Consumer AR Technologies in the Market

Kingsley, Clayton[a], Mansuripur, Masud[a]

[a]Wyant College of Optical Sciences, University of Arizona, 1630 E University Blvd. Tucson, AZ 85712

Committee Chair x______________________________

Dr. Masud Mansuripur

Committee Member x______________________________

Dr. Matt Dubin

Committee Member x______________________________

Dr. Yuzuru Takashima

ABSTRACT

The AR world was kickstarted 10 years ago by the announcement of Google Glass. In the decade since, the industry has grown to almost two billion dollars yet provided nearly nothing to show for it at a consumer level. Form and cost constraints remain the toughest obstacles for companies in the industry. Here, we analyze the technology necessary for AR systems to succeed in the market, demonstrate a proof-of-concept simulation proving that size reduction at cost is possible, and analyze the AR market in its current state.

Keywords: XR/AR/MR, CGH, MEMS-based laser projection, pupil tracking, waveguide combiners

1. INTRODUCTION

The emergence of extended reality (XR) is a fast-moving space in technology where traditional problems in optical, electrical, and mechanical engineering are even further intertwined than in most modern systems. From Meta's Oculus, to Microsoft's Hololens, to social media like Snapchat, the effort to combine the analog and digital worlds on our retinas is widespread. It can be well conveyed as a spectrum, as shown below in Figure 1.

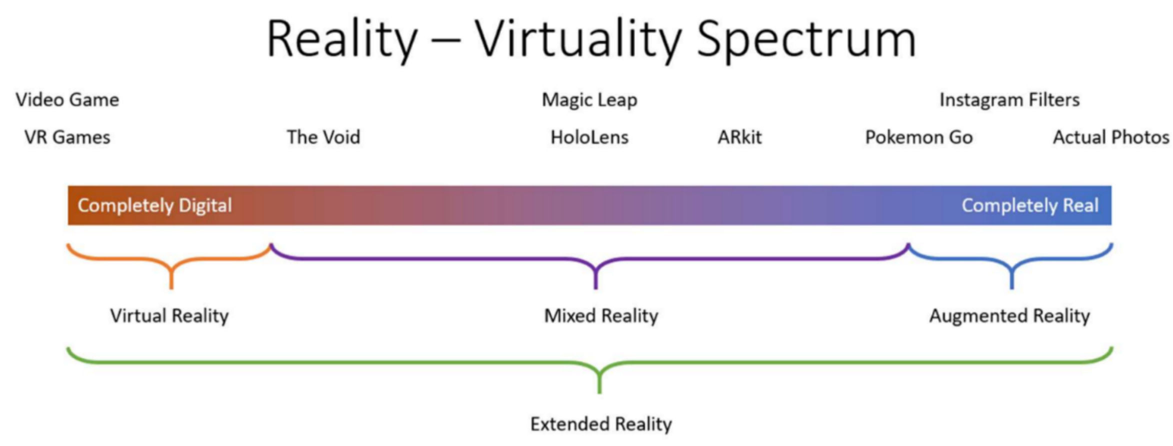


Figure 1. The reality-virtuality spectrum with examples from industry

This report will focus on the mixed reality (MR) portion of this spectrum and will use the terms MR and AR interchangeably, as they both refer to instances where the user's field of view is a combination of real-world information and digitally-created, -altered, and -overlayed media. The technology needed to enable widespread MR success at a consumer level is still immature, mostly in size and cost. These driving factors have led the industry to focus early efforts primarily in the commercial sector, where budgets are larger and form-factor less important. Widjanarko, et al. [1] summarize the pain-point design parameters of modern XR systems as follows:

- 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, color fidelity, reflections); and

- Spatial mapping, head tracking and latency.

Of the dozens of companies currently occupying the commercial AR space roughly half of them focus on the near inevitable ubiquity of wearable, near-eye displays while the other half trade-off potential mass adoption for specialization. This specialization allows companies looser design constraints and is due largely in part to the lack of maturity of the technology required to create a robust AR system suitable for day-to-day use. While products like Microsoft Hololens, Snapchat Spectacles, Lenovo ThinkReality A3, Epson Moverio BT300, Magic Leap 1, and Everysight Raptor differ dramatically in price, size, and functionality, the primary motive remains: overlay images into users' analog field of view.

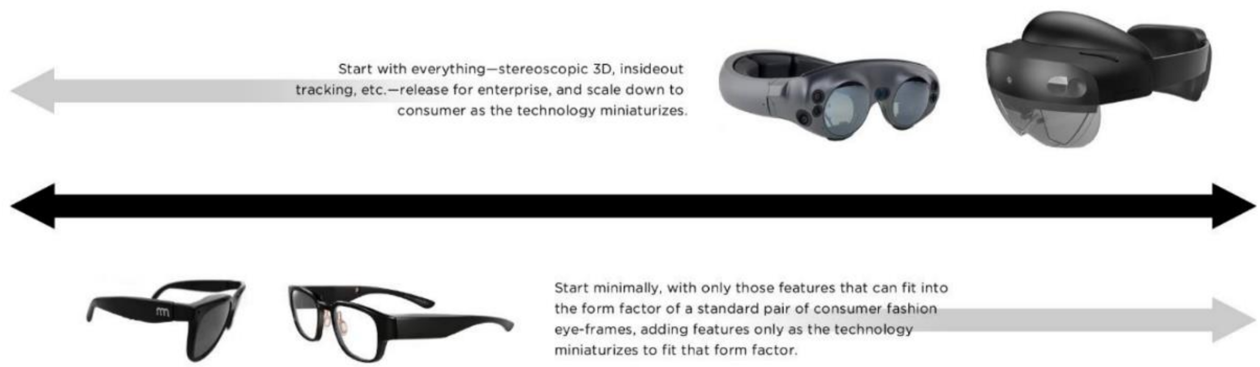


Figure 2. Two different product strategies of AR and MR [2]. Devices on the right side offer highest performance of the state-of the-art technology and become consumer friendly only if the technology miniaturizes. Devices on the left are minimalistic. Features are added step by step only when they can be included into the device within a consumer viable form factor.

