Analyzing phase masks for wide étendue holographic displays

Sagi Monin[1] Aswin C. Sankaranarayanan[2] Anat Levin[1]

Abstract—Spatial light modulator (SLM) technology forms the centerpiece of digital holographic displays. However, an inherent limitation of these devices is that their étendue, defined as the product of the display's eye box and field of view, is bounded by the number of pixel units. As a consequence, current SLMs are far from meeting the required field-of-view and eye box for the human visual system, which would require scaling the number of display units by a few orders of magnitude. Existing strategies for étendue-expansion rely on introducing a diffractive optical element (DOE), a fixed random phase mask whose pitch is much smaller than that of the original display, thereby spreading light over a wider angle. Displayed content is then optimized under perceptual constraints on the generated image. However, since the phase mask is fixed, the number of degrees of freedom does not increase and hence, the expansion in étendue necessarily comes with a loss of image quality. The trade-offs involved with such phase masks are not well understood.

This paper studies the space of phase masks that can be attached to an SLM to increase its angular range. It attempts to characterize what trade-offs are involved in étendue-expansion, and whatever specific phase mask designs would support better holograms.

Our theoretical results show that étendue expansion comes with a commensurate loss of contrast or resolution, depending on the specifics of the mask that we use. We show that while pseudo random masks support wide-étendue, they involve an inherent loss of contrast. Perhaps surprisingly, simple commonly-available phase masks like lenslet arrays provide near-optimal results that can largely outperform random masks.

Index Terms—Spatial light modulation, Holographic displays, Etendue

1 Introduction

Holographic displays [1], [2], [3], [4], [5], [6] are one of the promising approaches for generating three-dimensional (3D) visual content. A holographic display aims to replicate the wavefront generated by a real 3D scene; such a successful replication would produce a virtual scene that is *indistinguishable* from its real counterpart, and will successfully satisfy the numerous perceptual cues used by the human visual system for sensing depth.

In holographic near-eye displays we usually want the display to spread light over a wide field of view (FoV), but also allow the viewer to shift his pupil over a wide eye box. This eye box is equivalent to the spatial display area. Unfortunately these two quantities are tightly coupled with each other and both cannot be easily achieved using existing technology.

Digital holographic displays rely on spatial light modulation (SLM) technology, which are computer controlled arrays allowing an independent modulation of the phase in each pixel. Due to constraints imposed by Nyquist sampling, the maximal angle at which an SLM can bend light is bounded by its pixel pitch. As a result, controlling light at large angles requires display units smaller than a single micron, which is significantly below the pixel pitch of existing SLM arrays. It is possible to view the display using de-magnifying optics, which will reduce the pixel size. However, this also reduces the overall display area, contradicting the wide eye box goal. A short calculation shows that accommodating the human eye with a FoV that exceeds 120° and eye box that spans a centimeter, requires a display with 10 billion pixels, which is three orders of magnitude beyond current technology.

It can be shown that the *étendue* of the display, defined as the product of its spatial area and its maximal tilt angle, is only determined by the number of SLM pixels and is preserved through any lens system with which the SLM is combined. The étendue of existing SLMs is extremely limited.

Since étendue expansion is prohibitively daunting with the straightforward strategy of increasing pixel counts, recent works in this space have explored incorporating other optical elements into the display with the goal of breaking the aforementioned tradeoffs. In particular, it has been shown that introducing a diffractive optical element (DOE) or a phase mask in front of the SLM can have the effect of expanding its étendue [7], [8], [9], [10], [11]. Critically, the pitch of this phase mask is designed to be much smaller than the native SLM pitch; in turn, this allows the display to spread light over a much wider angular range, thereby increasing the FoV of the display with no change in its size or its eye box, thus increasing its étendue. However, a fixed phase mask does not provide any increase in the number of degrees of freedom in the display; hence, the expansion of the étendue *must* result in a loss of resolution or some other attribute of the displayed hologram. This characterization is hard for current designs in part due to the randomness of

- [1] *Department of Electrical Engineering, Technion.*
- [2] *Department of ECE, Carnegie Mellon University.*
- *sagim@campus.technion.ac.il, saswin@andrew.cmu.edu, anat.levin@ee.technion.ac.il*

the phase mask used.

This paper studies the space of étendue-expansion designs that are enabled by the addition of a diffractive phase mask. It attempts to answer the following two questions. First, what price do we pay for étendue expansion, and how can we characterize the space of holograms that can be created by combining a low resolution programmable SLM with a fixed higher resolution phase mask? The second question we address is the design of good phase masks. Assuming we are free to design and fabricate a DOE of our choice, what phase masks can create holograms of optimal quality?

Our analysis results in the following findings.

First, we provide a precise understanding on the loss of image quality that comes as a consequence of étendue expansion. Specifically, we show that étendue expansion by a certain amount results in a commensurate loss of contrast or resolution—a result that is consistent with the idea that the number of degrees of freedom in the device does not increase by the addition of a fixed phase mask.

Second, we show that random phase masks usually result in an inherent loss of visual quality, which is directly linked to their wide Fourier transform. This loss scales linearly with the expansion factor. Usually trading étendue for resolution results in fewer artifacts than trading it for contrast.

Third, we argue that the choice of optimal phase mask depends on the quality metric being measured. Previous works on étendue-expansion [7], [8] have measured the ability of the display to create 2D image targets of interest. We show that under this metric the ideal solution is attaching to the display a quadratic phase mask, or equivalently illuminating it with a spherical wave. This result is problematic, as it is well known that a quadratic phase mask cannot expand étendue, suggesting that metrics based on 2D targets are inadequate in capturing what we actually want from a wide-étendue display. Rather, we argue that the definition of good masks must involve metrics that take into account the 3D characteristics of the hologram. We adopt previous metrics such as a light field or a focal stack and use them to evaluate candidate phase masks.

Our fourth contribution is in showing that perhaps surprisingly, near optimal results can be obtained with a very simple phase mask equivalent to a lenslet array. Unlike the conventional usage of lenslet arrays in light-field displays, the lenslet array we suggest is placed in the Fourier domain, thus involving a significantly smaller number of lenslet units.

We hope the conclusions from this analysis will guide the design of future étendue expansion devices.

2 Related work

2.1 Holographic displays

Holography is one of the promising approaches for the design of near-eye displays for both virtual and augmented reality [12]. In a near-eye display, the SLM is placed close to the viewer's eye, and instead of displaying an intensity image, it displays a wavefront representing the propagation of a desired 3D scene to this plane [1], [13], [14], [15], [16], [17], [18], [19], [20], [21]. In theory the presented wavefront generate different images based on different positions and focus conditions of the viewer's eye.

A central challenge in holographic display design is content generation. While early work used 2D image objectives, recent ones [13], [14], [15] have looked at generating real holographic content by explicitly incorporating 3D focal stack content into the optimization. In this paper, we further emphasize that incorporating 3D content in the design stage is crucial.

2.2 Overcoming étendue limits

There has been several approaches for overcoming the limited étendue of holographic displays.

One approach tries to track the viewer's pupil and steer the hologram towards it [22], [23]. These approaches require an additional camera to track the viewer's gaze. Pupil tracking and steering light is not limited to holographic displays but also used in varifocal near-eye displays for dynamically foveated displays [24].

A different technique for étendue expansion tries to create multiple copies of the eye box allowing the hologram to be viewed from multiple eye positions. In [25] a holographic optical element (HOE) is placed in front of the eye that has the structure of a periodic lens array. Alternatively, multiple copies of the eye box are generated by utilizing higher diffraction orders of the SLM [26], [27]. However, such copies replicate the same hologram over a wider range, and do not allow the display of viewpoint dependent effects.

An alternative approach is to expand the étendue by attaching a static random mask, whose resolution is higher than the SLM. Different types of masks were proposed for étendue expansion: [10] suggested using a fully random phase mask and utilizing algorithms for focusing through scattering media, [9] used binary phase masks and [11] relied on a designed photon-sieve (pinhole mask). These works showed the potential of enlarging étendue by attaching a phase mask; however, they were only able to generate sparse images. Recently [7] was able to generate high quality étendue expanded images by utilizing an optimization framework and relying on the limited resolution of the human visual system. A follow up work [8] suggested a method for optimizing the shape of the phase mask, increasing the quality of the generated holograms. However, such a learning based approach can converge to local minima and its output is hard to analyze. In contrast, we design our mask using simple physical principles leading to explicit analytic results.

An orthogonal route for Étendue expansion is time multiplexing [28], [29]. In particular, Lee *et al*. [30] used a fast binary DMD combined with 6 laser sources illuminating it in different angles. As each source is tilted differently the hologram it creates fills a different sub-cone out of the desired FoV. However, due to the limited binary modulation of this display, multiple SLM patterns need to be multiplexed to create a single sub-cone, and hence the amount of temporal multiplexing that can be used to expand étendue is limited. A short calculation shows that given the temporal resolution of the human eye, the maximal étendue that a fast DMD can support by exploiting temporal multiplexing, is still lower than the étendue of slower liquid crystal based

SLMs, which simply contain 4K pixels. A similar calculation shows that the recent fast mirror based phase SLM of TI [31] is still not fast enough to provide a significant étendue expansion.

3 Background and Problem Formulation

In this section, we define étendue and introduce the basic expansion strategies.

Notation. We denote a 2D spatial position using $\vec{\boldsymbol{x}} = (x, y)$, and the first two coordinates of a direction (unit norm 3D) vector using $\vec{\boldsymbol{\theta}}$. Throughout this paper we target small angles, where the paraxial approximation implies $\sin(\theta) \approx \theta$, $\cos(\theta) \approx 1$.

3.1 SLM modulation and its limitations

We start with an overview of the spatial and angular limits of SLMs, following [21]. We model SLMs as arrays of pixels, whose phase can be independently controlled, resulting in a piecewise-constant phase mask $\phi_c(\lfloor\vec{\boldsymbol{x}}\rfloor)$, with $\lfloor\vec{\boldsymbol{x}}\rfloor$ denoting the rounding of $\vec{\boldsymbol{x}}$ to the nearest cell center. Assuming a pixel pitch of Δ_x, a display with of $\mathcal{N} \times \mathcal{N}$ pixels has a range $\vec{\boldsymbol{x}} \in [-\frac{1}{2}\Omega_x, \frac{1}{2}\Omega_x] \times [-\frac{1}{2}\Omega_x, \frac{1}{2}\Omega_x]$ with:

$$\Omega_x = \Delta_x \mathcal{N}. \tag{1}$$

When viewed from direction $\vec{\boldsymbol{\theta}}$ the display output can be computed as the Fourier transform of its modulation function

$$u_{out}(\vec{\boldsymbol{\theta}}) = \int e^{\frac{-i2\pi}{\lambda}\vec{\boldsymbol{\theta}}\cdot\vec{\boldsymbol{x}}} e^{i\phi_c(\lfloor\vec{\boldsymbol{x}}\rfloor)} u_{in}(\vec{\boldsymbol{x}}) d\vec{\boldsymbol{x}}. \tag{2}$$

The term u_{in} is the incident illumination, which we assume to be spatially uniform and, hence, a constant that can be dropped.

From Eq. (2) we see that to send light toward direction $\vec{\boldsymbol{\theta}}$ we need to display on the SLM a sinusoid of the form $\phi_c(\vec{\boldsymbol{x}}) = \frac{2\pi}{\lambda}\vec{\boldsymbol{\theta}} \cdot \vec{\boldsymbol{x}}$. In a SLM we can only display a piecewise constant approximation to this sinusoid. We can sample this sinusoid without aliasing only if its periodicity is wide enough to span at least two SLM pixels. Thus, Nyquist dictates that realizable tilts are limited to the range $\vec{\boldsymbol{\theta}} \in [-\frac{1}{2}\Omega_\theta, \frac{1}{2}\Omega_\theta] \times [-\frac{1}{2}\Omega_\theta, \frac{1}{2}\Omega_\theta]$ with

$$\Omega_\theta = \frac{\lambda}{\Delta_x}. \tag{3}$$

The maximal tilt angle is inversely limited by the SLM pitch. For example, the smallest pixel pitch among current commercially-available SLM is $3.74\mu m$, which limits θ to about $4°$, in visible wavelengths. This limited angular range is one of the largest barriers to the practical utility of digital holographic displays.

