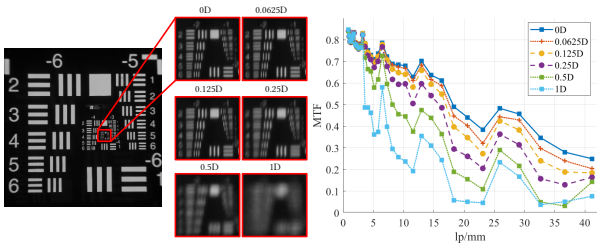


(a) Scene and content shown on phase SLM

(b) Images captured on lab prototype

Fig. 10. **Bokeh.** To characterize the focus and defocus blurs produced by the display, we use a scene shown in (a), consisting of a sparse 7×7 grid of 4×4 dots. The depth map associated with the scene is piecewise constant spanning a range from 0 D to 4 D. The resulting phase mask that we use on the phase SLM is shown on the right. In (b), we show images acquired with an observer camera at three different focus settings: the furthest (0D), the middle (2D), and the closest (4D). We observe that the in-focus spot is sharp while the defocus of other dots increases linearly with the difference in depth (measured in diopters). The bokehs observed are largely circular; the deviation from circular shape can be attributed to the shifts induced by the SLM as well as misalignments in system assembly. This result was captured in green color channel, and visualized as grayscale.

Comparison to time-multiplexed multifocal display. As we discussed in Section 4, one of the primary sources of aberrations in Split-Lohmann displays comes from leakage of light across depth discontinuities. To isolate the artifacts stemming from this, we compare against a time-multiplexed counterpart that is also implemented on the same hardware. Specifically, for the time multiplexing result, we divided the scene into 50 disjoint images, each associated with a different depth. We capture 50 images (for each focus setting of the observer camera), sweeping through the images corresponding to each focal plane; for each of these images, the SLM displays a global phase ramp with a constant gradient to place content at the desired focal plane. The captured images were summed up digitally



(a) Images of a res. chart placed at different depths (b) MTF plot

Fig. 11. **Characterizing depth resolution.** We display a USAF target shown at different depths, starting at 1 D and approaching 0 D, at smaller increments. The camera remains focused at 0 D, and insets of the images acquired with changing content depth is shown in (a). As we expect, the defocus blur reduces as we approach 0 D; yet, it is still distinguishable even when the content is at 0.125 D; this is also reflected in the MTF plots shown in (b). This suggests that our system can resolve focal planes separated by 0.125 D, which translates to 32 focal planes over a 4 D working range. The spatial frequencies in the MTF plots correspond to the OLED display plane; here, a frequency of 35 lp/mm, where we achieve MTF30 when focused on the content at 0D, corresponds to a spatial period of 4 pixels, which is only 2× worse than the display's native resolution.

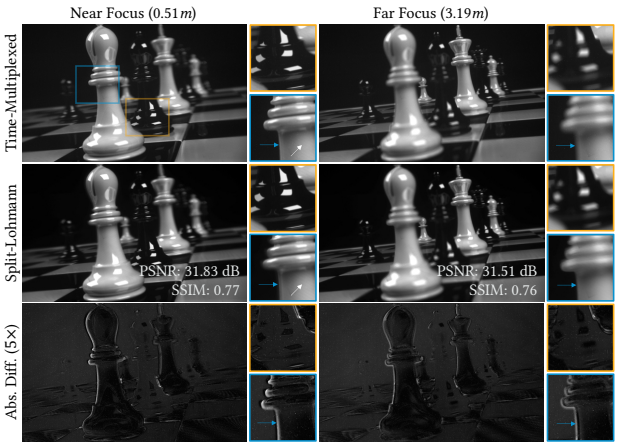


Fig. 12. **Comparison to time-multiplexed multifocal display.** To isolate the light leakage artifacts in Split-Lohmann, we compare against its time-multiplexed counterpart, implemented on the same hardware. The time-multiplexed result was obtained by sequencing through the content, one focal plane at a time by using a global phase ramp on the SLM; the acquired photographs are digitally summed. This process is repeated for each focus setting of the camera. The long exposures at high ISOs needed for these photographs results in a number of hot pixels. Split-Lohmann results have artifacts at depth discontinuities that we highlight in the insets with blue arrows and are easily seen in the absolute difference images (intensified 3× for visualization). The white arrows show the SLM's phase warping artifacts common to both methods. PSNR and SSIM values between the two systems are overlaid on the second row. Note that these values are affected by measurement noise in the acquired photographs. This result was captured in the green channel and visualized as grayscale. (Credits: 3D assets courtesy of "3DinHD" at TurboSquid)

to mimic the persistence of vision, and visualized in Figure 12 for two different focus settings of the observer camera.

