1 MATRIX FACTORIZATION DETAILS

In the following we describe in detail the individual steps of the multiplexing method. We focus on two DOEs, with a rank 1, but extensions to a larger number of DOEs are possible with tensor factorization, and dynamic DOEs with spatial light modulators can be modeled as higher rank matrices and tensors [Heide et al. 2014].

Newton’s update for solving sub-problems. As has been briefly discussed in the main text, for rank \( r = 1 \), our complex matrix factorization problem simplifies to vector form, illustrated as:

\[
\begin{align*}
\text{argmin}_{a,b} & \quad \frac{1}{2} \| T - W \circ ab \|_W^2 \\
= & \quad \text{argmin}_b \frac{1}{2} \| \text{diag}(W) t - \text{diag}(W) O_a b \|_2^2 \\
= & \quad \text{argmin}_b \frac{1}{2} f(b),
\end{align*}
\]

and

\[
f(b) = t^\dagger \text{diag}(W) t - 2t^\dagger \text{diag}(W) O_a b + b^\dagger O_a^\dagger \text{diag}(W) O_a b.
\]

In theory, it is reasonable to force the extracted phase component of every pixel to be unified for the amplitude all the time. While in practice, it is not necessary to normalize it at every step since we only need the final optimum to fit the physical model. Thus, during the optimization, our conditional projection (step 4 in Alg. 1) maintains as much of the data as possible while eliminating ill-conditioned cases, leading to not just faster but also better convergence. Once we approach convergence, the complex phase expression can then be forced to have unit amplitude.

2 FABRICATION DETAILS

As has been mentioned in the main text, the DOEs are fabricated by repeatedly applying photolithography and reactive ion etch (RIE) technology. Each round of fabrication results in a binary profile on the fused silica substrate. By repeating the process for \( n \) times, a \( 2^n \)-level phase surface can be fabricated finally.
In our work, the basic fabrication procedure is shown in Fig. 1. Four iterations are chosen to produce 16-level phase surface, which is capable of providing sufficiently high diffraction efficiency, as shown in Fig. 2. The principle wavelength is $\lambda = 550$ nm, and the corresponding total etching depth is 1195 nm. A final depth error of $\pm 10$ nm has been achieved. The microscope images of one of the fabricated DOE plates are shown in Fig. 3.

3 ADDITIONAL RESULTS

Analysis on air gap issue. To derive the encoding model, we make an assumption that the two DOE layers are in direct physical contact to enforce the factorization towards a well-conditioned optimization problem. In the following we analyze the air gap issue that is somehow inevitable in practice, especially where relative motion multiplexing is enclosed.

As shown in Fig. 4, the left one indicates a multiplexing configuration that the size of air gap is much smaller than a feature pixel so as to be reasonably neglected (e.g., static pairing scenarios). Further, the right one has an air gap whose size could be several times larger than that of a feature pixel. Based on Huygens–Fresnel diffraction principle, the wave emitted from a pixel on layer $A$ is spread out a bit when arriving layer $B$. Consequently, the synthesized phase modulation for this pixel is no longer one-to-one mapping, but to involve additional phase delays for neighbouring pixels (illustrated by different color wavefronts in Fig. 4). The two layer patterns encoded under the closely stacked assumption may fail. Fortunately, we can still derive a constrained range where the principle diffractive "rays" still travel through the target pixel. Then the encoded image can still be decoded while suffering from diffraction efficiency loss of some sense.

We reasonably consider a square aperture's diffraction (i.e. each cell in the DOEs is a square). The intensity of the diffraction of a square of layer $A$ on layer $B$ is mainly distributed in two orthogonal
directions [Goodman 2005], and can be expressed as:
\[ I(x, y) = I_0(x, y) \cdot \text{sinc}^2\left(\frac{k_0x}{2z}\right) \cdot \text{sinc}^2\left(\frac{k_0y}{2z}\right) , \]  
(7)

where \((x, y)\) is the coordinate of diffraction pattern, \(k = \frac{2\pi}{\lambda}\) is the wave number, \(\lambda\) is the principal wavelength of incident illumination, \(\omega\) represents the feature size (diameter) of this pixel, and \(z\) is the size of air gap.

Now we want to limit the wavefront spread of layer A within one pixel’s range on layer B by applying the condition that the half width of central order diffraction is with one pixel’s size, such that we derive the constraint as:
\[ z \leq \frac{\omega^2}{0.886\lambda}. \]  
(8)

Accordingly, we note that it won’t result in noticeable image degradation when the air gap size falls below 131 \(\mu m\) in our implementations (where the feature size 8\(\mu m\), wavelength 550nm). This value is feasible in engineering even for relative motion setups.