The *étendue* of the display marked as $\boldsymbol{\mathcal{E}}$ below, is defined as the product of its spatial and angular ranges. Putting together Eqs. (1) and (3), we observe that this quantity is a constant that depends only on the wavelength of light and the number of pixels on the SLM:

$$\boldsymbol{\mathcal{E}}^2 = \Omega_x^2 \cdot \Omega_\theta^2 = (\lambda\mathcal{N})^2. \tag{4}$$

Fig. 1(a) illustrates the SLM used in a simplified near-eye geometry. The SLM is placed at the focal plane of a lens with focal length f_1, generating a virtual image at distance f_1 after this lens. In the presence of fully coherent illumination this image corresponds to the Fourier transform of the wavefront displayed on the SLM:

$$\left|\mathcal{F}\left(e^{i\phi_c(\vec{\boldsymbol{x}})}\right)\right|^2. \tag{5}$$

A second eye-piece lens of focal length f_2 is used to project this image to optical infinity before it enters the eye. The eye-piece lens and the lens of the eye itself form a relay system allowing the eye to see the virtual image on the Fourier plane marked in (Fig. 1(a)). The FoV of the display corresponds to the maximal angle it can send to the eye. When viewed via the optics of Fig. 1(a) this angle is scaled to $\frac{f_1}{f_2}\Omega_\theta$. The eye-box is defined as the area of content before the eye, over which the viewer can shift his eye. If the original SLM area is Ω_x, it is scaled through the relay system to $\frac{f_2}{f_1}\Omega_x$. Thus the product of the FoV and the eye-box equals the étendue as defined in Eq. (4). This quantity is invariant to the choice of f_1 and f_2. If we wish to increase either FoV or eye-box, we unavoidably decrease the other.

As mentioned earlier, a good near-eye display should accommodate a FoV of $120°$ and an eye box of 1cm. This however, would require a pixel count three orders of magnitude higher than current technology. Such a large number of pixel units would result in angular resolution which is way above what the human retina can resolve. Yet, there is no simple way to trade resolution and étendue.

3.2 Étendue expansion with random phase masks

An approach to expand the étendue of the SLM, and thereby circumvent the limitations of Eq. (4), is to attach to the SLM a DOE realizing a random phase mask with a higher resolution [7], [8], [9], [10], [11]. Assuming the SLM and the mask are co-planar, we get a phase modulation of the form

$$\phi(\vec{\boldsymbol{x}}) = \phi_c(\vec{\boldsymbol{x}}) + \phi_m(\vec{\boldsymbol{x}}), \tag{6}$$

where ϕ_c is the programmable piecewise constant SLM phase and ϕ_m is a fixed phase mask. Let us denote the pitch of ϕ_m by Δ_x^m, and assume $\Delta_x^m < \Delta_x$, then the combined phase ϕ can send light over an extended angular range

$$\widehat{\Omega}_\theta = \frac{\lambda}{\Delta_x^m}. \tag{7}$$

For the same display area this increases the étendue by a factor $\boldsymbol{q_\mathcal{E}} \times \boldsymbol{q_\mathcal{E}}$ with

$$\boldsymbol{q_\mathcal{E}} = \frac{\widehat{\boldsymbol{\mathcal{E}}}}{\boldsymbol{\mathcal{E}}} = \frac{\widehat{\Omega}_\theta}{\Omega_\theta} = \frac{\Delta_x}{\Delta_x^m}. \tag{8}$$

Introducing a random phase mask complicates the relationship between the generated hologram and the phase pattern displayed on the SLM. To display a target intensity image I on the Fourier plane, we can optimize for the SLM pattern ϕ_c by minimizing the difference between the Fourier transform of the modulation function and the target image:

$$\underset{\phi_c}{\operatorname{argmin}} \left\| g * \left|\mathcal{F}\left(e^{i(\phi_c(\vec{\boldsymbol{x}})+\phi_m(\vec{\boldsymbol{x}}))}\right)\right|^2 - g * I \right\|^2, \tag{9}$$

where $*$ denotes convolution, and g is a low-pass filter. One motivation for including g is that the display resolution

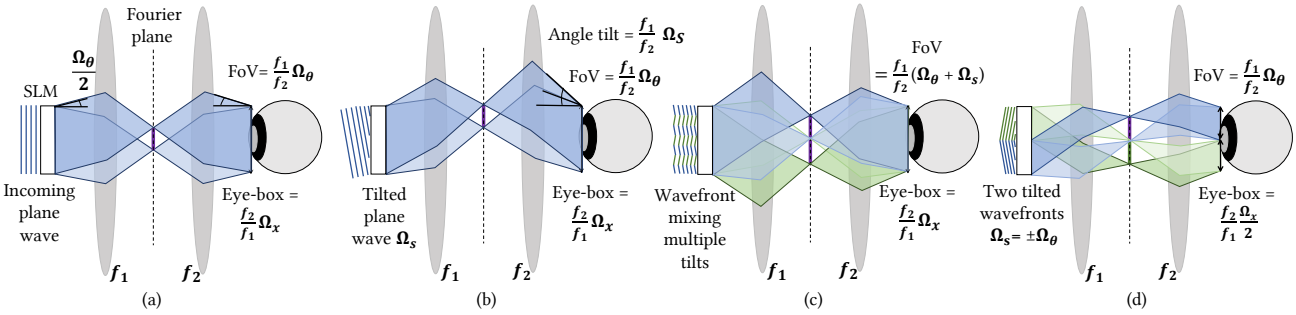


Fig. 1: Demonstration of étendue: (a) Near-eye holography setup where the FoV is the maximum diffraction angle of the SLM and the eyebox is determined by the size of the SLM. These properties can be traded by the ratio of focal lengths f_2/f_1. The purple line indicates the angular range of the SLM in the Fourier plane. (b) Illuminating the system with a tilted wavefront of angle Ω_s shifts the virtual image on the Fourier plane; however, it does not affect the étendue. (c) Illuminating the system with a piecewise linear wavefront mixing multiple tilt directions results in a larger FoV with the same eyebox size. In the illustration green and blue mark different slopes. (d) Dividing the SLM to two and illuminating each part with a tilted wavefront results in a larger FoV; however, each part of the image resides at only half of the eyebox.

is often higher than the angular resolution the human eye can resolve. Below we show that this low-pass filter has a fundamental role in the hologram design and analysis.

Usually we cannot find phase masks ϕ_c minimizing Eq. (9) with zero error. To see this we start with a simple count of the number of degrees of freedom. By the Nyquist principle, a normal piecewise constant display of pitch Δ_x and range Ω_x can produce a Fourier transform of range Ω_θ given by Eq. (3) and pitch

$$\Delta_\theta = \frac{\lambda}{\Omega_x}. \tag{10}$$

By plugging Eq. (1) it is easy to see that the ratio between angular range and angular resolution is equivalent to the number of display pixels $\Omega_\theta/\Delta_\theta = \mathcal{N}$.

With a random phase mask added to the display the angular resolution should follow Eq. (10), as the spatial range of the display did not change. The angular range was extended to $\widehat{\Omega}_\theta$. Thus the ratio between range and resolution has now increased to $\widehat{\Omega}_\theta/\Delta_\theta = \boldsymbol{q_\mathcal{E}}\mathcal{N}$, and target images can contain $(\boldsymbol{q_\mathcal{E}}\mathcal{N})^2$ entries. However, the programmable phase mask ϕ_c in Eq. (9) has only $\mathcal{N}^2$ degrees of freedom rather than $(\boldsymbol{q_\mathcal{E}}\mathcal{N})^2$. Hence, we do not expect to be able to generate arbitrary signals I with a resolution Δ_θ and range $\widehat{\Omega}_\theta$. In Sec. 4 we try to understand what family of signals we can generate and what do we trade for extending étendue.

3.3 Piecewise-linear phase modulation

To analyze phase modulations, in this paper we consider the family of piecewise-linear modulations. We express our phase masks as

$$\phi_m(\vec{\boldsymbol{x}}) = s(\lfloor\vec{\boldsymbol{x}}\rfloor)\cdot\vec{\boldsymbol{x}}, \tag{11}$$

where $\vec{\boldsymbol{s}}(\lfloor\vec{\boldsymbol{x}}\rfloor)$ is a 2D vector $\vec{\boldsymbol{s}} = (s_x, s_y)$ denoting the horizontal and vertical slope of pixel $\lfloor\vec{\boldsymbol{x}}\rfloor$. The overall phase mask we can express $\phi(x) = \phi_c(x) + \phi_m(x)$, is a combination of piecewise constant and piecewise linear components.

This family encompasses a wide range of phase modulations including those with random tilts at each SLM pixel, as well as others that are simpler to realize and produce higher quality results.

Symbol	Meaning
$\phi_c(\vec{\boldsymbol{x}})$	Piecewise constant phase.
$\phi_m(\vec{\boldsymbol{x}})$	Fixed mask phase.
$\mathcal{N}$	Number of pixel in one axis of the SLM.
Δ_x	SLM pitch.
Ω_x	SLM range width.
Δ_θ	Angular pitch.
Ω_θ	Angular range of original SLM.
$\widehat{\Delta}_\theta$	Low resolution angular pitch in expanded display.
$\widehat{\Omega}_\theta$	Expanded angular range.
$\mathcal{N}_{eb}$	Number of light field eye-box views.
$\Delta_\theta^{lf} = \mathcal{N}_{eb}\cdot\widehat{\Delta}_\theta$	Angular light field pitch.
$\boldsymbol{q_\mathcal{E}}$	Étendue expansion factor.

TABLE 1: Notations

Tilting the phase mask ϕ is equivalent to shifting its Fourier transform as illustrated in Fig. 1(b). Hence, by adding a range of tilt directions in front of the SLM (Fig. 1(c)) we expand the angular range at which we can display content.

Since the angular range of the wavefront is related to the slopes of the phase mask and not to their absolute phase values, our choice in Eq. (11) to parameterize the phase mask by their slopes simplifies analysis.

4 Analyzing Piecewise-Linear Phases

We analyze the range of producible holograms using piecewise linear phase modulations and characterize what we trade for étendue expansion. Specifically, we introduce below two strategies for the design of the fixed phase mask ϕ_m and the SLM content ϕ_c. For a target étendue expansion factor $\boldsymbol{q_\mathcal{E}} \times \boldsymbol{q_\mathcal{E}}$, the strategy in Sec. 4.1 comes at the price of a factor $\boldsymbol{q_\mathcal{E}^2}$ reduction in contrast; and the strategy in Sec. 4.2.2 comes the price of a factor $\boldsymbol{q_\mathcal{E}^2}$ reduction in resolution. Both strategies are outperformed by a direct optimization of Eq. (9), but their analysis will allow us to understand the trade-offs of étendue expansion.

As part of our study, we also ask what phase masks can support the display of better holograms. The construction in Sec. 4.1 leads to highly varying, pseudo random phase masks ϕ_m, and explains why such phase masks result

in severe loss of contrast. In many applications, trading étendue for spatial resolution may be more desired than trading étendue for contrast. With the resolution trade-off in mind, the analysis in Secs. 4.2 and 4.3 reveals other types of phase mask designs that can lead to higher quality results compared to random masks, including those that are easy to implement in hardware, such as a lenslet array with properly chosen pitch and focal length.

Our analysis also reveals that the 2D image quality metric of Eq. (9) may not be the right way to evaluate étendue-expanded holograms and we advocate metrics that take into account 3D content.