Figure 12 suggests little difference in the overall quality of images between Split-Lohmann and its time-multiplexed counterparts. A closer look does reveal some minimal ghosting artifacts near depth discontinuities. This is consistent with our simulation results in Figure 8. Since both Split-Lohmann and time-multiplexed results were obtained from the same hardware, this result isolates the artifacts that stem purely from the single-shot implementation; in other words, it compares the setups shown in Figure 2(c) and (d).

Field sequential vs. simultaneous color. The use of phase SLMs invariably results in chromatic aberrations when we work with multi-color inputs. We evaluate the performance of two color modes: *field sequential color*, where we time-multiplex across the three color channels, using the correct SLM gamma curve for each color to ensure maximal fidelity of phase modulation; and *field simultaneous color*, where we show all three color channels simultaneously on the OLED, with the SLM configured to an optimized gamma curve for green color. For field sequential color, changing the gamma curve on the SLM is sufficient as it changes the path length retardation so that the 0 - 255 input to the SLM maps to a 0 - 2π phase delay for the calibrated wavelength; this however requires displaying three images within persistence of vision. For field simultaneous color, the SLM is optimized only for the green wavelength; the other two wavelengths observe a phase ramp that is incorrect. Specifically, a phase gradient of v_0 designed for the green wavelength λ_g appears scaled as $v_0\lambda_g/\lambda$ when illuminated with light of wavelength λ. Hence, we expect red and blue wavelengths to focus respectively at a slightly further and closer depth, as compared to the desired depth, while the green channel would correctly focus.

Figures 13 and 14 look at differences between the two color modes, with the former using a high contrast scene and the latter using a natural scene. From Figure 13, we observe that each color is best resolved under its own gamma curve, with a peak value that is maximized for the correct retardation curve. The differences in peak value are not as bad as we would expect given that $\lambda_g/\lambda_r = 0.86$ and $\lambda_g/\lambda_b = 1.15$; this is likely a cause of the misalignments in the optical setup that likely reduces overall performance. From Figure 14, both qualitatively and quantitatively, we observe high fidelity between images from the two color modes.

The similarity in results between the two color modes leads us to choose field simultaneous color as our operating mode of choice for the rest of the results in this paper. In particular, it avoids the need for high frame-rate control of the OLED and SLM, which would require development of a customized control board for both units. This also enables us to showcase interactive VR scenes with little additional implementation overhead.

Parallax. Virtual scenes shown in multifocal displays do satisfy parallax cues, at least within the eyebox. Figure 15 and the supplemental video showcases this effect. However, our display does not implement any occlusion cues either via content optimization [Mercier et al. 2017; Narain et al. 2015] or using optical design [Chang et al. 2020]. This is a limitation of our display; we discuss this in more detail in Section 6.

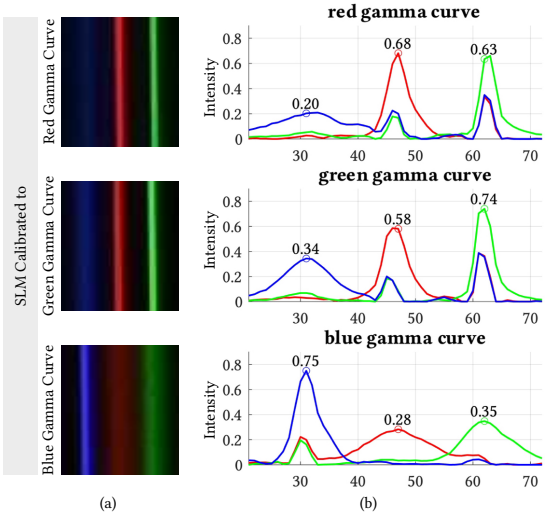


Fig. 13. **Comparison of focusing under different SLM gamma curves.** The scene in (a) is designed to test how the focus of the system changes as a function of wavelength. We display three colored line stripes at 0D with different gamma correction curves on the SLM. We can observe that each color lines become sharpest at the corresponding gamma curve. In (b), we average its intensity along the y direction. The color of the plot corresponds to the actual color channel. We can see that each peak reaches the maximum at the correct gamma curve. This provides a characterization of the difference in color performance between field sequential which would pick the peak for the best gamma curve, and field simultaneous which would be the peak corresponding to one of the rows.

Comparison to Focal Surface Displays. As mentioned in Section 2, focal surface displays [Matsuda et al. 2017] form the closest competitor to our idea, in that, both seek to implement a lens with a spatially-varying focal length. The stark difference in the nature of the two devices make a fair or direct comparison hard. Yet, it is illustrative to compare results on the same scene to understand differences in the quality of results and the reasons underlying them.

