Clearing Key Barriers to Mass Adoption of Augmented Reality with Computer-Generated Holography

T. Widjanarko[a], M. El Guendy[a,b], A. O. Spiess[a], D. M. Sullivan[a], T. J. Durrant[a], O. A. Tastemur[a], A. J. Newman[a], D. F. Milne[a], A. Kaczorowski*[a]
[a] VividQ Ltd., Research Division, Cambridge, UK;
[b] Centre of Molecular Materials for Photonics and Electronics, University of Cambridge, UK

ABSTRACT

Augmented and Mixed Reality promises another leap forward in productivity and lifestyle, offering benefits with a magnitude and impact matching that of the introduction of smartphones. However, to enable this, many significant technical challenges must be overcome.

Here we review the state of the art, identifying key challenges established in the literature to consumer-wearable devices. In particular, we discuss: vergence-accommodation conflict (the detrimental effect of overlays that are optically inconsistent with the real-world objects they augment), the need to present overlays visible against the vast dynamic range that the human eye can process, and constraints surrounding the scalability and cost of manufacture of optics.

We demonstrate that digital holography as a display mechanism not only provides an effective solution to the aforementioned challenges, but also that various hardware requirements become far less stringent. By operating in the Fourier Domain, holographic displays are freed of design compromises driven by the constraints of a pixelated screen.

However, the computational cost of CGH has previously been considered prohibitive. We demonstrate that for real-world applications the latest advancements made by VividQ deliver sufficient focal accuracy at a computational cost within reach of personal mobile devices.

We prove that it is now possible to clear the barriers preventing mass adoption of Augmented and Mixed Reality products with Computer-Generated Holography.

Keywords: Digital Holography, Computer-Generated Holography, 3D Holography, Augmented Reality, Mixed Reality, Vergence-Accommodation Conflict, Aberration Correction, Smartglasses

1. INTRODUCTION

Extended Reality (XR) is a highly anticipated product space[1,2] drawing attention from leading technology businesses. It is best understood as a range of solutions, from Virtual Reality (VR) where virtual content is experienced to the exclusion of reality; to Augmented Reality, in which the visible world is the focus of the experience and virtual content is presented alongside. The spectrum is shown in Fig.1.

The domain of 'Mixed Reality' (where the perception of the objective world is enhanced with a virtual overlay that provides context), information and engagement is of particular interest, and studies show promise of transformative impact in industries such as, education and training[4–7], design and manufacturing[8–10], healthcare (assisting surgery)[11] and archaeology (helping to reconstruct old ruins)[12,13].

Here we summarise findings on the challenges that must be overcome to deliver on these promises. The nature of the display technology used is shown to be a determining factor for a key set of issues in optical see-through head-mounted MR applications. For each such issue identified, research on the nature of the issue is presented, along with details of recent work in Computer Generated Holography that shows how the issue can be addressed.

1.1. Challenges in MR

Various views exist among researchers on the most critical technological barriers to the mass adoption of MR.

*ak@vivid-q.com; www.vivid-q.com

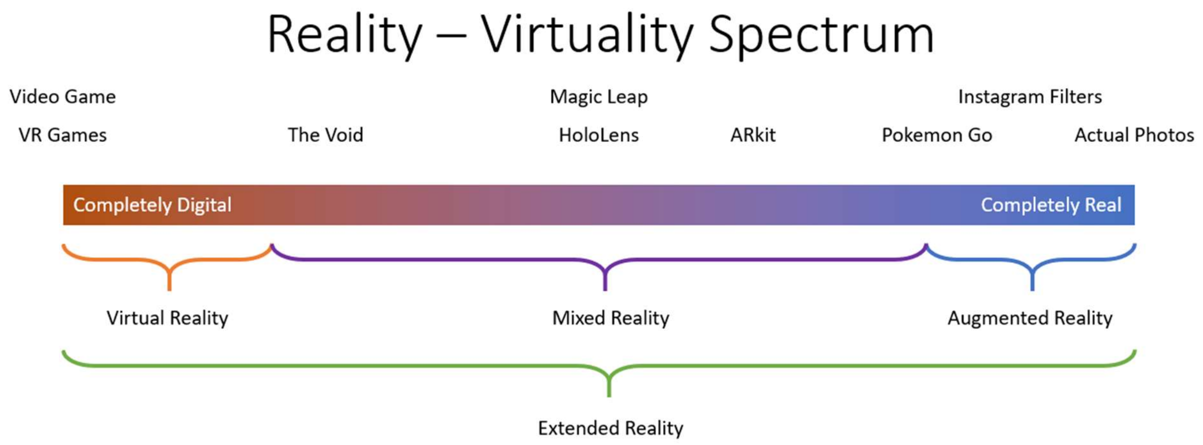


Figure 1. Reality-Virtuality spectrum[3] – spans from fully real to purely virtual

Azuma et al identify brightness and contrast, resolution, field of view, eye strain and form-factor[14] as the most important. D. Wagner et al[15] offer a more in-depth list: field of view, eye box size, brightness, contrast, uniformity, color quality, resolution, real-world distortions, virtual image distortions, eye safety, eye relief, peripheral vision, chromatic aberrations, depth perception, form-factor, optical efficiency, latency, and stray light. Kruijff et al[16] have compiled a comprehensive list of visual perception factors, specifically for AR and MR use-cases and considering display contexts from head mounted devices to head up displays, grouping the items into various issues including environment, capturing, augmentation, display, and others.

We believe that the aforementioned studies are representative of literature in the field, and suggest that issues particularly applicable to an optical see-through head mounted display for Mixed Reality applications are:

- Field of view;
- Ambient Contrast Ratio (visibility of the content against the environment) and eye safety;
- Eye strain, depth perception cues, and accommodation;
- Form factor, power consumption and battery life;
- Display properties (contrast, resolution, colour fidelity, reflections); and
- Spatial mapping, head tracking and latency.

1.2. Areas for Study

The issues identified can be separated into broad classes as follows:

- Factors that are largely independent of choice of display technology

  While the specifics of the display technology may significantly influence the nature of the integration, the key barriers that must be overcome are largely or entirely unrelated to the display technology used.

- Factors that are largely determined by the implementation of a display technology

  These factors depend sufficiently on the details of how a solution is implemented, that it is difficult or impossible to meaningfully assess how they depend on the choice of display technology.

- Factors that largely depend on the choice of display technology

  These factors are largely or entirely determined by the nature of the display technology chosen.

From this analysis, it is clear that the selection of core display technology is critical for the industry; this decision will govern how well consumer expectations on multiple key factors can be met, if at all. Therefore, the three factors in the right-hand column of Table 1 were selected as foci for investigation.

TABLE 1. HOW BARRIERS TO MR ADOPTION RELATE TO DISPLAY TYPE

Largely Independent of Display Technology	Determined by Implementation of Display Technology	Depends on Choice of Display Technology
Field of view	Display properties (contrast, resolution, colour fidelity, reflections)	Ambient Contrast Ratio and eye safety
Spatial mapping, head tracking and latency		Eye strain, depth perception cues, and accommodation
		Form factor, power consumption and battery life

1.3. Addressing Key Challenges for MR in see-through MR devices

This work presents recent developments that show significant promise in addressing the issues identified.

In particular, the benefits of Computer-Generated Holography (CGH) in addressing the Vergence-Accommodation Conflict and thus reducing eye strain are well understood, but we demonstrate that CGH can also address key challenges in ambient contrast and form factor. Since it enables each of the key factors above to be addressed, it may be of merit as a candidate in the question of the core display technology for the MR industry.

2. AMBIENT CONTRAST RATIO

2.1. Definition of Ambient Contrast Ratio

Contrast can be defined as the difference in colour and brightness of an object in comparison to others within the field of view[17] and is generally considered one of the most important aspects in perception of image quality.

Studies on human vision and perception confirm that there is a luminance threshold level above which the difficulty of an object recognition task stabilises, becoming invariant under any further increase in contrast[17]. Therefore, for content to be easily differentiated from the background, the contrast of the content against the background must at least reach this limit. Exceeding it may also cause the content to be perceived as more realistic[18].