4.1 Analyzing contrast-étendue tradeoff

We start by considering a simple strategy for setting the phase mask ϕ_c in an étendue-expanded display. This strategy produces any target I with a wide range $\widehat{\Omega}_\theta$ (the extended angular range) and high resolution Δ_θ (the idealized angular resolution, without low-pass filtering); since such targets have $(\boldsymbol{q_\mathcal{E}}\mathcal{N})^2$ degrees of freedom, we show that they can be produced with $\mathcal{N}^2$ SLM pixels *at the price of contrast reduction by a factor of* $\boldsymbol{q}_\mathcal{E}^2$. This will also allow us to understand the problems of fully random phase masks.

The proposed strategy relies on a simple observation: a linear phase ramp in the SLM plane leads to a shift of its Fourier transform. Since we wish to generate our image as the Fourier transform of the SLM, such phase ramps allow us to display content at different parts of the target. Hence, we can design content for different sub-parts of the target image separately and then spatially interleave them on the SLM to form a complete image.

Specifically, we divide the target image I into $K = \boldsymbol{q}_\mathcal{E}^2$ equal-sized sub-images $I^1, \dots I^K$ centered at a 2D grid of points of the form $\vec{\boldsymbol{\theta}}^k = (k_1, k_2)\Omega_\theta$ for integer (k_1, k_2) values in the range $|k_1| < \frac{1}{2}\boldsymbol{q_\mathcal{E}}, |k_2| < \frac{1}{2}\boldsymbol{q_\mathcal{E}}$. Each sub-image has an area $\Omega_\theta \times \Omega_\theta$ (the angular range of a basic piecewise constant display) and resolution Δ_θ. As the number of degrees of freedom in I^k is $\mathcal{N}^2$, which is also the number of degrees of freedom in the piecewise constant phase ϕ_c, we can solve a standard phase retrieval problem to find a phase mask ϕ_c^k producing I^k in the central square of the range, as illustrated in the top row of Fig. 2. By adding on the entire SLM a linear phase ramp with a fixed slope $\vec{\boldsymbol{s}}^k = \vec{\boldsymbol{\theta}}^k$, we can shift the position of I^k anywhere in the extended range $\widehat{\Omega}_\theta \times \widehat{\Omega}_\theta$, as illustrated in the 2nd row of Fig. 2. With this construction we achieve a complex field

$$u^k = \mathcal{F}\left(e^{i(\phi_c^k(\vec{\boldsymbol{x}}) + \vec{\boldsymbol{s}}^k \cdot \vec{\boldsymbol{x}})}\right), \tag{12}$$

such that for angles $\vec{\boldsymbol{\theta}}$ in the sub-range of I^k, satisfying $|\vec{\boldsymbol{\theta}} - \vec{\boldsymbol{\theta}}^k| \leq \frac{1}{2}\Omega_\theta$, we get:

$$|u^k(\vec{\boldsymbol{\theta}})|^2 \approx I^k(\vec{\boldsymbol{\theta}}). \tag{13}$$

To produce the full image I we randomly allocate each SLM pixel to one of the basis slopes $\vec{\boldsymbol{s}}^1, \dots \vec{\boldsymbol{s}}^K$, and denote by m^k the mask of pixels with slope $\vec{\boldsymbol{s}}^k$. m^k are binary and for each pixel $\vec{\boldsymbol{x}}$, $\sum_k m^k(\vec{\boldsymbol{x}}) = 1$. We combine the independent solutions ϕ_c^k into a single one using

$$\phi(\vec{\boldsymbol{x}}) = \sum_k m^k(\vec{\boldsymbol{x}}) \cdot (\phi_c^k(\vec{\boldsymbol{x}}) + \vec{\boldsymbol{s}}^k \cdot \vec{\boldsymbol{x}}), \tag{14}$$

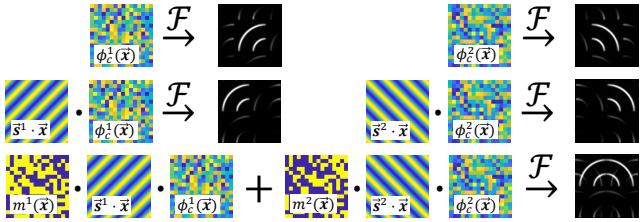


Fig. 2: Forming wide-étendue holograms from small ones. Top row: we solve the phase retrieval problem to find a piecewise constant phase mask generating a desired pattern of support $\Omega_\theta \times \Omega_\theta$, at the center of the target $\widehat{\Omega}_\theta \times \widehat{\Omega}_\theta$ area. Second row: By augmenting the piecewise constant mask with a linear ramp (visualized as a sinusoid) we can shift the content anywhere inside the wide $\widehat{\Omega}_\theta \times \widehat{\Omega}_\theta$ target area. Third row: we use a random mask and its complementary to merge the two holograms in the second row so that we now jointly display content at two sub-regions.

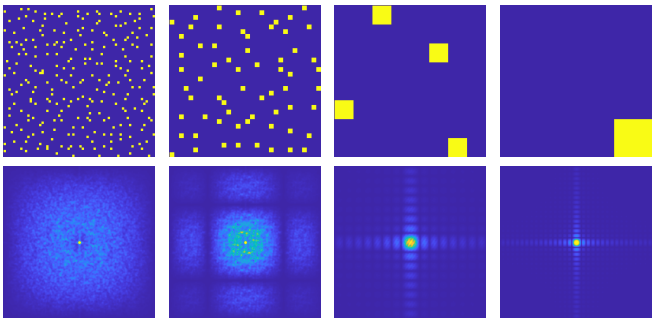


Fig. 3: The effect of bin size in the slope mask. The top row demonstrates 4 masks m^k (indicating the SLM pixels displaying a particular slope s_k) with different bin sizes, and the lower row the magnitude of their Fourier transforms M^k. With narrow bins the spectrum is a narrow impulse with a wide tail of lower power. As we increase bin size the tail support shrinks and eventually the spectrum becomes a low-pass filter. Note that all columns use the same resolution and range, and the zooming effect is solely due to the different patch size.

as illustrated in the last row of Fig. 2.

To quantify how well the Fourier transform of this mask explains the target image we find the scale α of the target image best explaining the hologram

$$\alpha = \underset{\alpha}{\operatorname{argmin}} \left\| \left| \mathcal{F}\left(e^{i(\phi_c(\vec{\boldsymbol{x}}))}\right)\right|^2 - \alpha I \right\|^2, \tag{15}$$

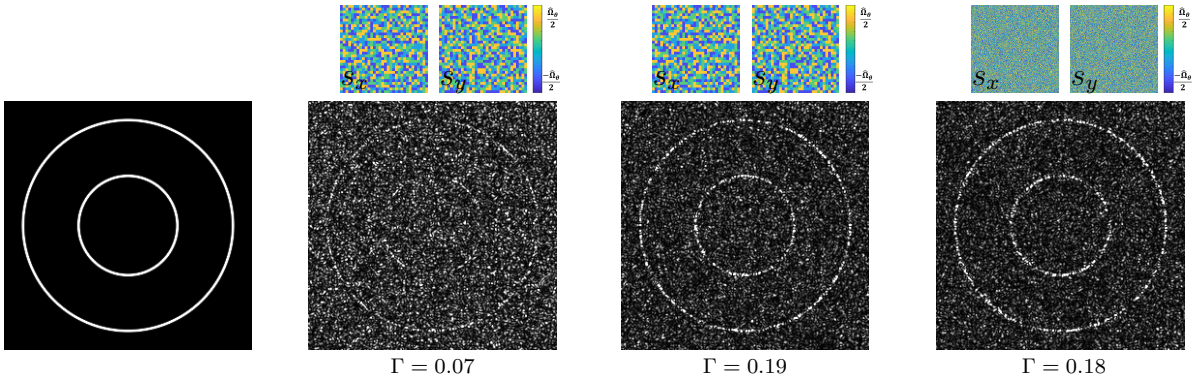
which can be computed in closed-form by projecting $\left|\mathcal{F}\left(e^{i(\phi_c(\vec{\boldsymbol{x}}))}\right)\right|^2$ onto I. We measure the signal to noise ratio

$$\Gamma(\phi) = \frac{\|\alpha I\|^2}{\left\| \left| \mathcal{F}\left(e^{i(\phi_c(\vec{\boldsymbol{x}}))}\right)\right|^2 - \alpha I \right\|^2}. \tag{16}$$

Below we also refer to the $\Gamma(\phi)$ metric as contrast because as we blur the hologram the noise appears as a global haze.

Claim 1. The combined phase mask $\phi(\vec{\boldsymbol{x}})$ of Eq. (14) produces I at the price of a factor $\boldsymbol{q}_\mathcal{E}^2$ contrast reduction.

A formal proof of this claim is provided in App. A.1. For the proof we characterize the randomness of the masks m^k,



$\Gamma = 0.07$ $\Gamma = 0.19$ $\Gamma = 0.18$

(a) Target (b) Indp. sub-squares (c) Piecewise constant opt. s.t. piecewise linear phase (d) Piecewise constant opt. s.t. random phase

Fig. 4: Wide-étendue content creation: We simulate étendue-expansion at factor 4×4, and compare strategies for content creation at the native pitch of the original display, Δ_θ (without any low pass filtering). As the number of degrees of freedom $\mathcal{N}^2$ is smaller than the number of entries in the target $(\boldsymbol{q_\mathcal{E}}\mathcal{N})^2$, all solutions involve some artifacts. (a) Target. (b) Naive content creation where we solve $\boldsymbol{q_\mathcal{E}} \times \boldsymbol{q_\mathcal{E}}$ independent phase retrieval problems and merge them using random masks as in Eq. (14). (c) Optimizing the piecewise constant part of Eq. (9) subject to the same fixed piecewise linear phase used in (b), improving contrast. (d) Optimizing the piecewise constant content subject to a random phase mask as in [7], with results similar to our random piecewise linear phase mask in (c). Colored squares above each result visualize the s_x, s_y slopes of the phase defined in Eq. (11). Numerical SNR contrast values using the Γ metric of Eq. (16) are reported under each image.

by assuming they are constant over bins of b pixels, where in a fully random mask $b = 1$; and bigger b values express mask areas with similar phases. Fig. 3 visualizes random masks m_k with different bin sizes and the absolute value of their Fourier transforms $M^k = \mathcal{F}(m^k)$. Multiplying the phase ϕ_c^k of the individual sub-squares in Eq. (14) with m^k is equivalent to convolving the targets u^k with their Fourier transform M^k. To achieve the desired target I^k, ideally M^k should be an impulse. In Fig. 3 the spectra includes a narrow impulse, but also content around it. Below we refer to any content in M_k around the central impulse as "noise tail". The width of this noise tail is governed by the bin size b. As part of the proof of Claim 1 we show that for an étendue-expansion factor $K = \boldsymbol{q}_\mathcal{E}^2$ the impulse contains only $1/K$ of the energy of M^k while the rest of the energy is wasted on the noise tail. Hence $M^k * u^k$ produces I^k plus a big noise term which is K times higher than the original signal.

Fig. 4(b) illustrates an image generated by this scheme, targeting an étendue expansion factor $\boldsymbol{q_\mathcal{E}} = 4$ in each axis. We can achieve the original angular resolution Δ_θ without any low-passing. Yet, a significant amount of noise is being added to the desired signal, reducing contrast. In the caption we quantify the contrast reduction using the metric of Eq. (16).

So far we considered optimizing each sub-image independently, In Fig. 4(c) we demonstrate that one can largely improve over the naive combination of independent solutions by directly optimizing for the piecewise constant part ϕ_c in Eq. (9), subject to the random slope masks $\phi_m(\vec{\boldsymbol{x}}) = \sum_k m^k(\vec{\boldsymbol{x}}) \cdot (\vec{\boldsymbol{s}}^k \cdot \vec{\boldsymbol{x}})$. Still, the contrast of the resulting image is reduced and speckle noise is present. In Fig. 4(d) we follow [7] and optimize the piecewise constant mask ϕ_c subject to a fully random phase mask of pitch $\Delta_x^m = \Delta_x / \boldsymbol{q_\mathcal{E}}$ defined in Eq. (6) with similar results. We have verified our results against an online implementation[1]. While these random phase masks are not piecewise linear, we empirically observe that their performance is similar to that of piecewise-linear phase masks where every pixel takes an independent slope.