Figure 16 shows focal stack results from prototypes of Focal Surface and Split-Lohmann displays. The results for the Focal Surface display is from their paper; we used the same scene captured with the similar focus settings. Both prototypes were optimized to operate over a 4D working range with similar fields of view, and both were operated under the field simultaneous color mode. The results for focal surface display are with three surfaces, time multiplexed into one image; ours is without any time multiplexing.

In general, we observe that Split-Lohmann has better focusing performance over the entire working range of the scene, with fewer artifacts. The main reason for this is the way we use our respective SLMs. Focal Surface displays use SLMs as their primary unit for enabling spatially-varying focal length; the relatively low bending power of commercial SLMs today becomes a significant bottleneck for this design even when the scene is split across three focal surfaces. Split-Lohmann, on the other hand, use the SLM to tilt light (in our prototype, we do not even push the SLM to its limits in terms of tilting light); the actual heavy lifting in depth placement is done

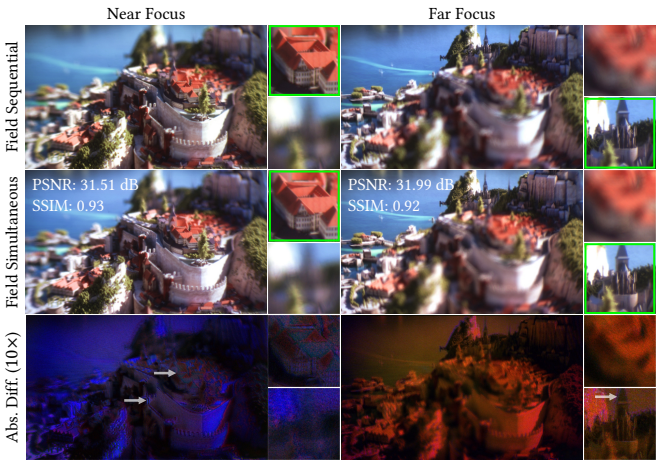


Fig. 14. **Comparison of color modes.** As with Figure 13, we compare field sequential and field simultaneous color modes, here on an extended scene over 0 D to 4 D. We observe an overhall high fidelity between the images captured from two modes, as shown in their PSNR and SSIM metric values. The absolute difference between the two color modes, intensified 10× highlight strong errors in blue and red channel; this is in addition to photon noise, that appears as a desaturated version of the original image. (Credits: "Scanlands" 3D scene courtesy of Piotr Krynski at Blender Studio)

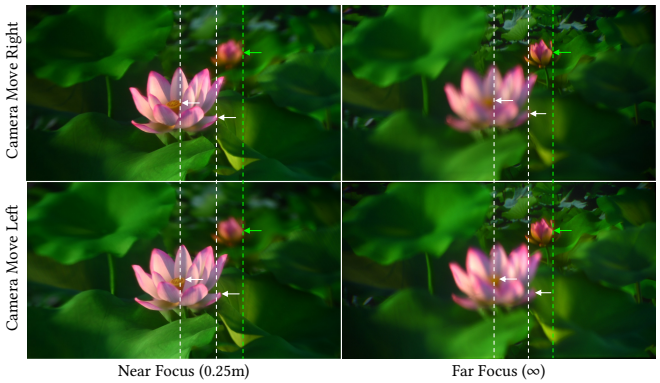


Fig. 15. **Observing parallax.** We show a scene with 4 D working range in field simultaneous color mode. Images were captured at two different focus settings: 0.25 m and ∞ m. The camera was moved horizontally by about 3 mm and we observe scene points closer to the camera translate a larger amount; the supplemental material has a video of this result. We overlay dotted lines and arrows as position reference—green for content at infinity and white for content at 0.25 m—to better highlight the amount of parallax. (Credits: 3D assets courtesy of "3D-ModelsArtist" at CGTrader)

using the cubic plates which are significantly thicker (50 µm) than phase SLMs. The differences in performance can also be attributed to ours being a table top prototype, and Focal Surface being a compact implementation. The supplemental material shows the same result, but in simulation.