At current stage of proof-of-concept, we reasonably rely on the closely stacked assumption. Experimentally, we find that although involving a small air gap around the thickness of two regular A4 papers (i.e. 200\(\mu m\)), the target image can still be decoded with a visually pleasing quality except for a slight loss of diffraction efficiency. Fig. 5 shows the synthetic images reconstructed from the multiple discrete pairings (second application) with different thickness settings. We found that when the air gap is set 131\(\mu m\) as calculated above, the image exhibits almost the same quality as the one without any air gap. When increasing the air gap up to 262 \(\mu m\), even to 500\(\mu m\), the images still exhibit visually pleasing quality, except for the occurrence of speckle noise and the diffraction efficiency loss. Consequently, we can release the constraint on Eq. 8 up to 2 physical pixels, then the tolerance of air gap size for a 8\(\mu m\) feature size’s design is 262\(\mu m\).

Nevertheless, this is an engineering issue that can be fixed with a modest amount of work.

However, to enforce a more robust image reconstruction, the extra phase modulation derived from noticeable air gap shall be considered. Consequently, the model \(T = AB^\dagger\) presented in Section 4 of the main text now becomes \(T = (A \cdot F_{c,r,y})B^\dagger\). Notice that the left term involves a spatial variant convolution corresponding to delays caused by the occurrence of air gap. Thus, the proposed factorization updates shall involve additional convolution and deconvolution steps, which may yield an ill-conditioned optimization. Notice that in phase modulation domain, the inverse formation model becomes more complex. This extension for arbitrary geometric alignment setting remains an open question and deserves further study.

Simulated results approximating pixelated misalignment. We quantitatively evaluate the perceptual quality of a reconstructed image using the offset pairing design as detailed in Section 5 of the main text. In the following we present the corresponding full resolution images (Fig. 7).

Full resolution experimental results of offset pairing scenarios. We have fabricated two pairs of offset layering DOEs that are encoded with two different gray-scale images, one “Einstein” and the other “Lena”. Partial results have been presented in Fig.6 of main text.
Fig. 5. Comparisons of synthetic images reconstructed from the 3 by 3 pairing examples subject to different air gap settings: 0 (left column), 131µm (center left column), 262µm (center right column), and 500µm (right column), respectively. The SSIMs are evaluated with respect to the image reconstructed without air gap.

Fig. 6. The synthetic comparison results of reference (ideal fabrication), 5% fabrication error, and 10% fabrication error, respectively.

Multiplexing is constant once the physical specification of DOEs are determined. Notice that the labels on bottom-left of each sub-figure indicates the frame number. At current stage, how to evaluate the maximum data amount that can be multiplexed subject to given hardware data capacity is still unclear. We yield this to future work.

One last thing, the frames are photographed after manually translating the top DOE layer to each target position. We further apply a small translation stage to record a video stream indicating well-defined animation effect. Compared to the discrete frame photographed above, the reconstructed running character in the video suffers from two constant “flare” due to the direct light leakage of the two alignment marks on the top and bottom sides of the patterns (refer to the photographs of the fabricated DOEs). Limited by the motion mechanics, we are unable to add additional cover to block these two marks when using a control motion stage. Nevertheless, this is not a fundamental issue of our method.

Possible color imaging solution. Conventional holographic projection suffers from the wavelength-dependency problem such that reconstructing a color image with a single pattern design is very tough. An alternative solution to obtain color projection is illustrated in Fig. 14. Again, following the holographic design scheme provides the design flexibility to merge or fusion multiple target holograms into one pattern design. Benefiting from the multiplexing scheme, the data bandwidth that can be encoded is drastically increased.

Application briefing for QR code encryption. We have mentioned that our multiplexing design can further boost new applications in security. Benefiting from the property that holograms intuitively
Supplementary Material:
Mix-and-Match Holography

• exhibit like random patterns, under some sense of customized multiplexing, the two phase DOEs can be thought of as an encrypted message and a key: both are required to decode target information. In this market, current shortcomings of wavelength-dependency and diffraction efficiency loss are not that critical. Here we briefly demonstrated an instance to reconstruct a QR code for message encryption (shown in Fig. 15). Notice that it can be multiplexed into the 4 application scenarios, especially the offset one (1-to-1) and the multiple pairing one (many-to-many), that have been detailed in the main tex. Please refer to the accompanying video.