4.2 Analyzing resolution-étendue tradeoff

In the previous section we have shown that despite the fact that the number of pixels in the SLM is smaller than the number of degrees of freedom in the target étendue-expanded image, one can generate the target at the original angular resolution Δ_θ, at the price of severe noise. However, the visual quality of this approach, as demonstrated in Fig. 4 is far from being compelling. Thus, we seek a different tradeoff. In this section we study the trading of étendue with resolution.

As analyzed above, if we increase the range of the target to $\widehat{\Omega}_\theta = \boldsymbol{q_\mathcal{E}}\Omega_\theta$ and keep the native angular resolution Δ_θ, we end up with $(\boldsymbol{q_\mathcal{E}}\mathcal{N})^2$ entries in the target but only $\mathcal{N}^2$ freedom degrees in the SLM. To match the degrees of freedom we can reduce the target resolution, or equivalently increase the pitch to

$$\widehat{\Delta}_\theta = \boldsymbol{q_\mathcal{E}}\Delta_\theta. \tag{17}$$

Thus, we select the width of the low-pass filter in the hologram generation cost of Eq. (9) as $\widehat{\Delta}_\theta$, and use the notation $g_{\widehat{\Delta}_\theta}$ to denote a low pass kernel with that width.

4.2.1 *The effect of different slope masks.*

In Fig. 5 we used gradient based optimization to minimize the basic 2D hologram score of Eq. (9) subject to a low-pass filter $g_{\widehat{\Delta}_\theta}$. We simulate 2×2 and 4×4 étendue expansions, targeting pitches $\widehat{\Delta}_\theta = 2\Delta_\theta$ or $\widehat{\Delta}_\theta = 4\Delta_\theta$, so that the

1. https://github.com/dongheon-yoo/Holographic-Display-etendue-Expansion

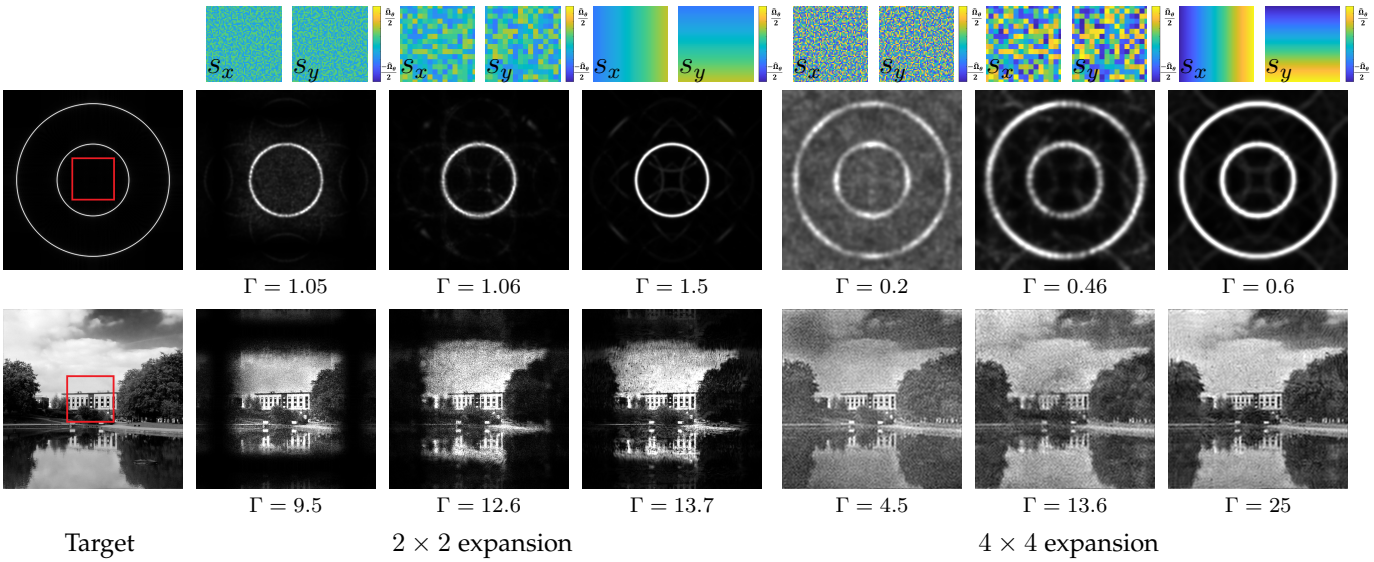


Fig. 5: Optimizing wide-étendue content: We simulate étendue-expansion at 2×2 and 4×4 factors, correspondingly targeting pitches $\widehat{\Delta}_\theta = 2\Delta_\theta$ and $\widehat{\Delta}_\theta = 4\Delta_\theta$. These larger pitches allow us to match the number of target entrees with the number of degrees of freedom, leading to higher quality results compared to Fig. 4. We compare piecewise linear slope masks with different bin sizes, where the largest bin size we can use is approximated by a continuously varying slope in the last column. The original range of the display without the piecewise linear phase mask is marked on the target using a red frame. Colored squares above each result visualize the s_x, s_y slopes of the piecewise linear phase as defined in Eq. (11). For the 2×2 expansion we used smaller slopes (see lower color range in the s_x, s_y images) and hence could generate content only at the central sub-frames. For the 4×4 expansion, slope masks involving big patches of constant slopes lead to better contrast. For the smaller expansion factor contrast loss is less of an issue, in agreement with Claim 1. Numerical contrast values using the Γ metric of Eq. (16) are reported under each image.

complexity of the target matches the number of freedom degrees. As we use larger pitches for the higher expansion factor, we effectively target a lower resolution.

We compare solutions that allocate patches of varying sizes of constant slopes. We start with fully random patches where each pixel has an independent slope, we increase to piecewise linear allocations with medium patch sizes, and finally monotonically varying slopes, which correspond to the largest patches one can use, as discussed in Sec. 4.2.2 below.

For 2×2 expansion, all masks achieve good contrast. However, in the 4×4 expansion, random masks with very small bins result in very low contrast. The global haze added to the images results from blurring the noise in the higher resolution holograms of Fig. 4. In agreement with Claim 1, the contrast reduction scales with the étendue-expansion factor, and in the 4×4 example it is more severe than in the 2×2 example. Higher expansion factors would involve an even lower contrast, and App. Fig. 13 demonstrates this using an 8×8 expansion example.

While by minimizing Eq. (9) we jointly optimize the content of the entire FoV rather than follow the simplified independent construction of Claim 1, its intuition can still help us understand the source of the contrast reduction. For that we revisit Fig. 3 demonstrating masks m^k with increasing bin sizes and their Fourier transforms M^k. For narrow bins, M^k is a narrow impulse plus a wide noise tail over the entire spectrum. However, as we increase the bin size, M^k becomes narrower and approaches a low pass filter. As a consequence, as wide bins result in less noise over the spectrum, the holograms they produce in Fig. 5 also have better contrast.

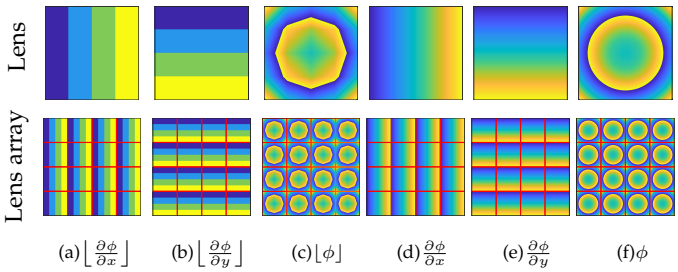


Fig. 6: Phase of lens and lens array. Top row: (a-b) The slope assignment of Claim 2, each sub-square on the SLM area displays a different slope, hence it generates content at a different sub-square out of the desired angular range. The integral of these slopes leads to the phase surface in (c). (d-e) A continuous approximation to the slopes of (a-b) with linearly varying slopes, whose integral is a quadratic phase mask of the form $\phi(\vec{\boldsymbol{x}}) = \alpha \|\vec{\boldsymbol{x}}\|^2$, displayed in (f). Lower row: lens-array, where each lenslet unit is constructed as in the top row.

Note that all examples in this paper are normalized to have maximal intensity 1, for results with lower contrast this usually reveals more background noise.

The random masks of [7] provide results similar to what we get with our smallest bins. We use the online implementation to show some results in App. Fig. 14. We have selectively chosen test images with a high dynamic range to better emphasize the problem. Contrast reduction is less noticeable in images of modest dynamic range, as the ones used in the original paper.

4.2.2 Optimality

We show that if one is interested in trading étendue for resolution there is a simple strategy for the fixed phase mask construction leading to near-optimal results.

Classically, spatial resolution in an imaging system is determined by the size of the aperture used in the pupil plane of the lens; the larger the pupil aperture, the higher the resolution of the image. Thus we can select the Fourier transform of the low-pass filter g in the hologram generation cost of Eq. (9) as a window function setting to zero any content on the SLM plane outside a square of width $\Omega_x/\boldsymbol{q_\mathcal{E}}$. The smaller area on the SLM naturally leads to a reduction in the resolution that is equal to the amount by which we expand the étendue, matching the pitch of Eq. (17).

Using this basic principle, we can formulate an alternate strategy for the design of the masks m^k associated with different linear tilts. We segment the SLM to $K = \boldsymbol{q}_\mathcal{E}^2$ non-overlapping sub-squares and assign each a different linear tilt. Hence each sub-square of the SLM generates content at a different sub-square of the Fourier transform, namely in a different sub-square of the target image. This slope assignment is visualized in the top row of Fig. 6(a-b), and summarized in the Claim below.

Claim 2. Relying on the above suggested mask we construct a minimizer for Eq. (9) with a resolution $\widehat{\Delta}_\theta = \boldsymbol{q_\mathcal{E}}\Delta_\theta$ and a contrast loss not worse than $\times 2$. The minimizer uses a piecewise linear slope mask dividing the SLM area into $K = \boldsymbol{q}_\mathcal{E}^2$ squares of size $\Omega_x/\boldsymbol{q_\mathcal{E}} \times \Omega_x/\boldsymbol{q_\mathcal{E}}$ each, and assigning them 2D slopes out of the discrete set $\vec{\boldsymbol{s}} = (k_1, k_2)\Omega_\theta$, for integer k_1, k_2 values, $|k_1| \leq \boldsymbol{q_\mathcal{E}}/2$,$|k_2| \leq \boldsymbol{q_\mathcal{E}}/2$.

The above construction assumes we solve K different optimization problems for each of the target sub-images I^k. This leads to a $2\times$ contrast loss because replicas caused by the piecewise constant phase modulation cause some percentage of the hologram energy to spread outside the target sub-square. In practice, by simultaneously optimizing the SLM content in all sub-squares together to minimize Eq. (9) we can largely improve on that. Replicas can be further reduced by considering a continuous approximation to the discrete slopes, and setting $\vec{\boldsymbol{s}}(\lfloor \vec{\boldsymbol{x}} \rfloor) = \frac{\lfloor \vec{\boldsymbol{x}} \rfloor \widehat{\Omega}_\theta}{\Omega_x}$ (Fig. 6(d-e)). The integral of this slope assignment is a quadratic phase mask, namely this strategy is equivalent to placing a field lens on the SLM plane (Fig. 6(f)) or illuminating it with a spherical wavefront [1].

Recently [8] have used a learning approach to seek a phase mask that would allow the display of high quality holograms. For that, they optimized the shape of the fixed phase mask to maximize the quality of the resulting holograms, using a metric similar to Eq. (9). The authors analyze the properties of the resulting mask showing that it meets the spectrum of natural images. However, the learning approach is subject to local minima and the result is not fully understood. In contrast, Claim 2 provides a simple analytic answer to the same question.