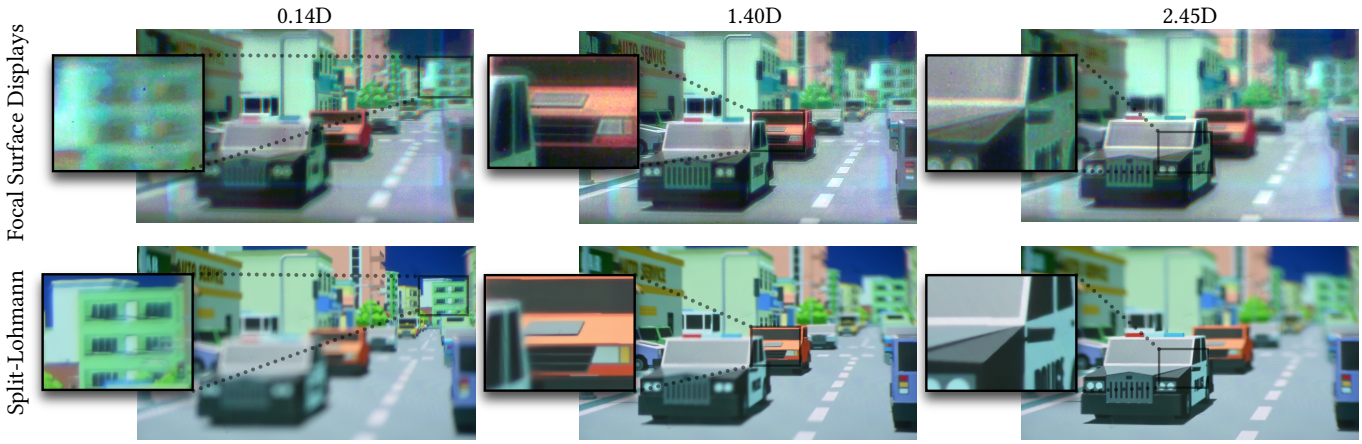


Fig. 16. **A qualitative comparison with Focal Surface display [Matsuda et al. 2017].** Results for both techniques are from real hardware prototypes, with results for Focal Surface Displays reproduced from their paper. We observe that Split-Lohmann exhibit sharper contents and a more natural defocus blur while having minimally observable noise. (Credits: "SimplePoly City" 3D scene courtesy of "VenCreations" at Unity Asset Store)

Gallery. A gallery of scenes is shown in Figure 17. All scenes occupy a 4D working range and captured under field simultaneous color mode. Focal stacks of these scenes can be found in the supplemental video.

5.3 Video and Interactive VR

Split-Lohmann displays has an extremely lightweight computational footprint for mapping RGBD content onto the the texture and phase patterns shown on the OLED and phase SLM, respectively. This allows to easily show videos and interactive VR scenes, like 3D games. Figure 18 shows stills from a video and an interactive 3D game; the supplemental material has a video recording of both scenes. The ability to easily port content with minimal overhead, while maintain high visual quality is one of the clear strengths of our technique.

6 DISCUSSION

We discuss some of the features of the proposed Split-Lohmann displays, including limitations that stem from various aspects of our design and implementation.

Leakage of defocus blur and content optimization. Multifocal displays typically suffer from the transparency of the focal planes, i.e., content on one depth does not block light from others. This typically manifests instead in the form of weakened occlusion cues due to the leakage of defocus blur from further away points through a content that is closer to the eye. Our technique is a "plain" realization of a multifocal display and inherits these limitations.

There are content optimization techniques [Choi et al. 2019; Mercier et al. 2017; Narain et al. 2015; Xiao et al. 2018a] that aim to resolve such issues. Such techniques jointly optimize the content shown on different focal planes, under a realistic image formation model of the multifocal display, so that the image seen by the eye as it focuses on different depths has realistic defocus and occlusion cues. Invariably, this requires the capability to have content at the same display pixel at multiple focal planes, especially for pixels near a depth edge. This is a capability that is beyond our current system *unless* we use temporal multiplexing; for example, focal surface displays benefit from content optimization when they temporal multiplex multiple surfaces. Given the complimentary nature of content optimization to the hardware design the we enable, we do believe this is a viable approach that can be incorporated into our system once we allow for temporally-multiplexed content.

Loss of sharpness in in-focus regions. A closer look at our lab prototype results, for example in Figure 17, shows a slight loss in sharpness at in-focus regions when compared to the image displayed on the OLED display. This is likely due to misalignments in the system; specifically, any non-diagonal shift between the two copies of the cubic wavefront will result in a non-spherical wavefront that the eye/camera cannot bring into sharp focus. This can likely be addressed using careful assembly and calibration. A second source of blur comes from aberrations introduced by the eyepiece (seen in the periphery of the field of view). This can potentially be resolved with spatially-varying calibration that, for each depth and spatial location, identifies ramps of different slopes and orientations.

Artifacts arising from the SLM. The use of a phase SLM results in a number of artifacts. There are two main effects we observe in our hardware results: phase warping artifacts, and unmodulated light.