The interplay of the human visual system, optics, and the natural world has never been so dynamic as it is when designing wearable AR. Weight plays a significant role in the ability of a product to be worn all day. Size, possibly the most difficult hurdle in the industry right now, is largely socially motivated, as the "visionary" systems, like those promised by Google 10 years ago, are to be worn socially. Furthermore, devices must be efficient, as excessive heat dissipation near the head will cause near immediate user discomfort. Another consideration that proves largely technically difficult is the variation of human interpupillary distances (IPD's). For this reason, systems must have large eye-boxes (exit pupils) to accommodate for differences amongst the population [3]. As we will discuss, the design necessities and constraints of HMD's are often contradictory, forcing teams to use ever-more creative approaches to further the market.

2. ESSENTIAL TECHNOLOGIES IN THE XR MARKET

2.1 MEMS-based Laser Projection

Micro-electro-mechanical systems (MEMS) are already popular in projection applications worldwide. One advantage of laser-based scanning with a MEMS device is the lack of a traditional object plane. This means that the law of etendue does not explicitly limit the system [4]. In comparison to panel-based displays, like OLED[5,6] or LCOS[7], laser based displays utilize temporal light modulation in conjunction with the pointing mirror to achieve pixel-by-pixel image creation. This allows for better performance in terms of brightness, contrast, power consumption, FOV, and distortion. [8,9,10,11]. The size—crucial trait—of a MEMS-based system can be decreased by utilizing one 2-axis mirror instead of two one-axis mirrors.

Additionally, employing a vacuum-sealed mirror, rather than one that operates at ambient pressure, can reduce space allocation for mirror actuators. The vacuum packaging also serves to reduce air-damping, aiding in power consumption and heat dissipation. Taking advantage of resonant actuation is another way to decrease the power needed to modulate light using MEMS devices. The most common methods employed to drive a MEMS scanner are electrostatic, electromagnetic, and piezoelectric. An electrostatic driver may employ a "capacitor comb" structure, where one row of capacitors is attached to the moving structure and another row is fixed to the frame. In this case, a voltage is applied across the row of fixed capacitors and the row of moving capacitors is attracted by electrostatic forces [4]. Electromagnetic actuation employs a similar method using inductors and varying current. One advantage of the electromagnetic case is the ability to produce both attractive and repulsive forces using the same actuator. Finally, the mirror may be driven by manipulating a piezoelectric material's thickness by varying the voltage across it. In each case, air resistance plays the largest damping role, and so it is crucial for the device to be vacuum sealed.

2.1.1 MEMS Packaging

With the necessity of a vacuum-sealed MEMS established, it is worthwhile to look into the optical challenges created by this packaging. The most obvious solution is to place a flat, transparent plate over the mirror. This leads to ghost images and a static parasitic reflection that causes a bright spot in the image (Figure 3, left) Naturally, then, one may tilt this cap to ensure that the static reflection is outside the exit pupil. This still does not alleviate or, more critically, control the ghost images caused by multiple reflections off of the cap's surfaces. (Figure 3, right) Due to the broad spectrum incident on a single HMD RGB display system, anti-reflective coatings are of little use here. One particularly novel solution, proposed by the firm Oqmented and dubbed "Bubble MEMS" packaging is to implement a spherical dome (Figure 3, center).

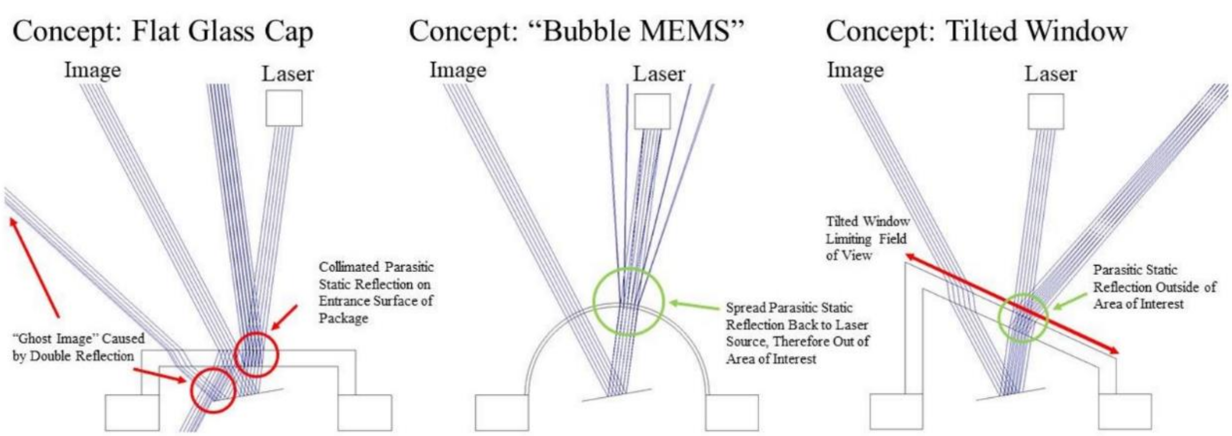


Figure 3. Comparison of different concepts for a hermetic glass package with respect to the reflection of incident laser beam. Left: Flat glass cap; Middle: Spherically shaped "Bubble MEMS"; Right: Tilted Window. Using a domed concept, the parasitic static reflection is directed back to the laser source and is always out of the area of interest.