The need for 3D scores. As mentioned, the use of low resolution SLM tiling, or equivalently a lens realizing a quadratic phase mask (top row of Fig. 6) provides a near-optimal solution to the cost in Eq. (9). However, it is well known that adding a lens to an optical system does not change its étendue. To see the problem, consider Fig. 1(d) which illustrates a $\times 2$ étendue expansion in flatland. As derived in Claim 2, it splits the SLM region into two equal-sized areas each displaying a fixed tilt. Effectively, content on the right side of the target image is only produced by the left part of the SLM where the corresponding slope is present. This means that the cone of light emerging from each part of the target image was reduced, or effectively, each part of the image is only viewed from half of the eye-box.

We note that covering all positions on the eye-box is less of an issue with a random phase mask [7] which anyway scatters light in all directions. Yet, even in this case, to achieve actual holographic properties such as focus and defocus cues, [7] explicitly optimize the 3D content produced by their hologram rather than 2D scores as in Eq. (9). The need for pupil-aware scores was recently also raised by [32].

As illustrated in the 6th column of Fig. 5, random slope patches of intermediate size allow for reasonable contrast and can serve as a way to balance contrast and eye-box coverage. However, below we derive an even better mask that explicitly accounts for eye-box coverage.

4.3 Scoring 3D content

Previous work on étendue-expansion has optimized the quality of 2D images as in Eq. (9). However, the above discussion suggests that the 2D score does not capture all the desired aspects of a wide-étendue display, in particular the fact that we want a wide coverage of the eye-box and not just a wide FoV. To this end, we consider below two popular optimization costs based on light fields and focal stacks. We show that these costs better score the holographic content, and in particular, they rule out the quadratic phase mask solution.

4.3.1 Light-field targets

We want to optimize the phase of the holographic display such that it generates a target light field rather than a 2D image. While classical hogel-based approaches [16], [17], [18], [19], [20], [21] to light field displays attempt to display the light field at the SLM plane, we aim to generate the light field as the Fourier transform of our phase mask. We parameterize the target light field as $L(\vec{\boldsymbol{\theta}}, \vec{\boldsymbol{y}})$ where $\vec{\boldsymbol{\theta}}$ is a viewing angle and $\vec{\boldsymbol{y}}$ is a position on the eye-box. As $\vec{\boldsymbol{\theta}}$ corresponds to the 2D dimension of the image we display to the viewer, we usually want to sample it with a much higher resolution than the one we use for $\vec{\boldsymbol{y}}$. When defining the target light field we again want to address the fact that we have a fixed number of freedom degrees in our SLM. Thus, if we aim to see $\mathcal{N}_{eb} \times \mathcal{N}_{eb}$ different eye-box bins (corresponding to different view points), we reduce the resolution in the FoV dimension, and target a wider pitch $\Delta_\theta^{lf} = \mathcal{N}_{eb}\widehat{\Delta}_\theta$ (see Eq. (17)).

To express the relationship between the SLM phase $\phi(\vec{\boldsymbol{x}})$ and the light-field, let us define by $w_{\vec{\boldsymbol{y}}}(\vec{\boldsymbol{x}})$ a box window of width $\Omega_x^{lf} = \Omega_x/\mathcal{N}_{eb}$ centered around $\vec{\boldsymbol{y}}$, where $\vec{\boldsymbol{y}}$ denotes a position on the SLM plane, or equivalently a position on the

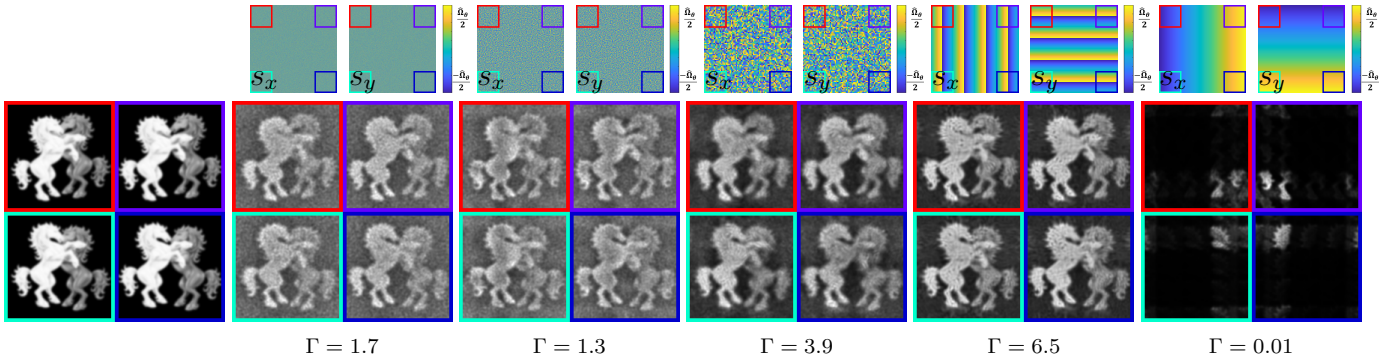


Fig. 7: Light field optimization: We simulate and demonstrate 4 view positions out of a 4×4 light-field, the presented view points (eye-box positions) correspond to the color frames marked on the mask planes at the top. Light field optimization targets image resolution $\Delta_\theta^{lf} = 4\widehat{\Delta}_\theta = 16\Delta_\theta$. (a) Target. (b-f) Piecewise constant SLM content is optimized subject to a fixed phase mask of a smaller pitch. The s_x, s_y phase mask slopes are visualized in color above the images. Numerical contrast evaluation (averaged over all viewpoints) is reported under each image. (b) Random phase mask by [7]. (c-d) Random piecewise linear slope assignment with different bin size. Wider bins achieve better contrast. (e) Optimal slope assignment equivalent to a lenslet array. (f) The optimal slope for 2D targets derived in Claim 2, here regions of consistent slope are too wide hence in each light-field view the SLM sends light only to a sub-region out of the target range.

eye-box plane as the eye-box is an image of the SLM plane, see Fig. 8. Thus, position $\vec{\boldsymbol{y}}$ of the light-field is expressed via $\mathcal{F}(w_{\vec{\boldsymbol{y}}}(\vec{\boldsymbol{x}}) \cdot exp(i\phi(\vec{\boldsymbol{x}}))$, that is, by setting the SLM content to zero anywhere outside the box $w_{\vec{\boldsymbol{y}}}$ and computing a Fourier transform.

With this goal in mind, we optimize for a phase mask minimizing

$$\min_{\phi_c} \sum_{\vec{\boldsymbol{y}}} \left\| g_{\Delta_\theta^{lf}} * \left| \mathcal{F}\left(w_{\vec{\boldsymbol{y}}}(\vec{\boldsymbol{x}}) \cdot e^{i(\phi_c(\vec{\boldsymbol{x}}) + \phi_m(\vec{\boldsymbol{x}}))} \right) \right|^2 - g_{\Delta_\theta^{lf}} * L(\vec{\boldsymbol{\theta}}, \vec{\boldsymbol{y}}) \right\|^2 . \tag{18}$$

We over-sample $2\mathcal{N}_{eb} \times 2\mathcal{N}_{eb}$ positions of the central $\vec{\boldsymbol{y}}$ viewpoint so that the windows $w_{\vec{\boldsymbol{y}}}$ overlap. To allow for variation in the viewer's pupil diameter we can also sum in the cost window functions $w_{\vec{\boldsymbol{y}}}$ of multiple supports, as in [32].

In Fig. 7 we expand étendue by a factor $\boldsymbol{q_\mathcal{E}} = 4$ in both axes and target $\mathcal{N}_{eb} = 4$ viewpoints (eye-box bins). We show 2×2 out of 4×4 eye-box position views in a light-field with pitch $\Delta_\theta^{lf} = 4\widehat{\Delta}_\theta = 16\Delta_\theta$. Using very small patches or the fully random phase mask of [7], we generate the desired light-field at a lower contrast. Bigger patches achieve much higher contrast.

In Fig. 7(f) we demonstrate the weakness of the seemingly optimal solution of Claim 2, corresponding to one big lens spanning slopes linearly over the SLM area. This mask can generate only a small region out of the desired target, because each eye-box window $w_{\vec{\boldsymbol{y}}}$ contains a very limited range of slopes, and these slopes map all SLM content into one sub-square of the desired range, as illustrated in Fig. 1(d).

Optimal slope mask for light-field display. Beyond random slopes, given a target resolution for our light-field we can derive an *optimal* setting for a slope pattern. For that, we divide the SLM area into $\mathcal{N}_{eb} \times \mathcal{N}_{eb}$ squares of size $\Omega_x^{lf} \times \Omega_x^{lf}$ each, where $\Omega_x^{lf} = \Omega_x / \mathcal{N}_{eb}$ is the full SLM width divided by the number of eye-box bins. A viewer moving along the eye-box can see squares of this size, so we need to ensure

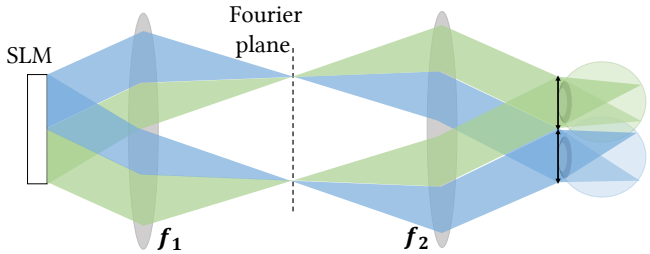
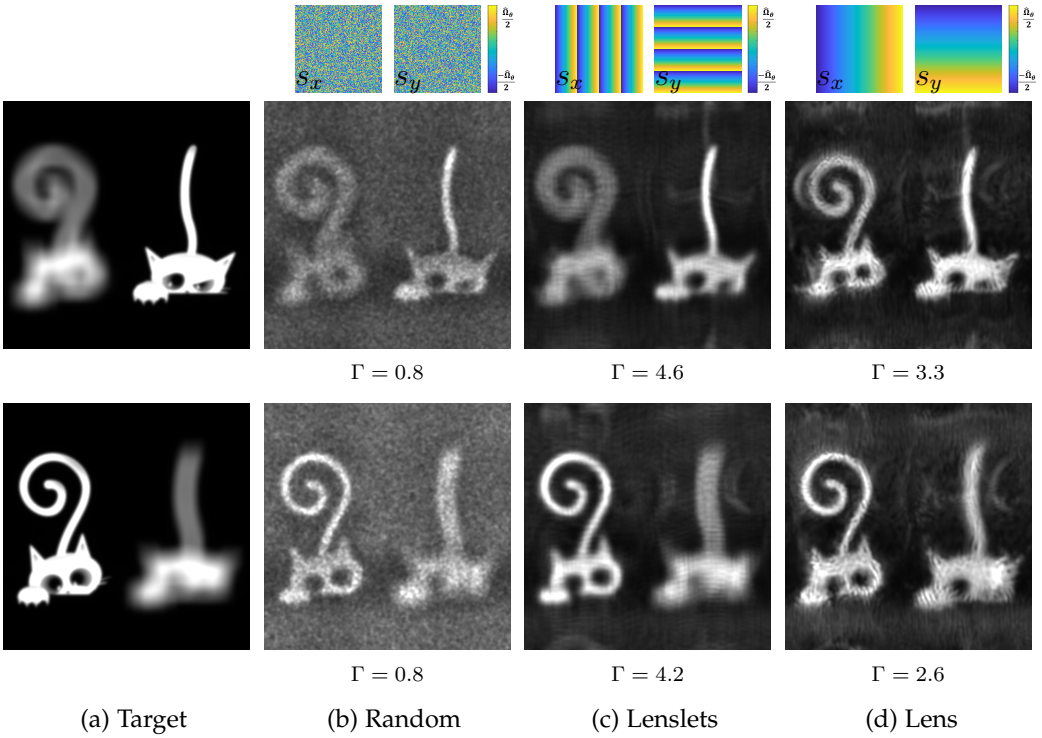


Fig. 8: Light-field viewing: If the eye pupil is smaller than the supported eye-box, the pupil moving across the eye-box filters a subset of the SLM area. The figure uses two different colors to demonstrate two such eye-box positions. To support the full field of view, each sub-square on the SLM should include all tilt directions.

that each such square contains all slopes so it can send light to the extended $\widehat{\Omega}_\theta \times \widehat{\Omega}_\theta$ FoV range, but we will use the biggest slope patches we can *within* each $\Omega_x^{lf} \times \Omega_x^{lf}$ block. That is, we subdivide each square into $\boldsymbol{q_\mathcal{E}} \times \boldsymbol{q_\mathcal{E}}$ sub-squares and assign to each of them a different slope, as illustrated in Figs. 6 and 8.