Phase warping artifacts arise due to inherent smoothness in the retardation induced by the phase SLM; so, when a sharp 2π to 0 phase warp occurs, the SLM induces a smoothened phase delay which results in visible artifacts that reveal the warping contour. An example of this can be seen in the chess pieces in Figure 12, highlighted by the white arrows in the insets. There are a number of ways to resolve this including time-multiplexing the warping locations around and letting the artifacts average away. In our results, the effects are subtle and so we chose not to resolve them.

Phase SLMs also have non-ideal fill factors which results in unmodulated light; for our prototype, about 5% of the light is unmodulated. The unmodulated light will create an image of the OLED at

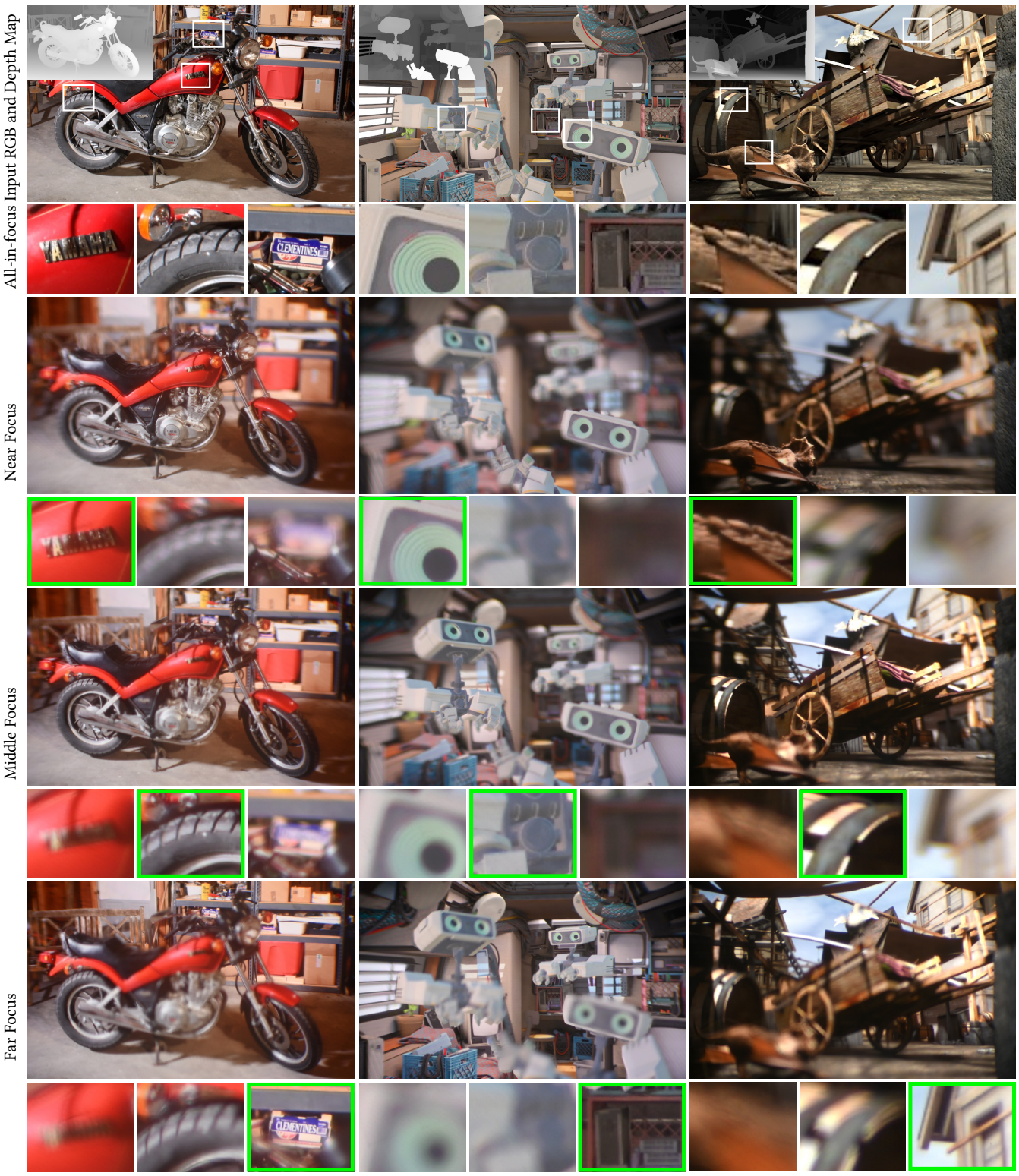


Fig. 17. **A gallery of static VR results.** (Credits: (left) RGBD content courtesy of the Middlebury 2014 Stereo Dataset [Scharstein et al. 2014]; (center) RGBD content courtesy of the DeepFocus dataset [Xiao et al. 2018b]; (right) Sintel 3D scene courtesy of Blender Studio)