Contrast for content presented against a black background can be effectively defined as the ratio of brightest to darkest adjacent points in the display. This is typically the case for conventional displays. In Mixed Reality applications, however, the content is presented against the variably lit objective world, and the metric used must account for this.

Ambient Contrast Ratio (ACR) is therefore defined as:

$$ACR = \frac{L_{on} + L_{background} \times T}{L_{off} + L_{background} \times T},$$

where L_{on} is the peak luminance of a point in the replay field from the display, $L_{background}$ is the luminance of the environment, and T is the transmittance of the combiner.

To illustrate this effect, Fig. 2 shows ACR for three hypothetical displays, chosen to be representative of the capabilities of leading display technologies, with L_{on} of 2,000 cd/m^2, 10’000 cd/m^2, and 300,000 cd/m^2 and contrast ratios of 10,000:1, 500:1 and 100:1 luminance contrast, respectively. T is assumed to be 25%.

The two curves for lower peak luminance start from the luminance contrast of the display and drop with rising ambient luminance. A very bright display with the worst contrast performs far better than other technologies from illuminance levels above 350 cd/m^2, which corresponds to a well-lit indoor environment or an outdoor environment on an overcast day.

It is then clear that, from the point of view of visibility, the peak brightness of the display is a far more important factor than the traditional L_{on} vs L_{off} contrast ratio from the point of view of an MR device.

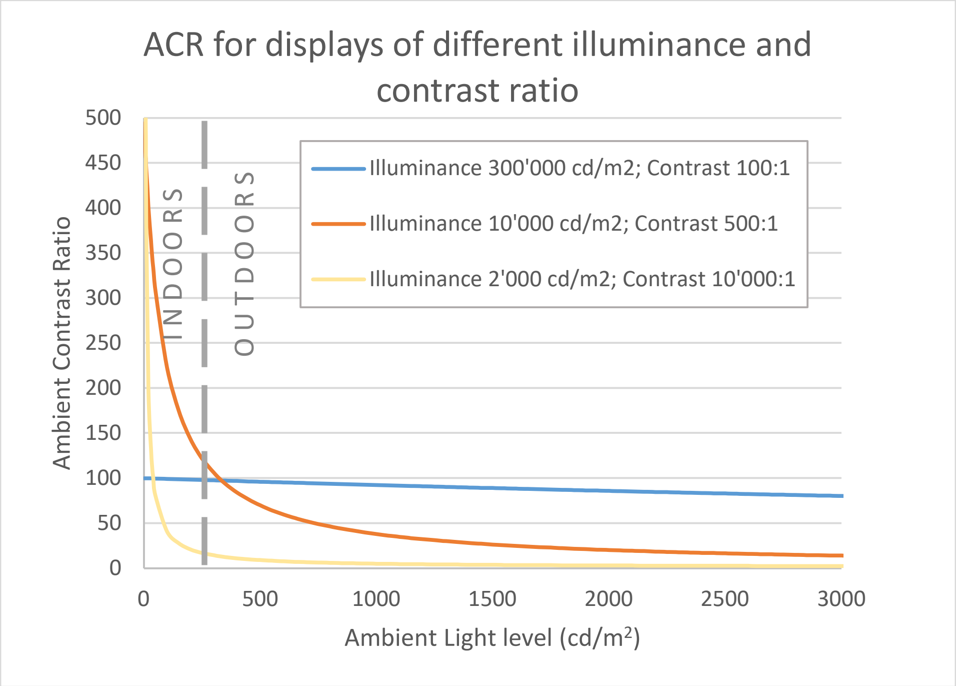


Figure 2. Plot of Ambient Contrast Ratio as function of Ambient Light Level

2.2. Ambient Contrast as a Key Issue for Optical See-Through MR HMD

Despite being recognised as an essential factor in AR and MR, no quantitative study on contrast requirements in virtual reality display has been reported. Gabbard et al[18] instigated a study of AR users in performing text identification tasks using an optical, see-through display. Emphasis is on the effect of three aspects: outdoor background textures, variation of outdoor illuminance and text drawing styles. They confirm that the performance of AR users in identifying text is strongly affected by the above factors and a combination of background texture and text drawing styles[18]. Background texture is also found to affect the local illuminance needed by the user to identify the text. However, this study does not explicitly state quantitative brightness figures required for the text or annotation identification in an AR display so that a sufficient contrast can be perceived to clearly identify the text for various illumination backgrounds.

2.3. Addressing poor Ambient Contrast Ratio within Computer-Generated Holography

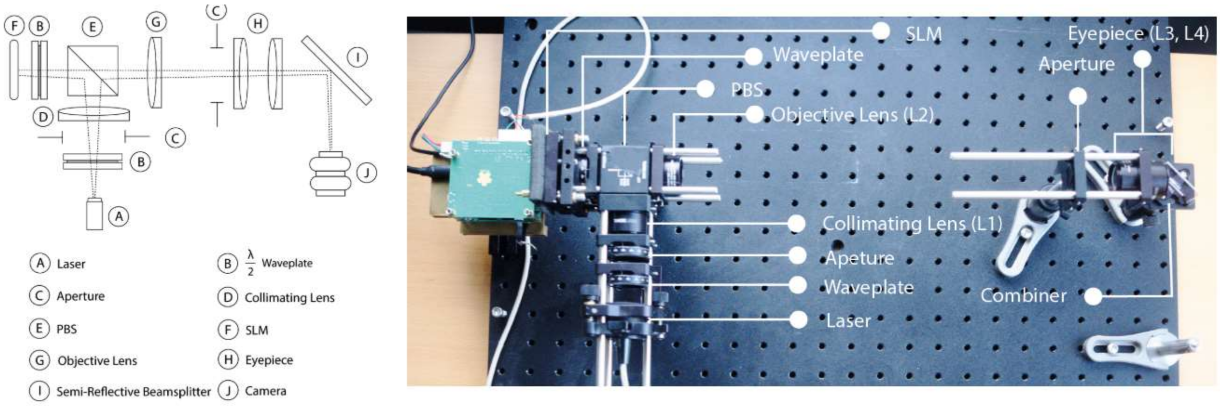
As previously shown, a satisfactory Ambient Contrast Ratio can only be achieved for a wide range of ambient light levels with a display module with a high peak brightness. In order to understand the capability of Computer-Generated Holography in this domain, we have conducted a study where we analysed the path from the initial light source to the viewer's eye. This work was presented at the EuroDisplay 2019 conference and is the subject of a forthcoming publication[19] which we summarise below.

A typical CGH approach uses a phase-modulating Spatial Light Modulator (SLM) to construct the holographic wavefront, compared to the amplitude modulation typical in conventional displays. Because the phase modulation approach acts to preserve the energy from the illumination source, the system can be much more efficient than those using amplitude-modulating displays.

In the conventional approach, utilizing an amplitude display, light is blocked or discarded in positions where no image overlay is needed. In a sparse scene, where only a small portion of the pixels in the source scene are occupied, the majority of the light is therefore lost, and converted into heat that must be dissipated. Holography, however, is performed on a phase device, and instead of discarding light, redistributes all the available energy to the relevant parts of the image. For this reason, even with the same illumination source and optical path as a conventional amplitude modulation display, optical efficiency of a holographic device will be multiple times higher for sparse content, such as the image in Fig 4b, below.

Further research is needed on the pixel occupancy requirements for valuable AR and MR use cases, but it is clear that where the intent of the application is to annotate or add informational overlays to the visible world, the pixel occupancy is necessarily a small fraction of the field of view, or the visible world becomes hard to perceive and the overlay becomes overwhelming, hence sparse scenes can be expected to be at least frequent. In Fig4b-c below, pixel occupancy is less than 10%.

In addition, CGH requires a coherent light source and typically uses laser illumination. This both has a very high luminance, enabling a high ACR even against bright backgrounds, and is typically physically separated from the light modulator element. Many common display technologies are temperature sensitive, leading to destructive feedback effects as thermal dissipation leads to loss of efficiency and further thermal dissipation. Separating the light modulator and illumination source allows better management of heat dissipation and helps prevent thermal run-away.