The reflection off a spherical dome is in the direction of the source, therefore decreasing formation of ghost images and, more critically, dumping the reflected light to an area that is inherently not a part of the image, no matter where the source is incident from. The bubble, then, is a very versatile packaging style. This dome, though, acts as a meniscus lens with negative optical power. While the incredibly thin nature of the dome makes the imparted beam divergence minimal, it is not negligible. Still, because the mirror is in the center of the domed cap, the effects of the optical power of the "bubble" are independent of mirror angle. Thus, the known effects of the dome can be compensated for without increasing the complexity of the system by appropriately choosing one's collimating elements. In larger-scale manufacturing applications, for example where the light source sub-architecture is outsourced, the bubble's effects can be compensated for with a single lens after reflection from the mirror. Of course, this comes at system-level tradeoff of size and weight.

2.1.2 MEMS Scanning Methods

There are two main scanning methods used in MEMS-based displays: raster and bi-resonant scanning. Raster scanning is traditional to the display industry and uses one fast axis, usually the horizontal axis, driven at resonance. The slow axis, usually the vertical axis, steps up and down across the image in a sequential, zigzagging pattern. The more emerging method, bi-resonant scanning, or Lissajous scanning [12], drives both the vertical and horizontal axes at resonance. This allows for greater FOV, particularly in the vertical direction, since driving the mirror at resonance is much more efficient than stepping it, as with raster scanning. The Lissajous scanning method requires a particular ratio between the resonance frequencies of the axes. This ratio is determined by the FOV of the system. Both methods, illustrated below, suffer from varying line densities at the edges. This line density difference is especially prevalent when Lissajous scanning is employed. Still, in most cases, such as for AR displays, the need for a larger FOV and faster refresh time outweighs this relative density issue.

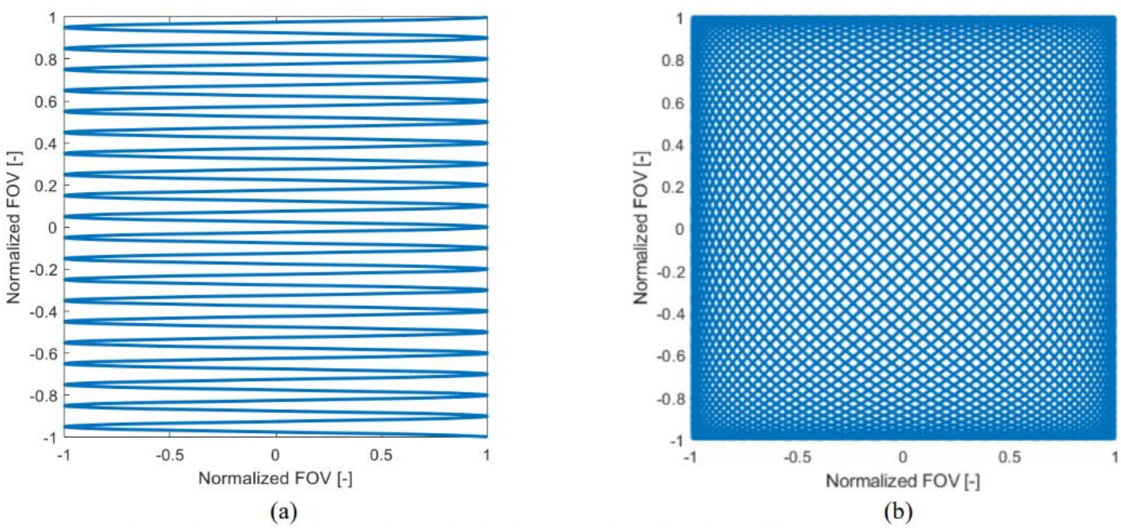


Figure 4. Comparison of two methods for pixel painting in LB micro-displays with (a) the raster scanning method where the horizontal axis is driven in resonance and the vertical axis in linear movement and (b) the bi-resonant Lissajous scanning method where both axes are driven in carefully chosen resonance frequencies.

As mentioned, there are two dominant mirror configurations. Either two 1D mirrors or one 2D mirror will be employed. In the case of two mirrors, raster scanning generally creates the image, though two mirrors may employ bi-resonant scanning. The first mirror will generally have a smaller diameter so as to allow for faster scanning. It will then reflect the 1D image onto the second, larger, slower mirror that will add the orthogonal dimension. This method is limited by the size-to-speed tradeoff of the second mirror [13]. Additionally, the use of two mirrors increases the size and weight of the system and requires more complex driving systems. One 2D MEMS mirror may be optimized for either raster or bi-resonant scanning. In either case, the use of a 2D mirror chooses the middle ground in the size vs speed trade. The advantages of having only one mirror are obvious: a more compact system, simpler driving electronics, and lower power consumption [4].