Fig. 18. **A gallery of video and interactive VR results.** (left) We show three frames captured with a *streamed movie* played on our prototype, with manual adjustment of focus. The green inset shown the region in focus in each image. We show frames from three time instances: $t_0 < t_1 < t_2$. In this video, the camera is moving forward along the optical axis, and so points out of focus in earlier frame come in to focus. For example, the bananas go from out-of-focus at t_0 to in-focus at t_1. (right) An *interactive scene* with a user playing the Minecraft game. As before, we manually change the focus of the camera while capturing the video game played. At t_0, the player is standing still and our camera is focused at the near white plant. Then the camera switches focus at t_1. The player then continued walking towards the in-focus distanced region at t_2. (Credits: (left) "Interior Kitchen" 3D scene courtesy of "Entity Designer" at Blender Market and animation courtesy of Mohamad Zeina; (right) Craft 3D game courtesy of Michael Fogleman)

the P_5 plane. These effects can be observed in high contrast scenes, but are too subtle in scenes with dense content.

Improving spatial selectivity of depth. The primary factor that limits how accurately we can select depth, especially near discontinuities, if the size of the PSF induced by the bicubic phase plate. One approach to improve this is to move beyond bicubic phase plates and instead optimize a phase plate, or even a pair of phase plates. We could, in principle, use the optimization framework of Heide et al. [2016] to learn such a pair of phase plates. An advantage of such an approach would be that we can go beyond diagonal displacement of the two phase plates and use the entire range of directions available on the Fourier plane.

Compact implementations. The current implementation of the Split-Lohmann display lacks portability, which is, perhaps, sufficient for its use in ophthalmology and robot surgery; however, there are numerous applications where portability of the display is a primary concern. We briefly discuss potential approaches to achieving portable implementations.

One step towards this is to use lenses with smaller focal lengths in the construction of the 4f system; for example, Chang et al. [2018] use 30mm lenses which reduces the physical size of the system as well as provide a 3× increase in the size of the eyebox. Alternatively, we could follow the recent work of Maimone and Wang [2020] where polarization-sensitive optics is used to reduce the physical footprint of a device without changing its optical footprint. However, 4f systems also require lenses with sufficiently large diameters to avoid vignetting; specifically, the diameter needs to be greater than sum of the image and Fourier plane fields diameters. So, a second step is required where we go beyond 4f systems to implement the system. Supplemental Figure S7 provides the schematics and ray diagrams for an implementation that avoids the use of 4f relays. Here, the 4f system is replaced with a single lens of focal length f_0 placed at $2f_0$ from the display as well as the phase SLM; the cubic phase plate is collocated with this lens. This system provides a 1:1 image relay between the display and SLM. Suppose that the SLM has a constant phase pattern on it; the lack of telecentricity would result in the reflected light to be significantly shifted when it appears again on the lens plane; this would mean that we would couple the cubic wavefront with a shifted copy, where the shift depends on the lateral location of the point on the display. To resolve this, we introduce a field lens on top of the SLM; this lens is designed such that the reflected light from each point at the SLM overlaps exactly with the aperture of the cubic phase plate; due to mirror reflection, we get the same wavefront but flipped on the SLM and so the contributions from the cubic phase plate cancels out as we would like when the SLM does nothing. To create spatially-varying focus effects, as before, we add linear phase ramps on the SLM which produces the appropriate shifts between the cubic plates.

Supplemental Figure S7c shows rendered images with this system; we used ray tracing for this simulation. We can clearly see defocus effects that depend on the amount of tilt that we induce at the SLM. The biggest advantage of this system is that we can implement it with lenses with small f_0, and smaller diameters than what we would be able to with a 4f system. Similar systems have been implemented in near-eye configurations (for example, see Hua [2017] and Hu and Hua [2014]). There are, however, clear challenges in any such implementation including significantly enhanced spherical and chromatic aberrations when we decrease the focal length of the lenses, reducing the effects of which would require sophisticated optical engineering.

Conclusion. This paper advances multifocal displays by providing a single-shot display that has a vastly enhanced spatial selectivity of depth for display pixels. The central technical contribution of our system is an optical arrangement where we achieve a spatially-varying focus tunable lens using a "split" version of Lohmann lens,

enabling independent local and non-mechanical control on the amount of shift between the two cubic phase plates that constitute the Lohmann lens. The multifocal display that we construct preserves the resolution of the display, allowing the depth selection across 25-50 focal planes without any temporal multiplexing.

The concept of Split-Lohmann potentially applies to systems well beyond 3D displays. At its core, it provides the ability to change focus planes associated with different spatial locations on an image plane; basic duality between displays and cameras should in principle permit adaptive and spatially-varying focus capabilities, and perhaps even all-in-focus captures with a reworked system.