(a) (b)

Figure 3. Holographic projector constructed for the study: (a) schematic diagram, and (b) setup photograph.

VividQ has recently demonstrated the brightness that can be achieved in practice with a holographic projector[19]. Figure 3 shows the layout of the projector. The laser light source in the setup is an RGB laser diode (Sumitomo Electric) which produces 234 lm/W (optical power) at D65 white and 7.8% electrical-to-optical conversion efficiency. The laser beam, which is expanded and collimated (D), illuminates the SLM (F) in reflection mode, and then is directed into the viewer's eye via a telescope (G) and (H). The function of the field aperture (C) is to clean up the image from repetitive diffraction orders as well as undiffracted light (the zero order). The magnification of the telescopic system results in a field of view approximately 16 by 8 degrees with an eyebox approximately 4mm by 3mm. The 3D image is then merged into the user's view via a 50:50 plate beamsplitter combiner (I). The camera (J) used to acquire the image is a Sony RX100IV, specifically selected due to its close similarity of its optical properties to the human eye, in terms of entrance pupil size focal length, aperture and apparent field of view. The camera is shown in Fig.4a.

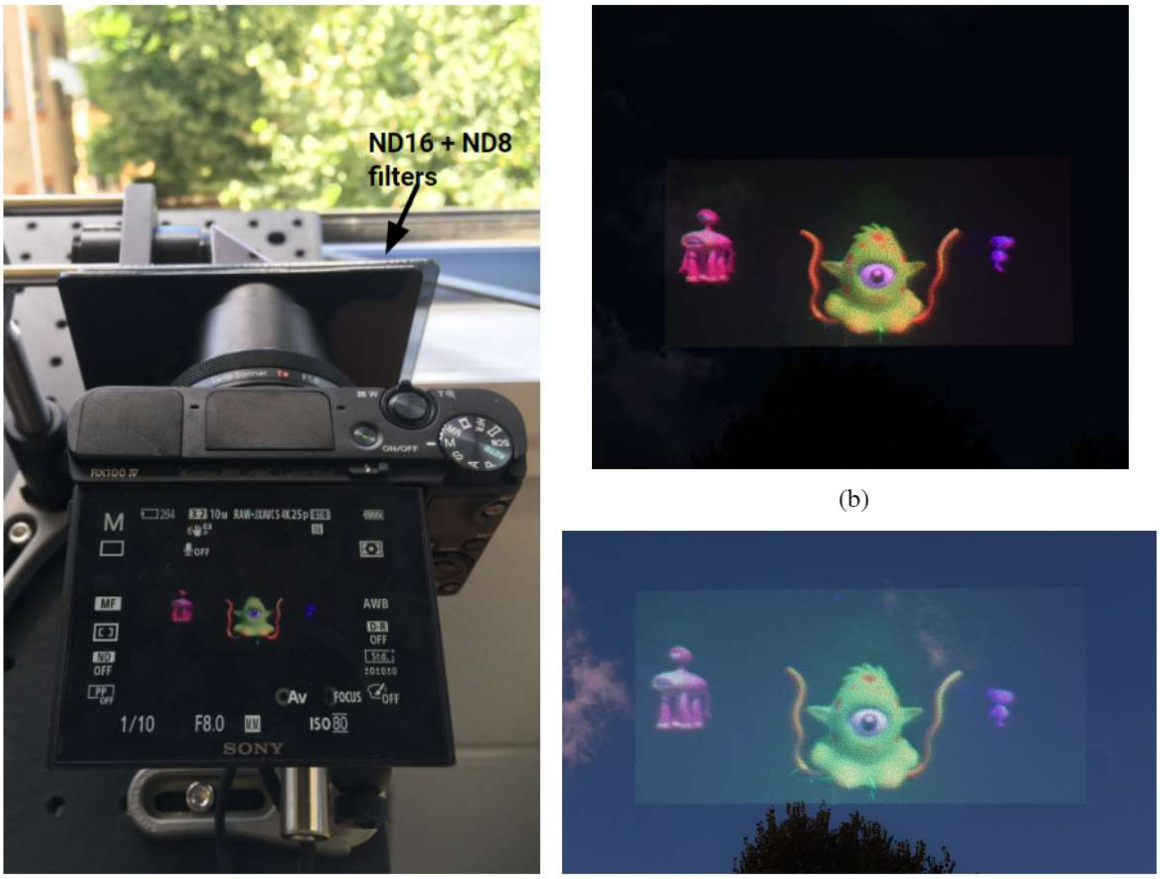
The optical losses in the system were divided into four distinct categories. The first, laser conditioning, specifies how much light generated by the laser ends up being available to the Spatial Light Modulator. The second, SLM illumination, is the efficiency and the losses at the microdisplay: efficiency of the modulation, and illumination as well as light discarded into various diffraction orders. The third, relay optics, are solely reflections from the optical surfaces, and the fourth, the combiner, is the final element that delivers the image to the eye. That may be a 50:50 beamsplitter (as used in this research) or a more sophisticated assembly of a waveguide and/or Holographic Optical Element which could perform additional optical functions like pupil replication.

By summarising all these factors (as seen in Table 2), we have concluded that the overall efficiency of the projector may vary from as much as 50% to as little as 0.05%. According to this analysis, the efficiency of 3% can be treated as typical. The particular setup we have implemented reached the efficiency of 2%. We achieved this by capturing a substantial portion of the beam (70% laser conditioning).

Images were captured with f=25.7mm at f/8, resulting in a pupil diameter of 3.2mm. A slow exposure of 1/10s was chosen to counteract the flicker due from SLM refresh rate. The slow exposure leads to saturation especially from the outdoor light background. Therefore, multiple Neutral Density (ND) filters were placed in front of the lens.

TABLE 2. LIST OF EFFICIENCY FIGURES OF EACH HOLOGRAPHIC PROJECTOR COMPONENTS[19]

	Best	**Typical**	**Worst**	**The setup[19]**
Laser conditioning	>90%	50%	<20%	**~70%**
SLM illumination	>70%	14%	<7%	**7%**
Relay optics	>90%	80%	<70%	**~90%**
Combiner	>90%	50%	<5%	**40%**
Total	**>50%**	**3%**	**<0.05%**	**~2%**



(b)

(a)

(c)

Figure 4. Experimental results from the holographic projector overlaying an image on the daytime sky. Picture was taken on 4th September at 2:47pm: (a) camera used in the holographic setup to capture the final images, with ND filters to prevent saturation, (b) image result from high power test and (c) the image result for low optical power setting[19].

An example captured image is shown in Fig.4a-b, which was taken at high optical power settings with measured power at the laser: R (red) = 0.89 mW, G (green) = 0.50 mW and B (blue) = 0.39 mW, resulted in a total optical power of 1.78 mW and luminous flux of 376 mlm. At this setting the achieved peak brightness is approximately 300,000 cd/m^2, which is above the brightness level of 10,000 cd/m^2 typically needed to display content over a daytime sky and reproducing a convincing real world scene[19]. In this setting the brightest clouds remain barely visible in the background. The lower power setting, i.e. measured laser power R = 0.10 mW, G = 0.09 mW and B = 0.07 mW, resulting in a total optical power of 0.26 mW and luminous flux of 59 mlm. The corresponding captured image is shown in Fig.4c. The resulting peak brightness is approximately 50,000 cd/m^2, which is again significantly above the minimum brightness required for outdoor light background. At the eye the corresponding luminous flux levels were 1.2 and 7.5 mlm for low and high optical power settings respectively, lying within the eye safety Maximum Permissible Exposure (MPE) level. The experiments show the high efficiency properties of holographic display in producing high brightness and contrast scenes.

In summary, VividQ has demonstrated a method that increases the brightness of the image, to over 10,000 cd/m^2, when using less than 5mW of laser power. Lasers therefore achieve a brightness of the image that is far superior to LED and OLED counterparts[20].