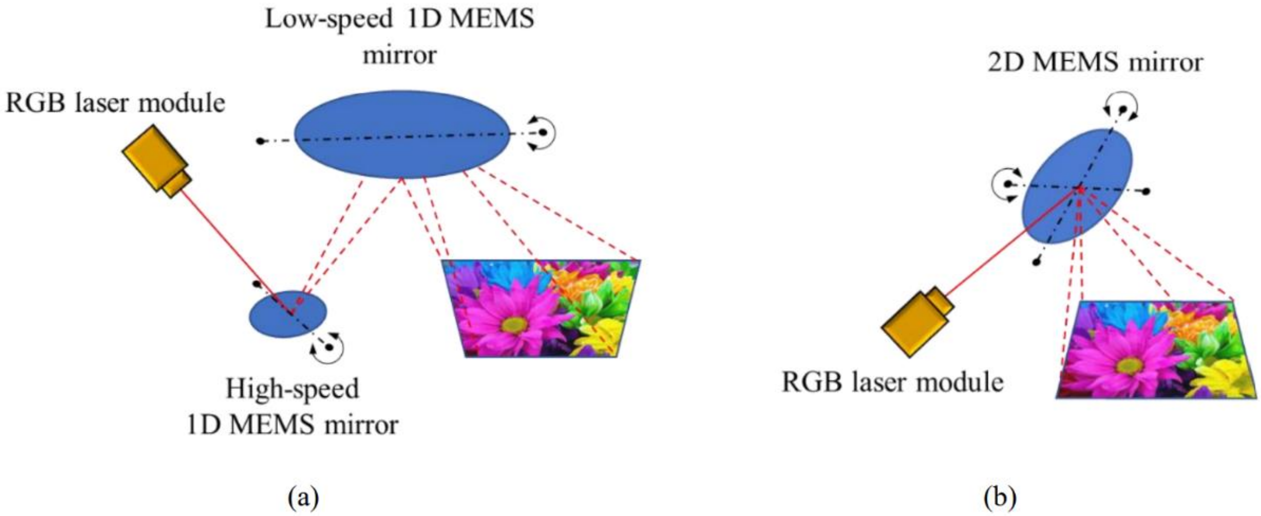


Figure 5. (a) LBS micro-display architecture based on two separate 1D MEMS mirrors and (b) architecture based on only one 2D MEMS mirror capable of moving independently in two axes.

In either configuration, image distortion is an inherent part of a MEMS image. This can be corrected for later in the system using lenses, CGH's, and other optical elements, of course at the cost of an increase in size and complexity.

2.2 Computer Generated Holograms

Holography is a lenseless imaging technique that can completely reconstruct optical fields [14]. Early holograms were single-use, as the sensing materials used were similar to film cameras. The advent of digital holography has allowed for robust, although computationally intensive, new imaging systems [15]. Today's CGH's are algorithmically generated. Their construction can be divided into three parts [14]:

1. Calculate: to allow the computer to digitally, instead of optically, calculate the interference fringes for a target object;
2. Encode: to determine the method to represent or encode the computation results;
3. Display: to display the encoded fringes on a suitable medium. [16]

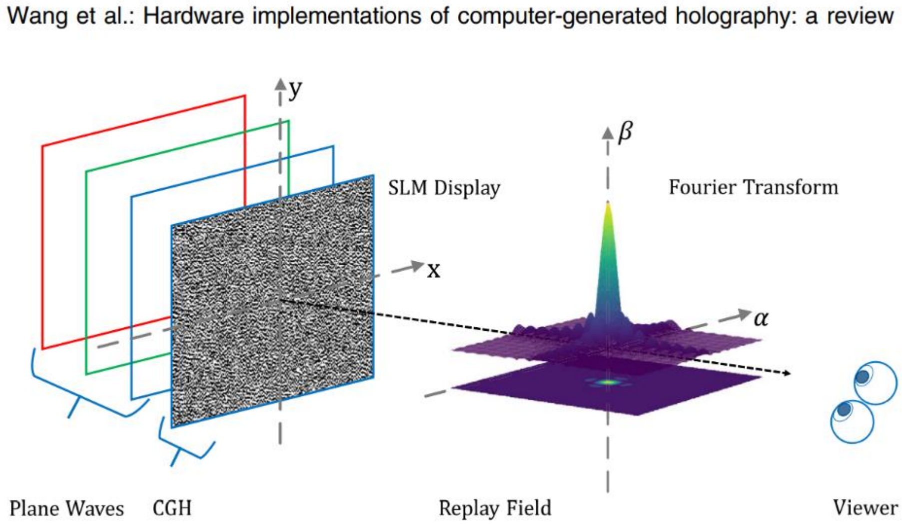


Figure 6. A typical system for CGH consisting of three main components: a light source, a computer or hardware platform for interference pattern calculation, and a device to display the hologram.[17]