ACKNOWLEDGMENTS

We thank Nathan Matsuda for sharing results from the Focal Surface display, Jen-Hao Rick Chang for valuable discussions and insights, and Anqi Yang for early help with cubic phase plate fabrication. This work was supported by the National Science Foundation under the awards 1730147 and 2008464.

REFERENCES

Kaan Akşit, Ward Lopes, Jonghyun Kim, Peter Shirley, and David Luebke. 2017. Near-eye Varifocal Augmented Reality Display Using See-through Screens. *ACM Transactions on Graphics* 36, 6 (Nov. 2017), 189:1–189:13.

Kurt Akeley. 2004. *Achieving Near-correct Focus Cues Using Multiple Image Planes.* Ph.D. Dissertation. Stanford University.

Kurt Akeley, Simon J. Watt, Ahna Reza Girshick, and Martin S. Banks. 2004. A Stereo Display Prototype with Multiple Focal Distances. *ACM Transactions on Graphics* 23, 3 (Aug. 2004), 804–813.

Luis W Alvarez. 1978. Development of variable-focus lenses and a new refractor. *Journal of the American Optometric Association* 49, 1 (1978), 24–29.

Luis W Alvarez and William E Humphrey. 1970. Variable-power lens and system. US Patent 3,507,565.

Sergio Barbero. 2009. The Alvarez and Lohmann refractive lenses revisited. *Optics Express* 17, 11 (2009), 9376–9390.

Praneeth Chakravarthula, Ethan Tseng, Tarun Srivastava, Henry Fuchs, and Felix Heide. 2020. Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. *ACM Transactions on Graphics* 39, 6 (2020), 1–18.

Jen-Hao Rick Chang, B. V. K. Vijaya Kumar, and Aswin C. Sankaranarayanan. 2018. Towards Multifocal Displays with Dense Focal Stacks. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 198:1–198:13.

Jen-Hao Rick Chang, Anat Levin, B. V. K. Vijaya Kumar, and Aswin C. Sankaranarayanan. 2020. Towards Occlusion-aware Multifocal Displays. *ACM Transactions on Graphics* 39, 4 (2020), 68:1–15.

Ai Hong Chen, Daniel J. O'Leary, and Edwin R. Howell. 2000. Near visual function in young children. Part I: near point of convergence. Part II: amplitude of accommodation. Part III: near heterophoria. *Ophthalmic and Physiological Optics* 20, 3 (2000), 185–198.

Suyeon Choi, Seungjae Lee, Youngjin Jo, Dongheon Yoo, Dongyeon Kim, and Byoungho Lee. 2019. Optimal Binary Representation via Non-convex Optimization on Tomographic Displays. *Optics Express* 27, 17 (2019), 24362–24381.

Wei Cui and Liang Gao. 2017. Optical mapping near-eye three-dimensional display with correct focus cues. *Optics Letters* 42, 13 (2017), 2475–2478.

Joseph W Goodman. 2005. *Introduction to Fourier optics. 3rd.* Roberts and Company Publishers.

Felix Heide, Qiang Fu, Yifan Peng, and Wolfgang Heidrich. 2016. Encoded diffractive optics for full-spectrum computational imaging. *Scientific reports* 6, 1 (2016), 1–10.

David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. 2008. Vergence–accommodation Conflicts Hinder Visual Performance And Cause Visual Fatigue. *Journal of Vision* 8, 3 (2008), 33–33.

Xinda Hu and Hong Hua. 2014. High-resolution Optical See-through Multi-focal-plane Head-mounted Display Using Freeform Optics. *Optics Express* 22, 11 (2014), 13896–13903.

Hong Hua. 2017. Enabling Focus Cues in Head-mounted Displays. *Proc. IEEE* 105, 5 (2017), 805–824.

Youngjin Jo, Seungjae Lee, Dongheon Yoo, Suyeon Choi, Dongyeon Kim, and Byoungho Lee. 2019. Tomographic Projector: Large Scale Volumetric Display with Uniform Viewing Experiences. *ACM Transactions on Graphics* 38, 6 (Nov. 2019), 215:1–215:13.

Paul V Johnson, Jared AQ Parnell, Joohwan Kim, Christopher D Saunter, Gordon D Love, and Martin S Banks. 2016. Dynamic Lens and Monovision 3D Displays to Improve Viewer Comfort. *Optics Express* 24, 11 (2016), 11808–11827.

Robert Konrad, Nitish Padmanaban, Keenan Molner, Emily A. Cooper, and Gordon Wetzstein. 2017. Accommodation-invariant Computational Near-eye Displays. *ACM Transactions on Graphics* 36, 4 (July 2017), 88:1–88:12.