3. EYE STRAIN, DEPTH PERCEPTION CUES & ACCOMMODATION

3.1. Vergence-Accommodation Conflict as the Cause of Eyestrain and Poor Perception of Depth

VAC arises when a flat (2D) representation of a virtual scene is presented to each eye at a fixed focal distance determined by the display optics. The difference in perspective in the virtual images presented to each eye satisfies the stereoscopic depth cue, and the nature of the scene presented may provide further depth information. The adjustments the eye makes to bring the image into focus is interpreted by the brain as an additional depth cue, and if the apparent distance of the image, as indicated by stereoscopic depth cues, does not match the focal distance of the 2D display, these depth cues will conflict. The stereoscopic cue for distance typically dominates, but a significant proportion of users are left with visual discomfort, nausea or headaches as a result[21–23].

Multiple studies have examined the effects of VAC: Shibata et al[24,25] performed a series of experiments attempting to record the effects of viewing distance and location of the virtual object on viewers' discomfort and fatigue. They conclude that with negative conflicts, i.e. when the apparent depth from stereo depth cues is further from the viewer than the focal depth, viewer discomfort is more significant at greater distances, while the opposite is true for positive conflicts. Ukai et al[27] review the health implications of viewing moving stereoscopic images. Kuze and Ukai[27] compared different types of moving images from two-dimensional movies to stereoscopic movies and conclude that the latter cause the most visual fatigue. Yano et al[28] demonstrate that visual fatigue is not as significant when viewing still stereoscopic images but that when images are moving, particularly in the depth plane, visual fatigue is more notable.

Typically, vergence-accommodation is considered a subjective assessment of people's reactions to the continual adaptation of an image's virtual focal distance. Lambooij et al[28], argues that while visual discomfort is a subjective reaction, visual fatigue is a quantifiable response.

Despite being a common problem, VAC affects people differently, and, in order to ensure AR's mass adoption, it is important to fully understand the effects of this condition on various demographics. Furthermore, analysis shows that different image patterns and depths produce different VAC effects[31].

While the impact of VAC may vary depending on use-case and demographic, and while the effects may not be pronounced in all individuals, studies have shown that negative experiences affecting a small proportion of the user population can have a disproportionate impact on the adoption of a new technology[34, 35]. Hence, user experience issues such as VAC could present a much more significant barrier to the introduction of MR devices than a particular user's impression might imply.

3.2. Addressing Vergence-Accommodation Conflict with a Holographic Display

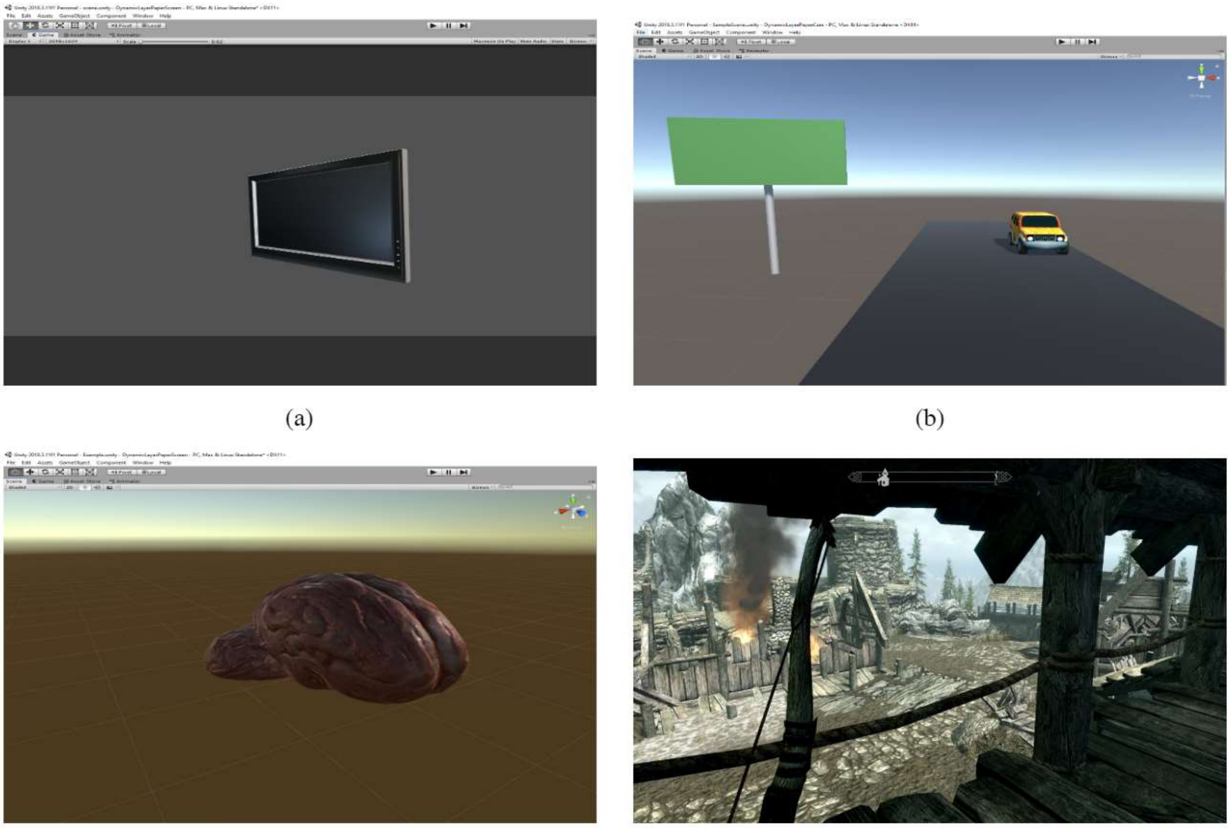
True 3D holographic displays are the most evident solution to address VAC, whereby viewers are presented with scenes having actual focal depth[30,31]. Hologram generation procedures frequently produce a 3D scene by diving the target image into discrete depth planes (layers) and producing a hologram that encodes all of these.

Algorithms have been developed to identify the required number of layers to adequately sample the object depth, i.e. to produce a convincing impression of continuous depth to the viewer. For example, Zabels et al[32] and Chen et al[33] use 6 and 10 depth layers per image, respectively. The fewer layers used, the faster the computation. The more depth layers used, the less the user will suffer from VAC. Thus, there is a trade-off between the number of layers in a holographic image and its computational speed, with an increased number of planes resulting in more accurate focal depth measurements, but also correspondingly increasing the processing cost.

An approach that shows promise is to be able to dynamically adjust the number of depth layers in depicting a 3D scene, allowing the focal accuracy and computational cost to be balanced according to the needs of the use-case. VividQ has developed such an algorithm, which can adaptively place sufficiently many layers on the object for a compelling user experience, but still remains maintains a reasonable load on the computational module[34].

For an adult with normal vision, a sequence of depth layers with a separation of 0.15 diopter will no longer be visually perceived as discrete but appear as a smooth and continuous depth of a 3D object[34]. To accurately represent the depth range covered by the typical human vision, which is 15 diopters, the required number of depth planes required would be 15/0.15 = 100, which would put a significant computational load on the processor. With a dynamic allocation approach a number of depth layers is determined that reliably falls within the computational capacity available, and these are used to cover the scene as completely as possible on a frame-by-frame basis.

The problem which the algorithm solves can be stated as follows: for a given list of input layers and a fixed number of output layers, place the output layers in a way that maximises the coverage of the object depth. The algorithm starts by placing output layers on top of the nearest and furthest point of the input. From here the algorithm finds the input plane with maximum error (in diopter space) and assign an output plane on that layer. The algorithm continues over the planes of maximum error until the maximum number of output layers is reached[34].


(a)

(b)

(c)

(d)

Figure 5. Several simulated scenes for evaluating the performance of the Dynamic Layer Allocation (DLA) algorithm[34]: (a) Wall TV, (b) Road view, (c) Brain and (d) Skyrim war zone.