Since CGH's require coherent light, they are almost always used with a laser or laser diode. This poses problems in the form of speckle and other artifacts. For this reason, CHG's are commonly combined with another spatial light modulator to counteract these artifacts; this inherently degrades the overall form factor of the system. It is the immense computational power, though, that poses the strongest challenge towards CGH's entrance into mainstream technology. Still, CGH's are commonly used in XR systems for a variety of applications including to increase ambient contrast ratio, minimize aberrations, and decrease vergence accommodation conflict:

2.2.1 Ambient Contrast Ratio & CGH's

In a purely display-based system, contrast can be easily defined as the as the ratio of the brightest to darkest adjacent points in the display. In AR applications, however, the images displayed must also combat with the high dynamic ranges of both the human visual system and the natural environment of day-to-day life. In turn, to contend with the brightest of conditions, the source of an AR display must be capable of brightness close to ambient sunlight, or 10,000 cd/m^2 [18]. This comes at the expense of heat, which we have established as a user-defined constraint. This is especially a problem when the image is pixel-sparse. Pixel-sparse images are ones whose primary content is real-world information. Given the common applications of AR in annotating and accenting the user's analog field of view, we can expect pixel-sparse images to occur frequently. In an amplitude-modulating design, such as liquid crystals, pixel-sparse images are especially inefficient with regards to total throughput and heat. One way to minimize the amount of light needed, and therefore heat, especially for pixel-sparse images, is to use a phase modulator rather than an amplitude modulator. CGH's are an up-and-coming method of phase modulation that have real promise in the MR space. CGH's require coherent light, so are typically used with laser sources. VividQ, an English company, has recently provided experimental data showing that a laser-based, CGH-modulated, near-eye projection system can provide over 10,000 cd/m^2 at less than 5mW of laser power. That is to say that images could be overlayed on a bright-sky scene using an eye safe level of optical power [18]. This has not been done with LED's or OLED's [19].

2.2.2 Aberrations & CGH's

If we consider aberrations to be essentially disturbances in the phase of a propagating light wave, then well-characterized aberrations may be corrected for with a CGH. The hologram simply needs to have the opposite phase disturbances as were imparted by the lens system we wish to correct for. Due to holography's additive (linear Fourier Transform) nature, field invariant aberrations can even be corrected for after the system is assembled [20]. This is incredibly efficient for XR systems already employing holography as a means of image formation [21]. It has also been shown that field-dependent aberrations, like distortion, can be reduced using CGH's, but at the expense of an increase in computational complications [22].

2.2.3 Vergence Accommodation Conflict and CGH's

Vergence-Accommodation Conflict (VAC) occurs when there is a discrepancy between the stereoscopic and focal depth cues sent to the brain. The stereoscopic cue is formed by the brains stitching of images from each retina while the focal cue is determined by the crystalline lens' accommodation. Typically, the stereoscopic cue wins out and, overall, depth can be determined. VAC is still a hefty problem as it causes visual discomfort and quick visual fatigue [23]. It has also been shown that VAC proves even more bothersome in moving scenes than static ones [24]. This is a common issue when dealing with virtual near eye images, as the stereoscopic cue will tell the brain where the analog scene is, but the focal accommodation cue will imply the distance to the projected image's focal plane. This becomes unsolvable with a 2D image when multiple parts of the analog scene are overlayed. 3D CGH's can be used to create multiple focal planes, allowing the projected image to lie in multiple focal planes provides the ability for stereoscopic and focal accommodation cues to agree on the presence of the overlayed digital image in the depth of field. While the focus accommodation of the human eye is a continuous 15D [25], a CGH can only create a finite number of focal planes. It has been shown that sequential depth layers of 0.15D can provide a perceivably smooth depth of focus [26]. Thus, for a CGH to entirely nullify VAC, 100 layers would be needed. Modern studies use only around 10 layers [27]. Thus, in dynamic cases, algorithms must decide where depth detail is needed on a frame-by-frame basis [26]. Again, this competes with an HMD's need to be light-weight and heat efficient. For exactly this reason, VividQ has developed a dynamic layer allocation algorithm that allows scenes to operate with as few 3D layers as possible depending on the depth of focus of the image being projected. Table 1 [1], below, shows the increase in complexity per frame when more layers are added.

TABLE 1. FLOPS REQUIRED FOR NUMBERS OF LAYERS USING DLA ALGORITHM [1]

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

2.3 Waveguides Combiners

Waveguide combiners are based on TIR propagation of the entire field in an optical guide, acting basically as a transparent periscope with a single entrance pupil and often many exit pupils. The crucial elements of a waveguide combiner are the input and output couplers. These couplers can take the form of simple prisms, embedded mirror arrays, analog holographic gratings, or other optical elements [28]. One can take the concept of a flat waveguide with single curved extractor mirror output coupler or freeform prism combiner—or even a curved waveguide with a curved mirror—to the next level by multiplying the mirrors to increase the eyebox or by fracturing metal mirrors into individual pieces. These methods are shown below in Figure 7. Other methods include grating and holographic waveguides. The development of waveguides, and particularly the bending of them, is the birthplace of heads-up projection. When the waveguide gets curved, though, everything gets much more complex, and the extractor mirror or lenses need to also compensate for the power imprinted on the TIR field at each TIR bounce in the guide.