Integrating this slope assignment is approximately a quadratic phase function in each block, see Fig. 6(f). Thus, this solution can be realized by attaching to the SLM a lenslet array, where each lenslet unit covers a square of size $\Omega_x^{lf} \times \Omega_x^{lf}$. Unlike classical light-field displays, here we use the SLM to generate the Fourier transform of the light-field and not the light-field directly. Thus the number of lenslet units we need is much lower, as it is defined by the desired number of eye-box bins $\mathcal{N}_{eb} \times \mathcal{N}_{eb}$ rather than by the number of pixels in the image we attempt to display. In Fig. 7(e) we compare the holograms produced by a lenslet array to other phase masks, demonstrating the higher quality light fields it can produce.

It should be emphasized that this solution is optimal



(a) Target (b) Random (c) Lenslets (d) Lens

Fig. 9: Focal stack optimization: We simulate a target scene which contains objects at two different planes and we demonstrate two images as will be viewed by an observer focusing his eye at these two planes. We consider three phase masks, a random slope assignment, a lenslet array, and a single quadratic lens. The lenslets lead to higher contrast and best results, as also indicated by the numerical contrast metric Γ under each image. A lens mask cannot generate a sharp image at one focal setting and a defocus one at another focal setting, as in each spatial position only a narrower subset of the angular cone is available.

only for a given light-field resolution Δ_θ^{lf}, since partitioning the SLM area into small periodic blocks does not allow for the viewing of the display at higher resolutions.

4.3.2 *Focal stack target*

As light fields are inherently coupled with angular and spatial resolution trade-offs, previous work [13], [14], [15] also optimized the SLM content to generate a focal stack. That is, the wavefront displayed by the SLM should display 2D images to a viewer's eye at a known position, but should allow the viewer to see realistic defocus blur as he focuses his eye at different depths. Focal stacks pose fewer constraints on the SLM phase compared to light-fields, and in principle, should support targets at a higher resolution. However, focal stacks are only optimized to display focal cues to a given pupil position, and may be sensitive to small pose variation of the pupil inside the eye-box. Despite this, focal stacks capture some of the 3D cues of the scene and we show below that as such, they already require not trivial DOE phase masks, and cannot be well displayed by the naive quadratic lens phase mask derived in Claim 2.

In App. A.2 we detail the content matching cost we use, which follows classical Frensel propagation principles exploited in previous work.

Fig. 9 visualizes results for two scenes. In this demonstration we target a resolution $\widehat{\Delta}_\theta = 4\Delta_\theta$, in a 4×4 étendue-extended display. Since we have constraints at multiple depth planes, in this particular example the number of constraints is larger than the number of freedom degrees and the result is not speckle free.

We compare three slope assignments, random slopes with small independent bins, the lenslet array solution mentioned above, and the naive single-lens solution. The lenslet array achieves better contrast and fewer speckle artifacts compared to random masks. We note that the lenslet array mask used in this simulation is designed for a lower resolution pitch Δ_θ^{lf} which is $\times 4$ wider than the $\widehat{\Delta}_\theta$ pitch we use here. Still, it leads to visually plausible results.

The single lens solution leads to poor results. As discussed above, when considering the entire FoV this phase mask spreads light over a wide angular cone, and gets light to the entire eye-box. However, in each spatial position of the image only a narrow sub-cone is available. As a result, effectively the depth of field is larger than intended and we couldn't generate an image which is sharp at one focal plane and blurred as the viewer focuses at a slightly different depth. In Fig. 9(d) the image is almost equally sharp at both depth planes. For an additional example and explanation on the effects of narrow sub-cones see App. A.2.

5 Discussion

In this work, we have considered extending the étendue of an SLM by attaching to it a DOE phase mask of a higher resolution. We attempt to understand what trade-offs are involved with this expansion and how to design the shape of the DOE to support higher quality holograms.

As the number of degrees of freedom in the display does not increase by the addition of a static DOE, étendue

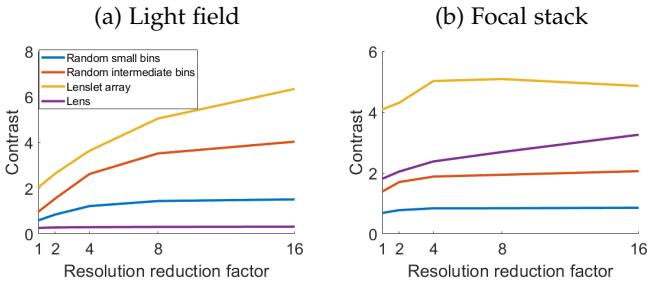


Fig. 10: Comparison of contrast vs reduction in resolution. (a) For the light field target of Fig. 7. (b) For focal stack target of Fig. 9. We used a low-pass filter g of width $q \times q$ with different q values, and plot contrast as a function of the resolution reduction factor q. Proposed lenslet outperforms random mask for all target resolutions, but the contrast gain increases as resolution reduces.

expansion must come with the loss of some image quality. We have explored two strategies for the design of a DOE and the SLM content and show they result in a loss of contrast or resolution which scales precisely with the étendue expansion factor. We explain that random masks can increase étendue, at the cost of contrast reduction. On the other hand, if one wishes to trade étendue for resolution, a simple quadratic phase mask corresponding to a field lens produces near-optimal results.

The optimality of a lens mask stands in contrast to the actual goal of a wide étendue hologram, and we conclude that previous metrics attempting to optimize the quality of the 2D images produced by the display are inadequate in capturing the full wealth of the holographic information. Thus, we suggest optimizing costs explicitly encoding 3D information such as the light field or focal stacks produced by the hologram. We show that, perhaps surprisingly, under these metrics commonly available phase masks such as lenslet arrays can support high quality étendue expansion, outperforming random phase masks.

As another way to evaluate the trade-offs between contrast and resolution, we plot in Fig. 10 the contrast metric Γ of Eq. (16) for the different masks analyzed in this paper, while varying the width of the low pass filter g. In agreement with the analysis, for both focal-stack and light field targets the best results are obtained by our proposed lenslet. The contrast gain increases as resolution reduces.

We hope this analysis will guide the design of future étendue-expansion systems, and in [29] we take a first step demonstrating some of these principles in a physical prototype.

Acknowledgments

We thank Marina Alterman, Rick Chang and Wei-Yu Chen for their valuable feedback on the manuscript. This research was supported by H2020 European Research Council (635537); Israel Science Foundation (1947/20); United States-Israel Binational Science Foundation (2008123/2019758); US National Science Foundation (2008464).

References

[1] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," vol. 36, no. 4, pp. 85:1–85:16, Jul. 2017.

[2] J.-H. Park, "Recent progress in computer-generated holography for three-dimensional scenes," *Journal of Information Display*, vol. 18, no. 1, pp. 1–12, 2017.

[3] S. Benton, "Experiments in holographic video imaging," in *SPIE*, vol. 08, 1991, pp. 247–267.

[4] P. St-Hilaire, S. A. Benton, M. Lucente, and P. Hubel, "Color images with the MIT holographic video display," pp. 73–84, 1992. [Online]. Available: http://dx.doi.org/10.1117/12.59642

[5] S.-C. Kim, J.-W. Moon, D.-H. Lee, K.-C. Son, and E.-S. Kim, "Holographic full-color 3D display system using color-LCoS spatial light modulator," pp. 223–233, 2005. [Online]. Available: http://dx.doi.org/10.1117/12.594455

[6] D. E. Smalley, Q. Y. J. Smithwick, V. M. Bove, J. Barabas, and S. Jolly, "Anisotropic leaky-mode modulator for holographic video displays," *Nature*, no. 7454, pp. 313–317, 2013. [Online]. Available: http://www.nature.com/nature/journal/v498/n7454/full/nature12217.html

[7] G. Kuo, L. Waller, R. Ng, and A. Maimone, "High resolution étendue expansion for holographic displays," *ACM Trans. Graph.*, vol. 39, no. 4, jul 2020.

[8] S.-H. Baek, E. Tseng, A. Maimone, N. Matsuda, G. Kuo, Q. Fu, W. Heidrich, D. Lanman, and F. Heide, "Neural Étendue expander for ultra-wide-angle high-fidelity holographic display," 09 2021.

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

[10] H. Yu, K. Lee, J. Park, and Y. Park, "Ultrahigh-definition dynamic 3d holographic display by active control of volume speckle fields," *Nature Photonics*, vol. 11, 01 2017.

[11] J. Park, K. Lee, and Y. Park, "Ultrathin wide-angle large-area digital 3d holographic display using a non-periodic photon sieve," *Nature Communications*, vol. 10, 03 2019.

[12] G. Koulieris, K. Akşit, M. Stengel, R. Mantiuk, K. Mania, and C. Richardt, "Near-eye display and tracking technologies for virtual and augmented reality," *Computer Graphics Forum*, vol. 38, no. 2, pp. 493–519, Jun. 2019.

[13] S. Choi, M. Gopakumar, Y. Peng, J. Kim, and G. Wetzstein, "Neural 3d holography: Learning accurate wave propagation models for 3d holographic virtual and augmented reality displays," *ACM Trans. Graph. (SIGGRAPH Asia)*, 2021.

[14] J. Zhang, N. Pegard, Z. Jingshan, H. Adesnik, and L. Waller, "3d computer-generated holography by non-convex optimization," *Optica*, vol. 4, p. 1306, 10 2017.

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

[16] D. Glasner, T. Zickler, and A. Levin, "A reflectance display," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 61:1–61:12, Jul. 2014.

[17] N. Padmanaban, Y. Peng, and G. Wetzstein, "Holographic Near-Eye Displays Based on Overlap-Add Stereograms," *ACM Trans. Graph. (SIGGRAPH Asia)*, no. 6, 2019.

[18] T. Yatagai, "Stereoscopic approach to 3-d display using computer-generated holograms," *Appl. Opt.*, vol. 15, no. 11, pp. 2722–2729, Nov 1976.

[19] M. Yamaguchi, H. Hoshino, T. Honda, and N. Ohyama, "Phase-added stereogram: calculation of hologram using computer graphics technique," *Proceedings of SPIE*, vol. 1914, pp. 25–31, 09 1993.

[20] H. Kang, T. Yamaguchi, and H. Yoshikawa, "Accurate phase-added stereogram to improve the coherent stereogram," *Applied optics*, vol. 47, pp. D44–54, 08 2008.

[21] L. Shi, F.-C. Huang, W. Lopes, W. Matusik, and D. Luebke, "Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3d computer graphics," *ACM Trans. Graph.*, vol. 36, no. 6, nov 2017.

[22] R. Haeussler, S. Reichelt, N. Leister, E. Zschau, R. Missbach, and A. Schwerdtner, "Large real-time holographic displays: from prototypes to a consumer product," *Proc. SPIE*, vol. 7237, 02 2009.

[23] C. Jang, K. Bang, G. Li, and B. Lee, "Holographic near-eye display with expanded eye-box," *ACM Trans. Graph.*, vol. 37, no. 6, dec 2018.

[24] J. Kim *et al.*, "Foveated ar: Dynamically-foveated augmented reality display," *ACM Trans. Graph.*, vol. 38, no. 4, jul 2019.