In order to examine the performance of the proposed algorithm, we have selected a range of sample scenes, which correspond to common use-cases of Augmented and Mixed Reality. The scenes are presented in Fig 5. Scene 1 (Fig 5a) is a television screen placed on the wall, which is not perpendicular to the user and occupies depth distance of 1.5 to 2 m. Scene 2 (Fig 5b) is relevant to Heads Up Display applications, showing the dashboard, street sign, and another car in the field of view. The objects are modeled as follows: the depth of dashboard spans between 0.8 to 0.9m, a flat sign is located at 3m, and the depth of distant car is between 9 to 10m. Scene 3 (Fig 5c) is indicative of high-value use cases for AR such as assisting surgery, or assembly of small and delicate parts. The display is expected to provide high resolution images with good focus accuracy, but most likely the object(s) span only over a small range of depth distances. The object is a brain and simulated to span in depth from 0.2 to 0.5 m (3 diopters). Finally, Scene 4 (Fig 5d) is the data taken from an existing computer game - Skyrim.

We have compared our proposed algorithm to a fixed placement of layers in equal distances in the dioptre space between 0.2m and 20m (which corresponds to the hyperfocal distance). For each of these scenes, we have compared the focus error of a standard algorithm to the proposed one and evaluated what percentage of the object(s) are in acceptable focus. This data is summarised in Table 3.

TABLE 3. COMPARISON OF FIXED AND DYNAMIC DEPTH LAYER PLACEMENTS

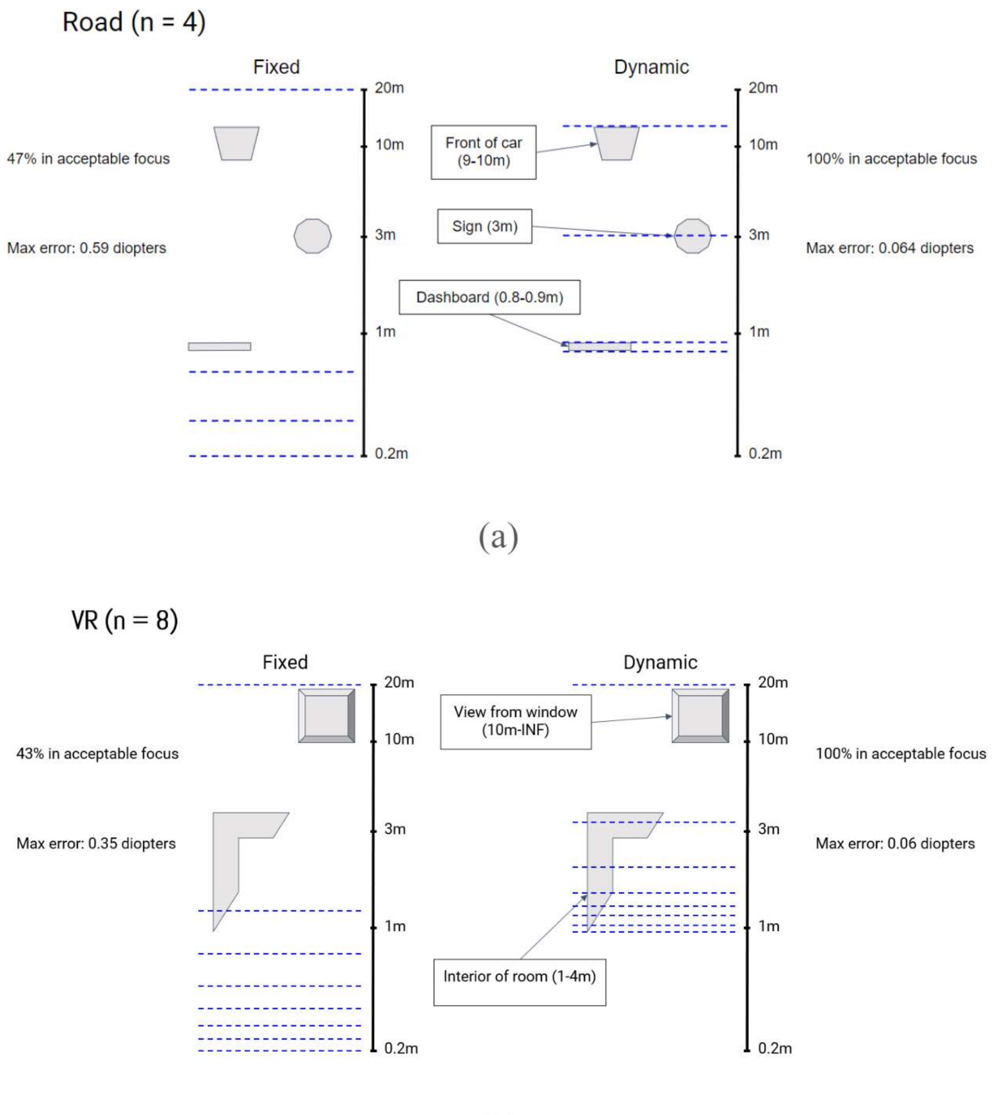
Algorithm	**n**	**Max. error (in diopters)**				**Coverage ratio (of acceptable focus)**			
		Scene 1	Scene 2	Scene 3	Scene 4	Scene 1	Scene 2	Scene 3	Scene 4
Fixed Layers	2	0.62	1.20	2.50	0.95	0%	47%	3%	33%
	4	0.62	0.59	0.81	0.83	0%	47%	6%	33%
	8	0.26	0.35	0.34	0.35	0%	47%	24%	43%
Proposed Algorithm[34]	2	0.08	0.23	1.50	0.47	90%	75%	7%	37%
	4	0.04	0.06	0.72	0.24	100%	100%	19%	69%
	8	0.02	0.01	0.34	0.06	100%	100%	38%	100%

It can be seen that the proposed algorithm is always performing better than the standard fixed placement of layers. For the first two scenes, we found 4 layers to be sufficient to reach the full coverage. The fourth scene reached the full coverage at 8 layers, while the third scene requires more than 8 layers to represent the object sufficiently. Nonetheless, the proposed algorithm reached a coverage of 38%, compared to the standard implementation which is only capable of 24% coverage.

The graphical illustration of these results can be seen in Fig 6. Fig 6a shows Scene 2 in the case of 4 layers, and Fig 6b shows Scene 4 covered by 8 layers. It should be noted that the focus error of the proposed method is approximately 10 and 5 times smaller than the standard method respectively.

One important point that has to be emphasized here is that there is a direct trade-off between the number of layers used and the compute requirement associated. The exact number of floating-point operations for generating a single hologram frame at 2K × 2K resolution (and displayed at 2K × 1.5K SLM) using VividQ's DLA real-time algorithm. The complexity of algorithm for layer-based hologram generation is approximately linear in the number of layers (as seen in Table 4).

VividQ's DLA algorithm[34] can be configured to vary the depth planes used, allowing framerate to be balanced against focal accuracy depending on the use case and the need, achieving greater than 60FPS with 12 depth planes for a binocular holographic display on consumer-available GPUs, sufficient for perceptually accurate focal depth for a wide range of content more than one metre from the eye. The implementation of multiple depth planes and the user perception of image depth mitigates the effects of vergence accommodation and thus contributes towards the mass adoption of 3D holographic displays in AR/MR/XR units.



(a)

(b)

Figure 6. (a) Results from the proposed and standard algorithms on Scene 2 for 4 layers - HUD demonstration (b) Scene 4 - Entirely Virtual environment, 8 layers[34].

TABLE 4. FLOPS REQUIRED FOR NUMBERS OF LAYERS USING DLA ALGORITHM[34]

Layers	**Complexity per frame (GFLOP)**
2	7
4	12
8	22
16	42

4. POWER CONSUMPTION, BATTERY LIFE AND FORM FACTOR

4.1. Impact of Optical Aberrations on Form Factor

As Beams et al demonstrate, display optics must balance weight, form factor and cost, with compromises to the optical performance of the display shown to be the main factor impacting on image quality in an XR devices[37].

Such compromises are primarily caused by optical aberrations from imperfect elements in the optical path. An ideal lens or mirror, for example, would have a surface that perfectly follows that modelled in the optical design, but manufacturing such an ideal component is impossible, and the more accurately a real component approaches it, the more the size, cost and weight of the component typically increases. The less perfectly a real component matches the optical model, the more aberrations, and the more significant the aberrations, it introduces.