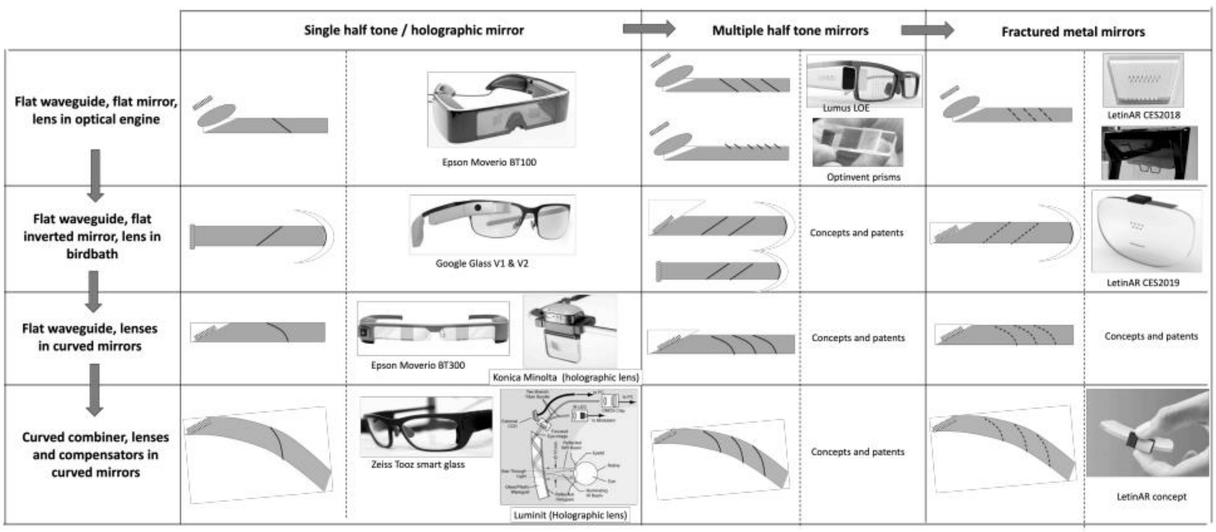


Figure 7: Multiplying or fracturing the extractor mirrors in flat or curved waveguides [28]

This coupling element is the most crucial feature of a waveguide combiner as the index of the coupling structure—if it is a hologram or grating and not a mirror—dictates the color uniformity over the field of the eyebox. The most obvious coupler is a prism with a slanted edge. Prisms make fine input couplers but are less effective as output couplers. One may use a prism as the input coupler with embedded cascaded mirrors with partially reflective coatings as the output coupler. When the output coupler is a reflective element, as in this case, excellent color uniformity can be achieved. As the see-through nature of a HUD is paramount, embedded cascaded mirrors have obvious drawbacks. Alternatively, microprism arrays as outcouplers can be index-matched to produce an unaltered see-through experience. [29]

Non-traditional optical elements are also used as combiners in industry, namely diffractive and holographic couplers. Thin reflective holographic elements can be used as both in and out couplers in waveguide combiners. This can be incredibly advantageous in the case of one full color hologram based on 3 phase multiplexed single color holograms. This architecture allows for a single plate waveguide, reducing weight, size, and cost. These holograms, though, do have quite low efficiency compared to their single-color counterparts [28] When the angular and spectral bandwidths are large, like for a large FOV or for RGB LED's, the linearly additive nature of holography allows thicker holographic combiners to be employed. The advantage in this case is that phase multiplexing can occur over many different holograms one on top of each other, allowing for multiple Bragg conditions to operate in concert to build up a wide synthetic spectral and/or angular bandwidth [30]. It is interesting to note that although the individual holograms acting in slivers of angular and spectral bandwidths alter the incoming spectrum as any other hologram, the spectral spread over the limited spectral range of the hologram is not wide enough to alter the MTF of the immersive image, and thus does not need to be compensated by a symmetric in- and out-coupler as with all other grating or holographic structures. This allows this waveguide architecture to be asymmetric, such as having a strong in-coupler as a simple prism [28]. Figure 8, below, shows the operation of thin and thick holograms used as combiners in reflective and transmissive modes.

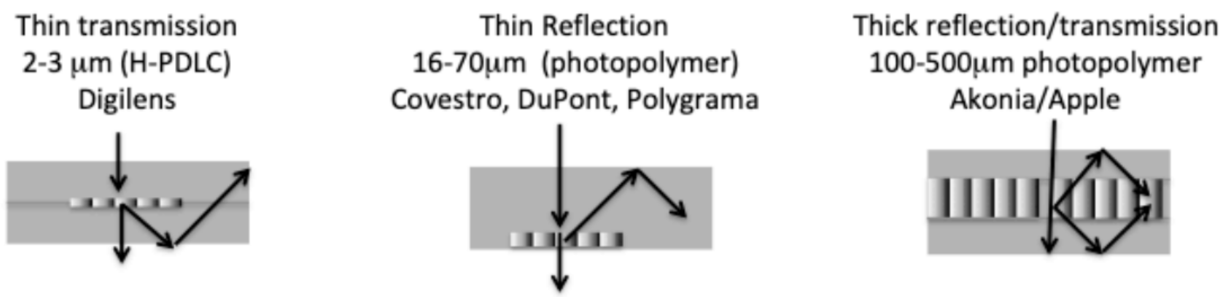


Figure 8: Different types of volume holograms acting as in and out couplers in waveguide combiners

2.4 Pupil & Eye Tracking

The ability to track a human's line of sight (LOS) is incredibly desirable and has been researched extensively in recent years [31]. In a non-display context, LOS tracking can be implemented to advance national security and or to further consumer understanding in brick-and-mortar establishments. In display systems, LOS tracking provides the ability for the display to be interactive with hands-free control capability. The advancement of HMD technology has further propagated the interest in display-context LOS tracking. While HMD's have been employing head-tracking systems to approximate LOS, this is not very robust. When the user changes their focus to a laterally faraway portion of the field, the user's eyes will complete movement before the head, meaning that there is delay in the head-tracking system. If the user moves focus laterally to a nearer point in the field, the head will not move at all, and no LOS change will be detected. Furthermore, if the HMD slips or moves at all on the users head, the LOS calculation will fail. Thus, it is desirable to track both head movement and eye movement in an HMD.