George Alex Koulieris, Kaan Akşit, Michael Stengel, Rafał K Mantiuk, Katerina Mania, and Christian Richardt. 2019. Near-eye display and tracking technologies for virtual and augmented reality. In *Computer Graphics Forum*, Vol. 38. 493–519.

Sheng Liu, Dewen Cheng, and Hong Hua. 2008. An Optical See-through Head Mounted Display with Addressable Focal Planes. In *IEEE/ACM International Symposium on Mixed and Augmented Reality*. 33–42.

Adolf W Lohmann. 1970. A new class of varifocal lenses. *Applied Optics* 9, 7 (1970), 1669–1671.

Gordon D Love, David M Hoffman, Philip JW Hands, James Gao, Andrew K Kirby, and Martin S Banks. 2009. High-speed Switchable Lens Enables the Development of a Volumetric Stereoscopic Display. *Optics Express* 17, 18 (2009), 15716–15725.

Andrew Maimone, Andreas Georgiou, and Joel S. Kollin. 2017. Holographic Near-eye Displays for Virtual and Augmented Reality. *ACM Transactions on Graphics* 36, 4 (July 2017), 85:1–85:16.

Andrew Maimone and Junren Wang. 2020. Holographic optics for thin and lightweight virtual reality. *ACM Transactions on Graphics* 39, 4 (2020), 67–1.

Nathan Matsuda, Alexander Fix, and Douglas Lanman. 2017. Focal Surface Displays. *ACM Transactions on Graphics* 36, 4 (July 2017), 86:1–86:14.

Olivier Mercier, Yusufu Sulai, Kevin Mackenzie, Marina Zannoli, James Hillis, Derek Nowrouzezahrai, and Douglas Lanman. 2017. Fast Gaze-contingent Optimal Decompositions for Multifocal Displays. *ACM Transactions on Graphics* 36, 6 (Nov. 2017), 237:1–237:15.

Rahul Narain, Rachel A. Albert, Abdullah Bulbul, Gregory J. Ward, Martin S. Banks, and James F. O'Brien. 2015. Optimal Presentation of Imagery with Focus Cues on Multi-plane Displays. *ACM Transactions on Graphics* 34, 4 (July 2015), 59:1–59:12.

Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A Cooper, and Gordon Wetzstein. 2017. Optimizing Virtual Reality for All Users Through Gaze-contingent and Adaptive Focus Displays. *Proceedings of the National Academy of Sciences* 114 (2017), 9.

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.

Yingsi Qin, Wei-Yu Chen, Matthew O'Toole, and Aswin C. Sankaranarayanan. 2023. Project Page: Split-Lohmann Multifocal Displays. https://imaging.cs.cmu.edu/split_lohmann.

Kishore Rathinavel, Hanpeng Wang, Alex Blate, and Henry Fuchs. 2018. An Extended Depth-at-field Volumetric Near-eye Augmented Reality Display. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (2018), 2857–2866.

Jannick P Rolland, Myron W Krueger, and Alexei Goon. 2000. Multifocal Planes Head-mounted Displays. *Applied Optics* 39, 19 (2000), 3209–3215.

Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling. 2014. High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth. In *Pattern Recognition*.

Liang Shi, Beichen Li, and Wojciech Matusik. 2022. End-to-end learning of 3d phase-only holograms for holographic display. *Light: Science & Applications* 11, 1 (2022), 247.

Guanjun Tan, Tao Zhan, Yun-Han Lee, Jianghao Xiong, and Shin-Tson Wu. 2018. Polarization-multiplexed multiplane display. *Optics Letters* 43, 22 (Nov 2018), 5651–5654.

David George Voelz. 2011. *Computational fourier optics: a MATLAB tutorial.* Vol. 534. SPIE press Bellingham, Washington.

Austin Wilson and Hong Hua. 2019. Design and demonstration of a vari-focal optical see-through head-mounted display using freeform Alvarez lenses. *Optics Express* 27, 11 (2019), 15627–15637.

Lei Xiao, Anton Kaplanyan, Alexander Fix, Matthew Chapman, and Douglas Lanman. 2018a. DeepFocus: Learned Image Synthesis for Computational Displays. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 200:1–200:13.

Lei Xiao, Anton Kaplanyan, Alexander Fix, Matthew Chapman, and Douglas Lanman. 2018b. DeepFocus: Learned Image Synthesis for Computational Displays. *ACM Transactions on Graphics* 37, 6 (dec 2018).

Tao Zhan, Jianghao Xiong, Junyu Zou, and Shin-Tson Wu. 2020. Multifocal displays: review and prospect. *PhotoniX* 1, 1 (2020), 1–31.