There are various types of aberrations, most notably chromatic aberration, occurring in the lenses, focusing different colour wavelengths at different points, and spherical aberration, where light rays focus at different points due to an optical component's spherical surface.

Aberrations in an optical system, with a proper and careful design, can be minimised and even eliminated, using a complex, multi-element lens train. However, this conflicts significantly with commercial pressures on volume, price, and weight. Most current AR and MR HMDs use single lens optical systems in order to lower cost and improve form factor. As headset manufacturers attempt to meet user preferences for wider field of view, further trade-off occurs by compromising lens focal length. For a single element optical system this trade-off potentially increases distortion and spherical aberrations.

4.2. Holographic Correction of Optical Aberrations

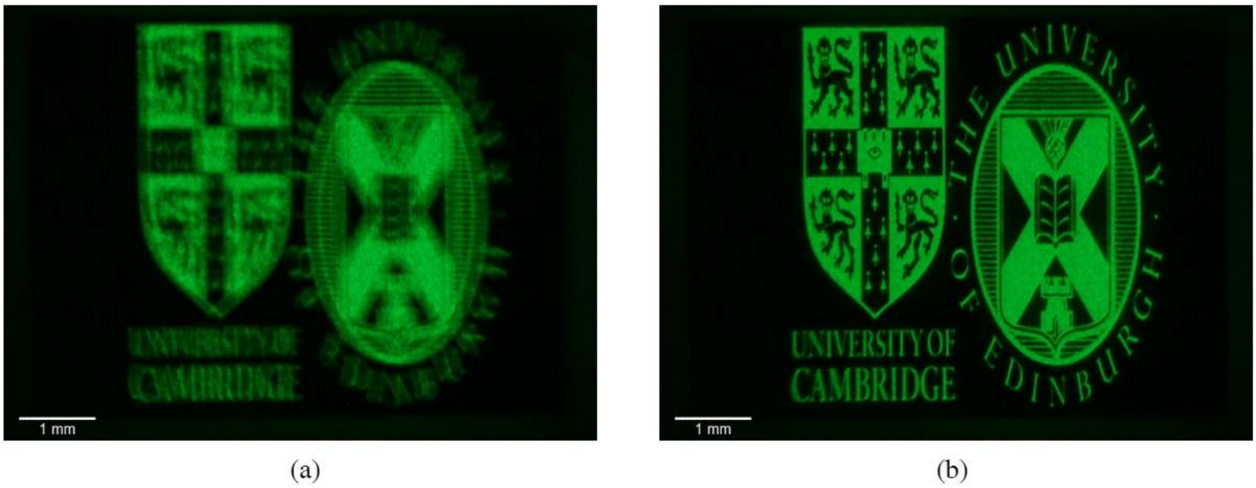
Ghebremichael et al[36] and Porfirev et al[37] consider low order Zernike polynomials to represent aberrations in an optical system. The lowest orders correct the most dominant aberrations.

Kaczorowski et al indicates that aberrations are essentially the disturbances in phase of the lightwave[40]. If the disturbance can be characterised for the lens train then the same disturbance (but out-of-phase compared to the original wave) can be introduced in advance to counter the wave distortion and correct the aberrations. The counter-distortion of the wave can be synthesised as a phase mask presented on a Spatial Light Modulator, which, due to the additive nature of holography, (which is a linear Fourier System) can simply be combined with a holographic interference pattern used to generate the intended image.

Computer-Generated Holography therefore enables correction of monochromatic aberrations in MR, allowing optical fidelity to be maintained without driving up cost and weight of the total system. This approach aligns well with the latest developments in display optics, where a Holographic Optical Element (HOE) is used to direct the beam and reduce form factor, such as the AR headset reported by Jang et al[39].

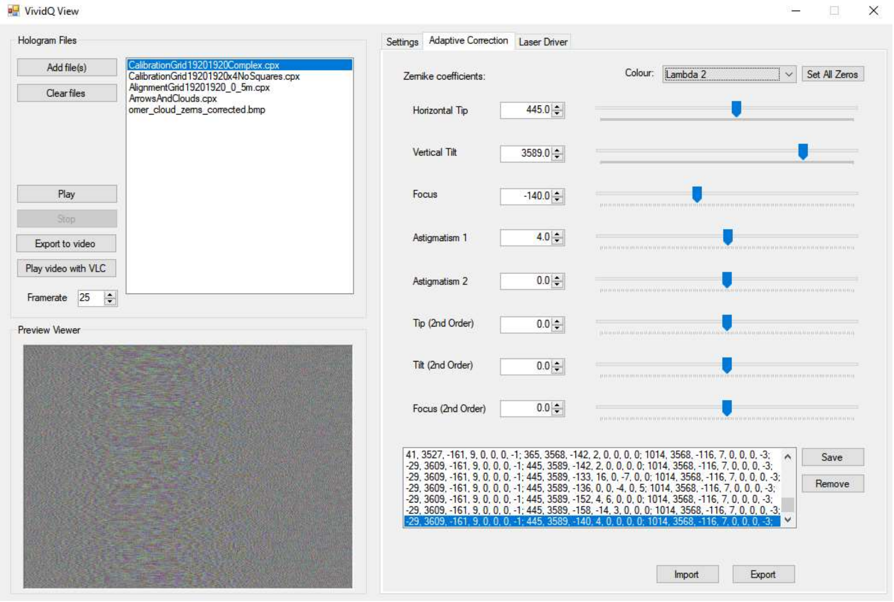
Kaczorowski et al[40] showed that field-invariant aberrations (such as non-uniform laser profile or SLM non-flatness) can be characterized and fully corrected even after the assembly process. Field-dependent aberrations can also be corrected, but at an expense of significantly increasing the compute[40,41].

An example correction of a large aberration can be witnessed in Fig 7. The example was performed on a 2D holographic projector, optimized for maskless lithography. However, the same methods and similar results are achieved in direct-view holographic systems. In order to unlock that functionality, VividQ developed VividQ View software module that can perform adaptive aberration correction on any holographic optical system (Fig 8).



(a) (b)

Figure 7. Aberration correction using holographic methods: (a) prior (b) after the correction[38].



```
CalibrationGrid19201920Complex.cpx
CalibrationGrid19201920x4NoSquares.cpx
AlignmentGrid19201920_0_5m.cpx
ArrowsAndClouds.cpx
omer_cloud_zerns_corrected.bmp
```

Framerate: 25

Zernike coefficients:

Colour: Lambda 2

Horizontal Tip: 445.0

Vertical Tilt: 3589.0

Focus: -140.0

Astigmatism 1: 4.0

Astigmatism 2: 0.0

Tip (2nd Order): 0.0

Tilt (2nd Order): 0.0

Focus (2nd Order): 0.0

```
41, 3527, -161, 9, 0, 0, 0, -1; 365, 3568, -142, 2, 0, 0, 0, 0; 1014, 3568, -116, 7, 0, 0, 0, -3;
-29, 3609, -161, 9, 0, 0, 0, -1; 445, 3589, -142, 2, 0, 0, 0, 0; 1014, 3568, -116, 7, 0, 0, 0, -3;
-29, 3609, -161, 9, 0, 0, 0, -1; 445, 3589, -133, 16, 0, -7, 0, 0; 1014, 3568, -116, 7, 0, 0, 0, -3;
-29, 3609, -161, 9, 0, 0, 0, -1; 445, 3589, -136, 0, 0, -4, 0, 5; 1014, 3568, -116, 7, 0, 0, 0, -3;
-29, 3609, -161, 9, 0, 0, 0, -1; 445, 3589, -152, 4, 6, 0, 0, 0; 1014, 3568, -116, 7, 0, 0, 0, -3;
-29, 3609, -161, 9, 0, 0, 0, -1; 445, 3589, -158, -14, 3, 0, 0, 0; 1014, 3568, -116, 7, 0, 0, 0, -3;
-29, 3609, -161, 9, 0, 0, 0, -1; 445, 3589, -140, 4, 0, 0, 0, 0; 1014, 3568, -116, 7, 0, 0, 0, -3;
```

Figure 8. Aberration correction routine performed using VividQ™ View.