Eye tracking methods can largely be characterized as imaging or non-imaging. In both cases, methods largely disregard the eye's ability to move about the visual axis and instead track only the eyes horizontal and vertical movement. Non-imaging methods rely on the electromagnetic signals involved in eye movement. Methods include electro-oculography, where transducers are placed on either side of the user's head [32], and the sclera search coil method, in which an induction coil or contact lens is implanted in a user's eye [33]. Due to the invasive nature of both methods, they are not popular. Thus, researchers have taken advantage of the eye under near infrared (NIR) illumination. NIR illumination is advantageous because of its lack of perception by the user and the eyes reflective response to these wavelengths. Notably, the first Purkinje image, created by the anterior corneal surface is known to create a glint that can be tracked algorithmically in an on- or off-axis LED setting. One issue associated with glint tracking is the loss of illumination with high angles of eye rotation and a tight trade-off between field of view and resolution. Additionally, tracking only the glint does not allow for head-movement. To combat illumination and resolution issues, researchers have proposed multiple LED configurations and pan-tit-zoom camera configurations [34, 35]. To allow head movement within the system, one can track multiple eye features or an eye feature and a facial feature. Adding LED's and camera hardware compromise the compactness and weight of the system. Tracking a facial feature in addition to an eye feature requires a greater field of view, which decreases resolution. Still, it can be shown that tracking the glint with a single LED can provide meaningful LOS estimation results.

3. A PHYSICAL OPTICS-BASED PUPIL TRACKING SIMULATION

3.1 Motivation

The motivation of this simulation is to verify in a physical optics engine the results obtained by a purely mathematical simulation performed by Hua et. al. [31]. While prior research [34] focuses primarily on a four LED design, this simulation is significant because it provides proof of concept of single-LED configuration for pupil and head tracking. Furthermore, the LED can be placed on or off axis, allowing for multiple architectures and intricate design choices. We present analysis of eye-feature movement for one LED on and off axis including resolution, speed, and range of the configuration. The physical optics-based results show expected differences from the purely mathematical simulation due to the inclusion of realistic LED modeling, aspheric eye features, and higher order optical phenomena such as scattering and ray-splitting.

3.2 Hua's Mathematical Model

The mathematical model we here seek to extend into a physical optics simulation was performed in MATLAB under ideal conditions. It utilized the Arizona Eye Model [25] with conic constants ignored, a rotational axis of 12.5 mm, and a pupil diameter of 4mm. Aside from the differences in eye modeling, the geometry of the CCD, LED and imaging system are consistent in the mathematical model and my physical simulation. The geometrical configuration is detailed in Figure 9 below.

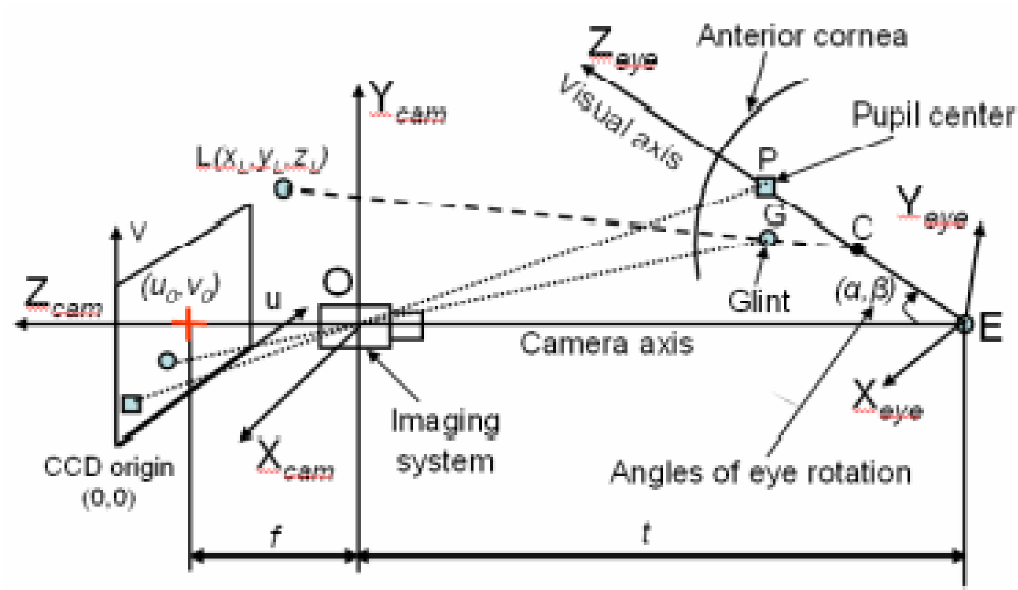


Figure 9. Geometry of coordinate systems. [31]