[25] X. Xia, Y. Guan, A. State, P. Chakravarthula, T.-J. Cham, and H. Fuchs, "Towards eyeglass-style holographic near-eye displays with statically expanded eyebox," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020, pp. 312–319.
[26] M. Kim, S. Lim, G. Choi, Y. Kim, H. Kim, and J. Hahn, "Expanded exit-pupil holographic head-mounted display with high-speed digital micromirror device," *ETRI Journal*, vol. 40, 05 2018.
[27] M.-H. Choi, Y.-G. Ju, and J.-H. Park, "Holographic near-eye display with continuously expanded eyebox using two-dimensional replication and angular spectrum wrapping," *Optics Express*, vol. 28, pp. 533–547, 01 2020.
[28] S. Choi, M. Gopakumar, Y. Peng, J. Kim, M. O'Toole, and G. Wetzstein, "Time-multiplexed neural holography: A flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators," in *Proceedings of the ACM SIGGRAPH*, 2022, pp. 1–8.
[29] A. C. S. Sagi Monin and A. Levin, "Exponentially-wide Étendue displays using a tilting cascade," *ICCP*, 2022.
[30] B. Lee, D. Yoo, J. Jeong, S. Lee, D. Lee, and B. Lee, "Wide-angle speckleless dmd holographic display using structured illumination with temporal multiplexing," *Opt. Lett.*, vol. 45, no. 8, pp. 2148–2151, Apr 2020.
[31] T. A. Bartlett, W. C. McDonald, and J. N. Hall, "Adapting Texas Instruments DLP technology to demonstrate a phase spatial light modulator," in *Emerging Digital Micromirror Device Based Systems and Applications XI*, vol. 10932. SPIE, 2019, pp. 161 – 173.
[32] P. Chakravarthula, S.-H. Baek, E. Tseng, A. Maimone, G. Kuo, F. Schiffers, N. Matsuda, O. Cossairt, D. Lanman, and F. Heide, "Pupil-aware holography," *arXiv*, 2022.

Sagi Monin Sagi Monin received his B.Sc. and M.Sc. (Summa cum Laude) degrees in Electrical Engineering from the Technion. He is currently a Ph.D. student at department of Electrical Engineering, Technion, Israel, supervised by Prof. Anat Levin.

Aswin C. Sankaranarayanan Aswin C. Sankaranarayanan is a professor in the ECE department at CMU, where he is the PI of the Image Science Lab. His research interests are broadly in compressive sensing, computational photography, signal processing and machine vision. His doctoral research was in the University of Maryland where his dissertation won the distinguished dissertation award from the ECE department in 2009. Aswin is the recipient of the best paper awards at CVPR 2019 and ICCP 2021, the CIT Dean's Early Career Fellowship, the Spira Teaching award, the NSF CAREER award, the Eta Kappa Nu (CMU Chapter) Excellence in Teaching award, and the Herschel Rich Invention award from Rice University

Anat Levin Anat Levin is an Associate Prof. at the department of Electrical Engineering, Technion, Israel, doing research in the field of computational imaging. She received her Ph.D. from the Hebrew University at 2006. During the years 2007- 2009 she was a postdoc at MIT CSAIL, and during 2009-2016 she was an Assistant and Associate Prof. at the department of Computer Science and Applied Math, the Weizmann Inst. of Science.

Analyzing phase masks for wide étendue holographic displays

Supplementary appendix

APPENDIX A

A.1 Spectrum of binary random masks

Below we prove Claim 1 of the main paper.

Claim 1. The combined phase mask $\phi(\vec{x})$ of Eq. (14) produces I at the price of a factor $\boldsymbol{q}_{\boldsymbol{\mathcal{E}}}^2$ contrast reduction.

Proof: Let us denote the Fourier transform of the slope mask as $M^k = \mathcal{F}(m^k)$. As multiplication in the image domain translates to convolution in the frequency domain we get

$$\mathcal{F}\left(e^{i\phi(\vec{x})}\right) = \sum_k M^k * u^k. \tag{19}$$

Since we divide the SLM area into K disjoint masks m^k, each of them is non zero over $1/K$ of the SLM area. Let us assume the random masks m^k are constant over bins of b pixels, where in a fully independent random mask $b = 1$. We show below that their Fourier transforms $M^k = \mathcal{F}(m^k)$ are random variables with expected power

$$|M(\vec{\boldsymbol{\theta}})|^2 = S_c(\vec{\boldsymbol{\theta}}) + S_t(\vec{\boldsymbol{\theta}}), \tag{20}$$

with:

$$S_c(\vec{\boldsymbol{\theta}}) = \frac{1}{K^2\Delta_\theta^2}\operatorname{sinc}^2\left(\frac{\vec{\boldsymbol{\theta}}_x}{\Delta_\theta}\right)\operatorname{sinc}^2\left(\frac{\vec{\boldsymbol{\theta}}_y}{\Delta_\theta}\right), \tag{21}$$

$$S_t(\vec{\boldsymbol{\theta}}) = \frac{b^2}{K\Delta_\theta^2\Omega_\theta^2}\operatorname{sinc}^2\left(\frac{b\vec{\boldsymbol{\theta}}_x}{\Omega_\theta}\right)\operatorname{sinc}^2\left(\frac{b\vec{\boldsymbol{\theta}}_y}{\Omega_\theta}\right). \tag{22}$$

We visualize random masks m^k and their spectrum M^k in Fig. 3, for a few different bin widths. The S_c term of Eq. (20) is a narrow impulse whose width is the angular pitch Δ_θ. In addition we receive a noise-like signal of random values spread over the entire spectrum expressed by the term S_t. For small bins (small b values) the noise is spread over the entire spectrum with low values. When the bin size b is increased, the support over the spectrum is decreased; however, the values of the random noise are increased.

We can compute the total energy of the impulse and noise-like terms

$$\int S_c(\vec{\boldsymbol{\theta}})d\vec{\boldsymbol{\theta}} = \frac{1}{K^2\Delta_\theta^2}, \tag{23}$$

$$\int S_t(\vec{\boldsymbol{\theta}})d\vec{\boldsymbol{\theta}} = \frac{1}{K\Delta_\theta^2}. \tag{24}$$

That is, the overall energy of the noise is a factor K higher than the central impulse. This ratio is independent of the bin size and holds even if noise tail values are low ($b = 1$). Thought, note that for very large bins the tail is narrow, merging with the central impulse. We recall that $|u^k|^2 \approx I^k$. Hence, for small b values, $M^k * u^k$ produces I^k plus a big noise term which is K times higher than the original signal.

By that we have completed the proof, but we are left with the task of computing the expected spectrum of the random masks M^k. To this end we consider random variables z_j corresponding to binary boxes of size $b\Delta_x \times b\Delta_x$ centered at a random position on the SLM plane. We define the random mask m^k as the sum of N independent random variables

$$m^k = \sum_{j=1}^{N} z_j \tag{25}$$

Since m^k is non zero over fraction $1/K$ of the SLM area, the expected number of such variables is

$$N = \frac{\Omega_x^2}{Kb^2\Delta_x^2}, \tag{26}$$

which is also

$$N = \frac{\Omega_\theta^2}{Kb^2\Delta_\theta^2}. \tag{27}$$

We note that this definition ignores the fact that the boxes z_j cannot collide on the same SLM pixels, and this simplification will introduce some bias in the result we derive below.

Let Z_j denote the Fourier transform of z_j. Then the mask spectrum M^k corresponds to $M^k = \sum_j Z_j$. We want to compute the expected spectrum $E[|M^k|^2]$. For that let us define $\mu = E[Z_j]$. Since M^k is the sum of N independent random variables, we can compute its spectrum as

$$E\left[|M^k|^2\right] = NE\left[|Z^j - \mu|^2\right] + N^2|\mu|^2 \tag{28}$$

$$= NE\left[|Z^j|^2\right] + N(N-1)|\mu|^2 \tag{29}$$

$$\approx NE\left[|Z^j|^2\right] + N^2|\mu|^2 \tag{30}$$

Thus to compute the spectrum we need to compute $E[|Z^j|^2]$ and $|\mu|^2$. As z^j is a box function, its spectrum Z^j is a sinc and

$$|Z^j(\vec{\boldsymbol{\theta}})|^2 = \frac{b^4}{\Omega_\theta^4}\operatorname{sinc}^2\left(\frac{b\vec{\boldsymbol{\theta}}_x}{\Omega_\theta}\right)\operatorname{sinc}^2\left(\frac{b\vec{\boldsymbol{\theta}}_y}{\Omega_\theta}\right) \tag{31}$$

Also the average of z_j is a uniform spread over the SLM, and the SLM area is Ω_x^2, thus its Fourier transform is

$$|\mu(\vec{\boldsymbol{\theta}})|^2 = \frac{1}{K^2N^2\Delta_\theta^2}\operatorname{sinc}^2\left(\frac{\vec{\boldsymbol{\theta}}_x}{\Delta_\theta}\right)\operatorname{sinc}^2\left(\frac{\vec{\boldsymbol{\theta}}_y}{\Delta_\theta}\right) \tag{32}$$

Substituting Eqs. (31) and (32) in Eq. (28) we get the desired result:

$$|M(\vec{\boldsymbol{\theta}})|^2 = S_c(\vec{\boldsymbol{\theta}}) + S_t(\vec{\boldsymbol{\theta}}), \tag{33}$$

with,

$$S_c(\vec{\boldsymbol{\theta}}) = \frac{1}{K^2\Delta_\theta^2}\operatorname{sinc}^2\left(\frac{\vec{\boldsymbol{\theta}}_x}{\Delta_\theta}\right)\operatorname{sinc}^2\left(\frac{\vec{\boldsymbol{\theta}}_y}{\Delta_\theta}\right), \tag{34}$$

$$S_t(\vec{\boldsymbol{\theta}}) = \frac{b^2}{K\Delta_\theta^2\Omega_\theta^2}\operatorname{sinc}^2\left(\frac{b\vec{\boldsymbol{\theta}}_x}{\Omega_\theta}\right)\operatorname{sinc}^2\left(\frac{b\vec{\boldsymbol{\theta}}_y}{\Omega_\theta}\right). \tag{35}$$

■

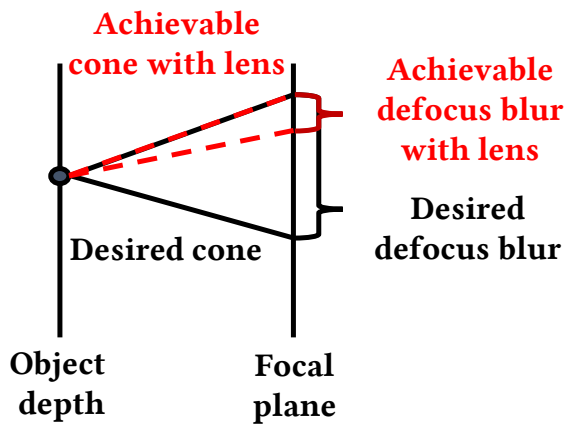


Fig. 11: Undesired cone effects: When generating a focal stack we aim to achieve a cone that mimics the correct defocus blur depending on the objects distance. However, when we use a lens as phase mask we are only able to achieve part of the desired cone, resulting in a narrower, shifted blur image.

A.2 Focal stack optimization

To optimize SLM content to generate a focal stack target we follow the Frensel propagation models used in previous work. The image I_z that an SLM produces at focal depth z (which is now different than the Fourier plane of Fig. 1) can be computed using the Frensel propagation operator:

$$\mathcal{F}\left(h_z(\vec{\boldsymbol{x}})e^{i\phi(\vec{\boldsymbol{x}})}\right), \tag{36}$$

with

$$h_z(\vec{\boldsymbol{x}}) = e^{\frac{2\pi i z}{\lambda}} e^{i\pi\lambda z\|\vec{\boldsymbol{x}}\|^2}. \tag{37}$$

With this goal in mind, we search for SLM content minimizing the error with a target focal stack

$$\underset{\phi}{\operatorname{argmin}} \sum_z \left\| g * \left|\mathcal{F}\left(h_z(\vec{\boldsymbol{x}}) \cdot e^{i(\phi_c(\vec{\boldsymbol{x}})+\phi_m(\vec{\boldsymbol{x}}))}\right)\right|^2 - g * I_z \right\|^2. \tag{38}$$

Fig. 9 in the main paper visualizes the outcome of this optimization.