4.3. Power Consumption

Weight and volume are key consideration for MR head-mounted optical see-through applications. While limited head movement might be acceptable in some applications, consumer expectations will demand a compact form factor with total volume not exceeding that of a current mobile phone, facilitating a high degree of head mobility. In the absence of a fundamental breakthrough in battery technology, this places a harsh limit on the amount of energy that can be made available to power a full day's use. Reducing power consumption to meet this target while maintaining the high-quality virtual experience of the users is a challenging goal.

Choi et al studied energy consumption of Magic Leap One (a commercially available untethered MR headset) with 2-3 hour battery life[42] in typical scenarios. They found that more than 40% of the total power comes from gaze-tracking sensors, gyroscopes, and the displays, with the display consuming almost 20 times the power of the wireless module, depending on data rate.

In order to meet consumer expectations of form factor and battery life, significant optimisations will need to be found in computational efficiency across the entire platform, including connectivity and SLAM as well as display.

4.4. Power Consumption Improvement

Laser-based display systems, such as CGH cannot match emissive displays such as OLED for power efficiency at low luminance. However, unlike CGH, which is shown to be able to achieve acceptable Ambient Contrast at all background lighting levels, there is no clear path for OLED or comparable technologies to scale to the peak pixel luminance needed by daylight environments, and it is not clear that the efficiencies they offer at low luminance could be maintained in this context.

This is notable because it is the optical power element of the power budget that is always going to be least susceptible to optimisation. While some use-cases may be viable in dimly lit environments, and thus may not require a high peak luminance, majority acceptance of AR and MR will only be achieved when clear contrast against any background can be taken for granted - even outdoors and in bright sunlight. However, it is clearly the case that whatever optical efficiency CGH can deliver today, it is more than offset by the power drawn to drive Liquid Crystal on Silicon (LCoS) Spatial Light Modulators and in the calculation of the holographic interference patterns.

While the power cost of current LCoS devices is indeed prohibitive for cordless HMD use cases, this is the state of a nascent technology for which power efficiency has not yet been a significant optimisation target. Given that the amplitude LCoS is currently undergoing developments, targeting reducing power requirements, we expect these advancements to naturally translate into phase LCoS panels.

Similarly, the focus of development in algorithms for true 3D Computer Generated Holography for head-mounted AR and MR use-cases has primarily focused on demonstrating that a viable image quality is achievable within the processing capability of generally available hardware, and power optimisation has not yet been a focus of research. It is reasonable to assume that significant gains can be made, as with many other technologies that are now ubiquitous, but which had a prohibitive power cost when first developed.

Further optimisation in optical efficiency, however, is a much more challenging proposition. Generation of luminance is a process that is strictly governed by thermodynamics, and that has been the subject of a great deal of study; while a breakthrough might be achieved through targeted compromise in a specific application, significant gains in efficiency cannot be expected. The same is true for the efficiency of optical systems, where the principles are well-understood.

We argue that the consumer expectations for brightness, contrast against background and battery lifetime of AR or MR device are only going to be met by a solution that starts with a viable *optical* power budget, and then finds the necessary efficiencies in the rest of the display system.

4.5. Future Work - Roadmap to Mobile Compute

VividQ's pipeline was developed on NVIDIA CUDA platform firstly because, GPUs are widespread, easily available and already used by a majority of businesses for a multitude of applications, and secondly because parallel architecture of GPUs is well suited to the computation of computer-generated holograms.

However, desktop GPUs are impractical for a portable, battery-powered solution. In the near term, solutions such as an NVIDIA Tegra System-on-Chip (SoC) could be considered for a tethered system small enough to be easily carried, but it is envisaged that a consumer untethered headset could be achieved through the use of an Application-Specific Integrated Circuit (ASIC) or SoC with a dedicated Holographic processor. In general, one may attempt to characterize the efficiency of a given compute architecture by taking two key factors into account: the peak compute factor, and how much power the architecture uses. For NVIDIA's platform this value is in the range of 50 GFLOPS/W, but architectures are already reported in the literature that achieve 130 GLOPS/W[43] and there are indications that this number can soon be surpassed and reach up to 200 GFLOPS/W.

Additionally, it is known that there are many optimizations that have not yet been explored, such as low-level fine-tuning of the Fourier Transform routines or reusing portions of computed data between adjacent frames. Having performed a preliminary analysis, which will be published in upcoming work from VividQ, we believe that a significant reduction in compute power is easily achievable with today's technology and available silicon processes.

5. CONCLUSIONS

The potential of Augmented and Mixed Reality solutions to transform working practice across a broad range of industries, to create new social dynamics and to create new consumer value and solutions is clear and the interest taken by the largest global technology businesses has been widely reported across the media.

However, this has built correspondingly high consumer expectations, in a domain where the technical challenges are at least intimidating - even to businesses able to dedicate billions of dollars to research and development - and even considered prohibitive by some educated commentators.

As this paper has shown, there has been only limited research into the factors that will actually govern consumer-satisfaction; and far more study must be done if AR and MR are to achieve the potential they promise. Display technologies that have succeeded - such as video compression, or touch screen interfaces - have always built upon deep understanding of the human factors involved, and a failure to understand human factors has always been a factor in display technologies that have struggled to succeed.

For Augmented/Extended Reality technologies to break through to the consumer market in volume, and to deliver the anticipated transformations in market and lifestyle, in particular, issues such as VAC and ambient contrast must be much better understood and addressed:

- For industrial applications, is it practical or acceptable to dismiss or exclude 1% of the workforce if the AR/MR solution introduced causes them nausea, or if eye strain or cognitive conflict impairs their function?
    - Existing studies are sufficient to establish this as a legitimate concern. Further study is needed to determine the extent of the concern.
- Can consumer value be delivered by an MR display that can't be used outdoors on a sunny day?
    - The capability and power cost to achieve the necessary ambient contrast must be understood for each display solution if a viable choice is to be made.

Figure 9. VividQ’s binocular holographic headset prototype.

Solutions such as Computer-Generated Holography offer compelling answers to some of the most daunting challenges facing development of the truly transformative AR and MR devices that will deliver the next display revolution. The challenges remain significant, and it is clear that a great deal more work must be done, both in understanding the human factors involved and developing the complete end-to-end systems that will address them, but the path to that solution now seems clear.

In order to demonstrate the above factors and show that through the contributions made by VividQ, Computer Generated Holography now offers a clear route to a consumer-acceptable solution, given sufficient investment, we have developed a prototype binocular holographic headset, as depicted in Fig 9. The compactness is achieved through embracing certain types of optical aberrations, which are fully corrected within the software. The headset is capable of showing bright and vivid imagery due to its optical efficiency, and Vergence-Accommodation Conflict is eliminated, by using CGH to place virtual objects at the correct focal depth.

REFERENCES

[1] Conway, N., Soro, A., Brown, R. and Turkay, S., “Exploring Uses of Augmented Reality in Participatory Marketing,” Ext. Abstr. 2019 CHI Conf. Hum. Factors Comput. Syst., 1–6, Association for Computing Machinery, Glasgow, Scotland Uk (2019).

[2] Wong, R. Y. and Mulligan, D. K., “When a Product Is Still Fictional: Anticipating and Speculating Futures through Concept Videos,” Proc. 2016 ACM Conf. Des. Interact. Syst., 121–133, Association for Computing Machinery, Brisbane, QLD, Australia (2016).

[3] “Mixed reality vs augmented reality: what’s the difference? - Aniwaa blog.”, Aniwaa (2018).

[4] Yuen, S., Yaoyuneyong, G. and Johnson, E., “Augmented Reality: An Overview and Five Directions for AR in Education,” J. Educ. Technol. Dev. Exch. JETDE 4(1) (2011).

[5] Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y. and Liang, J.-C., “Current status, opportunities and challenges of augmented reality in education,” Comput. Educ. 62, 41–49 (2013).

[6] Lee, K., “Augmented Reality in Education and Training,” TechTrends 56(2), 13–21 (2012).