Application briefing for emissive transparent display. One unique advantage that diffractive optics has over refractive optics is its higher light efficiency when used in short wavelength’s spectra (e.g. ultraviolet (UV)). Most glass medium have inevitable high absorption on UV spectrum (i.e. $\lambda < 400$ nm). To guarantee relative high transmittance, expensive material and special coating need to be involved when designing UV-based refractive optics.

We have tested one of our DOE pair on an emissive fluorescent film (provided by [Sun et al. 2013]). This specially designed fluorescent film absorbs UV light and emits visible light, making it particularly promising in a wide range of VR/AR applications. As shown in Fig. 16, directly illuminating a pair of our DOEs with 405nm UV light source, and accordingly adjusting the projection distance, we successfully observe a red color reconstructed “cat” on an optical-clear film. Notice that the relative low image contrast and the noisy background are due to the large wavelength deviation (i.e. our DOEs are optimized at 550nm). Here we only show an example application. Combining the investigated multiplexing schemes with this emissive transparent display technique, a wide range of applications can be enabled.

Accompanying video. We recommend the readers to refer to the accompanying video to comprehensively catch the spirit of our mix-and-match holography solution.

REFERENCES
Fig. 7. Comparisons of synthetic images reconstructed from the offset pairing subject to different implementation conditions— (a) perfect condition, (b) weak noise, (c) strong noise, (d) small rotation, (e) large rotation, (f) strong noise + small rotation one (g) directional shift, (h) dual directional shift, (i) strong noise + large rotation, respectively. Notice that all images have been resampled and integrated over 200µm per pixel subject to human eye’s angular resolution at a viewing distance of 50cm. (Lena image source from Wikipedia)
Fig. 8. Comparisons of synthetic (left) images and experimental images (right two) reconstructed from the offset pairing. Notice that all the images have been resampled and integrated over 200µm per pixel subject to human eye’s angular resolution at a viewing distance of 50cm. We also note that the intensity distribution of a target image has an impact on the encoding performance, further reflected as the speckle noise level in the decoded image. Specifically, the hair of Einstein exhibits a nosier distribution than that of Lena’s hat. Overall, the two decoded gray-scale images are visually recognizable with details preserved. (Albert Einstein Photograph by Orren Jack Turner, Lena image source from Wikipedia)
Fig. 9. Hologram visualization results (top row) and their corresponding synthetic reconstruction results (bottom row) of the banking account encoding application scenario. The parameters are the same as in Fig. 8 of the main text.
Fig. 10. Comparisons of synthetic images reconstructed from the offset pairing subject to two factorization update rules. Specially, we provide zoom-in close-ups of the second pair, as an instance. Notice that all images have been resampled and integrated as in Fig. 7.
Fig. 11. Synthetic results (top) reconstructed from the offset pairing of factorizing the phase profile of a Fresnel lens, and its factorization convergence evaluation (bottom).
Fig. 12. The experimental (top) and corresponding simulated (bottom) results of projecting 26 animated frames through translational pairing of two DOEs.
Fig. 13. The synthetic comparison results of multiplexing 13 frames versus 26 frames onto translational pairing of two DOE with the same data bandwidth. Readers may notice the difference of image contrast and background noise level.
Fig. 14. An alternative solution of projecting a dual-color image through the offset pairing of DOEs with coded light sources. Left: diagram of set-up, where two binary masks are to be embedded in the light path to encode the incident illumination such that only those sub-regions designed for green light are illuminated by green light, meanwhile only those sub-regions designed for red light are illuminated by red light. Right: green channel, red channel, and dual-color reconstruction, respectively. Notice that this fusion can be multiplex into other application scenarios that have been detailed in the main text, by carefully constructing the target matrix incorporating the “filtering” process of coded light sources.

Fig. 15. An alternative application proof of incorporating mix-and-match multiplexing holograms with QR code encryption. Notice that the two DOEs are individually with the volume of 8mm by 8mm by 0.5mm only, so as to be easily assembled onto any kind of electronic device. With spot light or planar light illumination, a visualization with much larger size (5cm by 5cm at 1m distance) can be decoded.

Fig. 16. A specific application proof of incorporating mix-and-match multiplexing holograms with UV emissive display. From the working status shown we observe a reconstructed scene as well as the background scene. The photographed scene is shown in the center close-up as a reference.