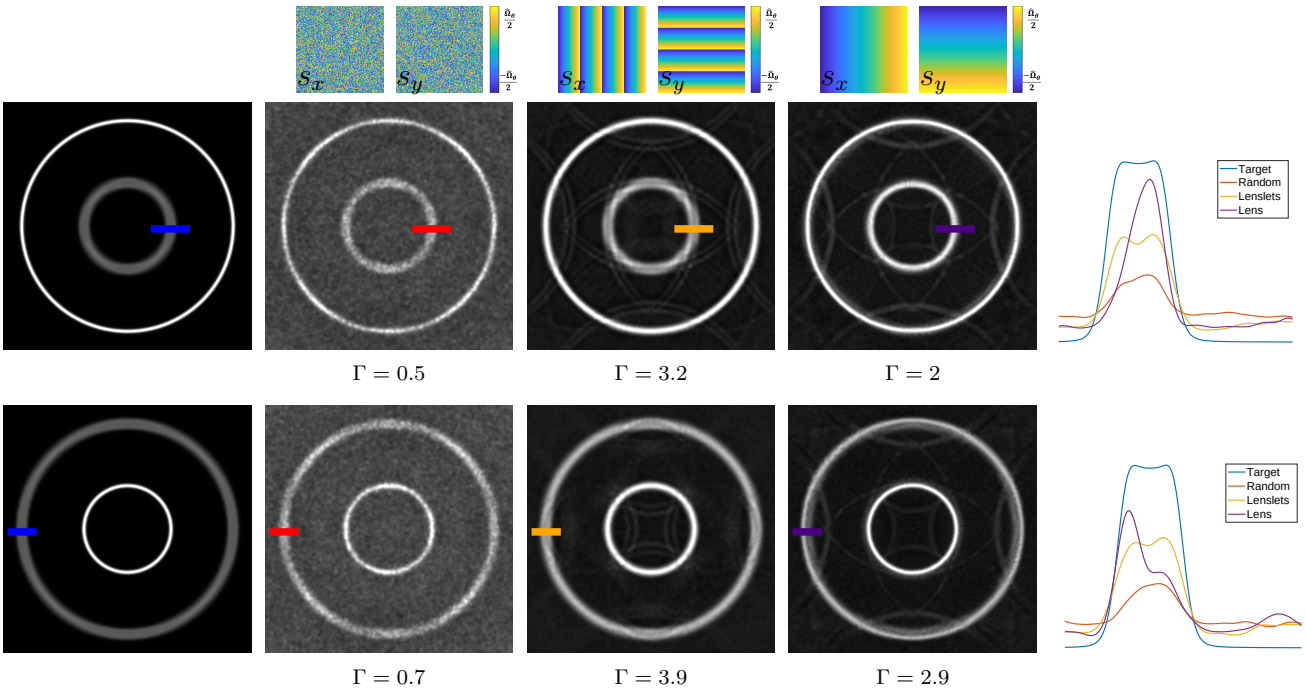
Fig. 11 illustrates the problem of using a lens array as a DOE to increase étendue with a focal-stack target. In each spatial position the lens DOE only allows a narow sub-cone out of the full desired angular range. Since there is a narrow cone of angles emerging from a point, when this cone intersects a different focal depth the defocus blur it generates is narrower than the desired one. If the sub-cone is tilted, the defocus is also shifted. In Fig. 12 we demonstrate this problem with a simple scenario with two circles at two different depths. As seen in the figure, the defocus blur produced by the lens DOE is narrower than the target, and the image is shifted in dependence of the location of the object in the image. This is a result of the direction of the cone which depends on its origin on the lens.

A.3 Additional results

In Fig. 13 we demonstrate another synthetic example from Sec. 4.2 comparing a 2D hologram with different phase masks at different étendue-expansion factors, going from 2×2 to 4×4, and even to 8×8. In agreement with Claim 1, the contrast reduction of the fully random masks becomes worse for higher expansion factors. Random masks with medium patch sizes or the optimal quadratic lens generate images with better contrast.

In Fig. 14 we simulate étendue expansion by a factor 16 (4×4), for the fully random masks used in previous work [7] using the online implementation[2]. From these results we can see that when the image has a wide dynamic range the loss of contrast is more significant. For images whose original dynamic range is limited (as the ones included in the original paper), contrast reduction is less of an issue.

2. https://github.com/dongheon-yoo/Holographic-Display-etendue-Expansion



(a) Target (b) Random (c) Lenslets (d) Lens (e) Cross-section

Fig. 12: Focal stack optimization: Target scene contain two circles at different planes and we demonstrate two images as will be viewed by an observer focusing his eye at these two planes. We consider three phase masks, a random slope assignment, a lenslet array, and a single quadratic lens. The lenslets lead to higher contrast and best results, as also indicated by the numerical contrast metric Γ under each image. A lens mask cannot generate a sharp image at one focal setting and a defocus one at another focal setting, as in each spatial position only a narrower subset of the angular cone is available.

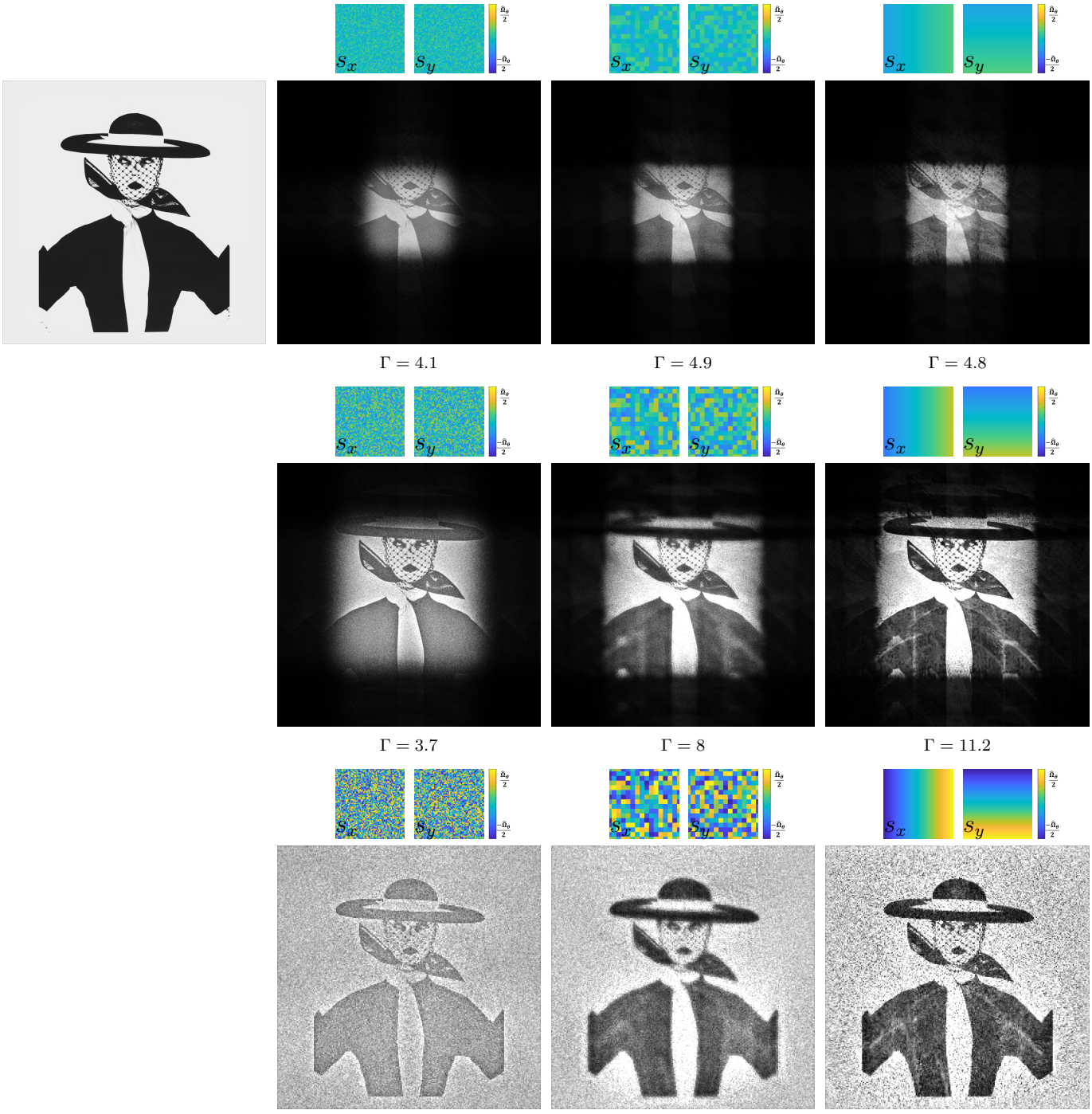


Fig. 13: Comparing masks for 2D étendue expansion: We visualize étendue expansion by increasing factors 2×2, 4×4 and 8×8. As the expansion factor increases the contrast loss of the random mask is larger, in agreement with the prediction of Claim 1.

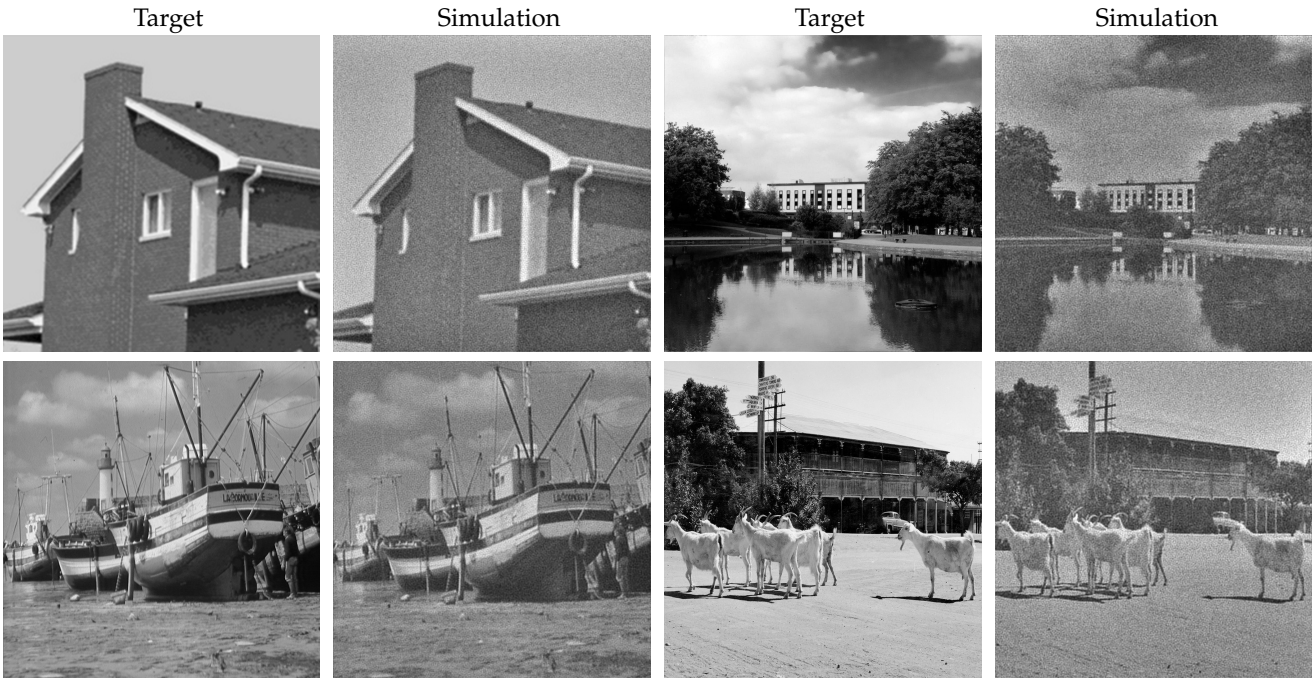


Fig. 14: We present results following [7] implementation, with Étendue expansion of 16 (4×4). We show that for images with limited dynamic range the reduced contrast is not as significant; however, when the dynamic range is higher (goats and river) the contrast reduction is more noticeable.