[7] Akçayır, M. and Akçayır, G., “Advantages and challenges associated with augmented reality for education: A systematic review of the literature,” Educ. Res. Rev. 20, 1–11 (2017).

[8] Nee, A. Y. C., Ong, S. K., Chryssolouris, G. and Mourtzis, D., “Augmented reality applications in design and manufacturing,” CIRP Ann. 61(2), 657–679 (2012).

[9] Novak-Marcincin, J., Barna, J., Janak, M. and Novakova-Marcincinova, L., “Augmented Reality Aided Manufacturing,” Procedia Comput. Sci. 25, 23–31 (2013).

[10] Büttner, S., Mucha, H., Funk, M., Kosch, T., Aehnelt, M., Robert, S. and Röcker, C., “The Design Space of Augmented and Virtual Reality Applications for Assistive Environments in Manufacturing: A Visual Approach,” Proc. 10th Int. Conf. PErvasive Technol. Relat. Assist. Environ., 433–440, Association for Computing Machinery, Island of Rhodes, Greece (2017).

[11] Moline, J., “Virtual Reality for Health Care: a survey,” [Virtual Reality in Neuro-Psycho-Physiology], IOS Press, Amsterdam, Netherlands (1997).

[12] Morandi, S. and Tremari, M., “Make the invisible visible: Virtual reconstruction of a submerged medieval church (S. Siro, Como, Italy),” 2016 22nd Int. Conf. Virtual Syst. Multimed. VSMM, 1–4 (2016).

[13] Matsushita, R., Higo, T., Suita, H. and Yasumuro, Y., “[POSTER] On-site AR Interface with Web-Based 3D Archiving System for Archaeological Project,” 2015 IEEE Int. Symp. Mix. Augment. Real., 208–209 (2015).

[14] Azuma, R. T., “Making Augmented Reality a Reality,” Imaging Appl. Opt. 2017 3D AIO COSI MATH PcAOP 2017 Pap. JTu1F1, JTu1F.1, Optical Society of America (2017).

[15] “Why is making good AR displays so hard? | LinkedIn.”, <https://www.linkedin.com/pulse/why-making-good-ar-displays-so-hard-daniel-wagner/> (24 January 2020).

[16] Kruijff, E., Swan, J. E. and Feiner, S., “Perceptual issues in augmented reality revisited,” 2010 IEEE Int. Symp. Mix. Augment. Real., 3–12 (2010).

[17] Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E. and Malach, R., “Contrast Sensitivity in Human Visual Areas and Its Relationship to Object Recognition,” J. Neurophysiol. 87(6), 3102–3116 (2002).

[18] Gabbard, J. L., Swan, J. E., Hix, D., Schulman, R. S., Lucas, J. and Gupta, D., “An empirical user-based study of text drawing styles and outdoor background textures for augmented reality,” IEEE Proc. VR 2005 Virtual Real. 2005, 11–18 (2005).

[19] Newman, A. J., Spiess, A. O., Milne, D. F. and Kaczorowski, A., “An Intrinsically Eye Safe Approach to High Apparent Brightness Augmented Reality Displays using Digital Holography,” (Under Preparation).

[20] Haas, G., “40-2: Invited Paper: Microdisplays for Augmented and Virtual Reality,” SID Symp. Dig. Tech. Pap. 49(1), 506–509 (2018).
[21] Banks, M. S., Kim, J. and Shibata, T., “Insight into vergence/accommodation mismatch,” Head- Helmet-Mounted Disp. XVIII Des. Appl. 8735, 873509, International Society for Optics and Photonics (2013).
[22] Pei, C., Yan, X., Zhao, K. and Jiang, X., “Method for generating full-parallax holographic stereograms without vergence-accommodation conflicts,” J. Electron. Imaging 23(6), 061109 (2014).
[23] Nguyen, M., Tran, H. and Le, H., “Exploration of the 3D World on the Internet Using Commodity Virtual Reality Devices,” Multimodal Technol. Interact. 1(3), 15 (2017).
[24] Shibata, T., Kim, J., Hoffman, D. M. and Banks, M. S., “The zone of comfort: Predicting visual discomfort with stereo displays,” J. Vis. 11(8), 11–11 (2011).
[25] Shibata, T., Kim, J., Hoffman, D. M. and Banks, M. S., “Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accommodation conflict,” Stereosc. Disp. Appl. XXII 7863, 78630P, International Society for Optics and Photonics (2011).
[26] [Ukai, K. and Howarth, P. A., “Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations,” Displays 29, 106–116 (2008).
[27] Kuze, J. and Ukai, K., “Subjective evaluation of visual fatigue caused by motion images,” Displays 29(2), 159–166 (2008).
[28] Yano, S., Emoto, M. and Mitsuhashi, T., “Two factors in visual fatigue caused by stereoscopic HDTV images,” Displays 25(4), 141–150 (2004).
[29] Lambooij, M., Fortuin, M., Ijsselsteijn, W., Evans, B. and Heynderickx, I., “Measuring visual fatigue and visual discomfort associated with 3-D displays,” J. Soc. Inf. Disp. 18(11), 931–943 (2010).
[30] Takaki, Y. and Yokouchi, M., “Accommodation measurements of horizontally scanning holographic display,” Opt. Express 20(4), 3918–3931 (2012).
[31] Candry, P. and Maximus, B., “Projection displays: New technologies, challenges, and applications,” J. Soc. Inf. Disp. 23(8), 347–357 (2015).
[32] Zabels, R., Osmanis, K., Narels, M., Gertners, U., Ozols, A., Rūtenbergs, K. and Osmanis, I., “AR Displays: Next-Generation Technologies to Solve the Vergence–Accommodation Conflict,” Appl. Sci. 9(15), 3147 (2019).
[33] Chen, J.-S. and Chu, D. P., “Improved layer-based method for rapid hologram generation and real-time interactive holographic display applications,” Opt. Express 23(14), 18143–18155 (2015).
[34] Tastemur, O. A. and Newman, A. J., “Dynamic Layer Allocation: Analysis of the trade-off between accommodation accuracy versus computational load in computer-generated holography,” (Under Preparation).
[35] Beams, R., Kim, A. S. and Badano, A., “Transverse chromatic aberration in virtual reality head-mounted displays,” Opt. Express 27(18), 24877–24884 (2019).
[36] Ghebremichael, F., Andersen, G. P. and Gurley, K. S., “Holography-based wavefront sensing,” Appl. Opt. 47(4), A62–A69 (2008).
[37] Porfirev, A. P. and Khonina, S. N., “Experimental investigation of multi-order diffractive optical elements matched with two types of Zernike functions,” Opt. Technol. Telecommun. 2015 9807, 98070E, International Society for Optics and Photonics (2016).
[38] Kaczorowski, A., “Adaptive Aberration Correction for Holographic Projectors,” PhD thesis, University of Cambridge (2017).
[39] Jang, C., Bang, K., Moon, S., Kim, J., Lee, S. and Lee, B., “Retinal 3D: augmented reality near-eye display via pupil-tracked light field projection on retina,” ACM Trans. Graph. TOG 36(6), 190:1–190:13 (2017).
[40] Kaczorowski, A., Gordon, G. S. D. and Wilkinson, T. D., “Adaptive, spatially-varying aberration correction for real-time holographic projectors,” Opt. Express 24(14), 15742–15756 (2016).
[41] Kaczorowski, A., Gordon, G. S., Palani, A., Czerniawski, S. and Wilkinson, T. D., “Optimization-Based Adaptive Optical Correction for Holographic Projectors,” J. Disp. Technol. 11(7), 596–603 (2015).
[42] Choi, J., Park, H. J., Paek, J., Balan, R. K. and Ko, J. G., “LPGL: Low-power graphics library for mobile AR headsets,” MobiSys 2019 - Proc. 17th Annu. Int. Conf. Mob. Syst. Appl. Serv., 155–167, Association for Computing Machinery, Inc (2019).
[43] Milder, P., Franchetti, F., Hoe, J. C., and Püschel, M., "Computer generation of hardware for linear digital signal processing transforms." ACM Transactions on Design Automation of Electronic Systems (TODAES) 17, no. 2 (2012): 1-33.