The results of Hua's simulation which we wish to test show slight nonlinearity for glint crated by an on axis LED and high non-linearity for glint's created by LED's off axis. The resolution achieved by this simulation is ~.7-1 degrees per pixel for an on axis LED within +-20 degrees of diagonal eye movement (ie. $\alpha=\beta=20$deg). [31]

3.3 The Eye Model

The eye model utilized in the Zemax simulation consists of the nonsequential model provided by OpticStudio. It consists of 6 elements: the cornea, anterior chamber, iris, vitreous chamber, lens, and retina. The key differences between Hua's model and this model are the inclusion of conic constants, the reflective and transmissive properties of the elements, and the axis of rotation of the eye. The model is shown below in Figure 10 while Table 2 shows the reflective and transmissive properties of all elements at 780nm. The ocular transmission data was collected from the experimental results of Boettner and Wolter [36].

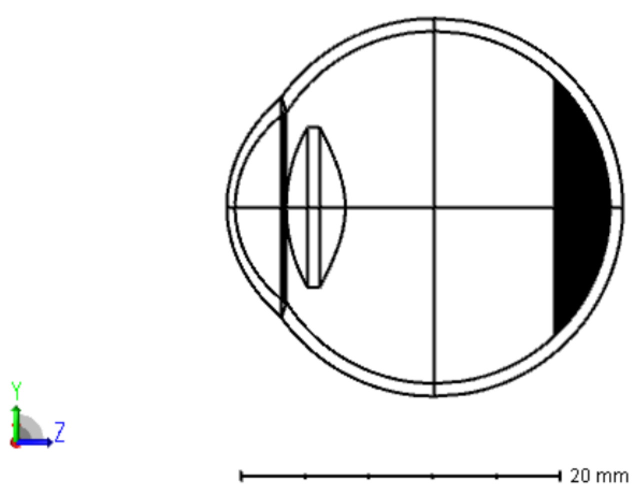


Figure 10. The nonsequential eye model used in Zemax simulation.

```
IDEAL cornea .93 .07

IDEAL aqueous .96 .04

IDEAL lens .95 .05

IDEAL vitreous .93 .07
```

Table 2. The reflectance and transmittance of each eye elment at 780nm. Given in each row as ‘IDEAL <element> <transmittance> <reflectance>’

3.4 Illumination and imaging model

The imaging optics consist of one lens of focal length 10mm located on the eye’s optical axis 50mm from the anterior corneal surface and a CCD detector located at the focal point of this lens. The focal length of the lens was chosen to allow a field of view just large enough to image the eye at 30deg diagonal rotation. The area of interest in the plane of the eye is defined to be a rectangle of 32mm by 24mm, based on general population data. One could increase this area to track a facial feature or accommodate a larger population, but that would decrease the tracking resolution of the system. Aberrations due to a real lens were not present in Hua’s mathematical model and therefore must be considered. The aberrations inherent in the lens are shown in Figure 11. The only notable aberration is spherical aberration which poses no problem to tracking as the glint is tracked as the center point of the first Purkinje image. The CCD is a standard 1/3-inch NTSC detector with a width of 4.8mm and a height of 3.6mm. It has 640 pixels horizontally and 480 pixels vertically. In Zemax, this is modeled as a detector surface with the parameters specified above. The lens is a direct simulation of a commercial-off-the-shelf (COTS) lens manufactured by Archer Optics. The choice to use COTS parts is intentional so as to allow future research to easily replicate this exact simulation. Table 3 below provides the prescription for the imaging lens used.

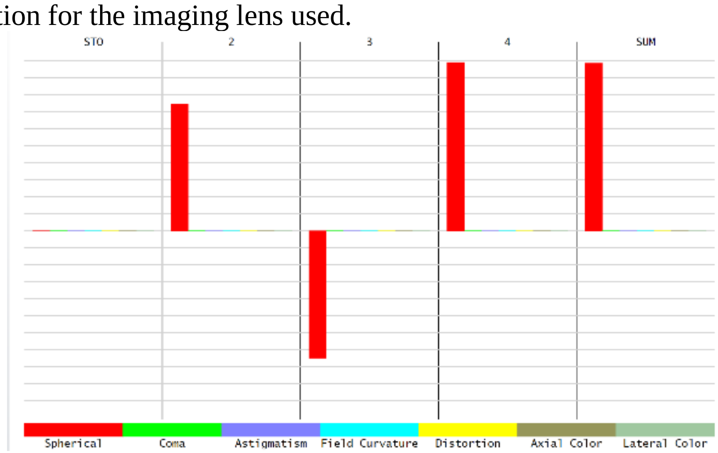


Figure 11. Characterization of aberrations in imaging lens.

	Surface Type		Commen	Radius	Thickness	Material	Coating	Clear Semi-Dia		Chip Zone	Mech Semi-Dia	Conic	TCE x 1E-6
0	OBJECT	Standard ▾		Infinity	Infinity			0.000	U	0.000	0.000	0.0...	0.000
1	STOP (aper)	Standard ▾		Infinity	0.000			3.500	U	0.000	3.500	0.0...	0.000
2	(aper)	Standard ▾	ADP-8-1...	6.087	4.500	BK7		4.000	U	0.000	4.000	0.0...	-
3	(aper)	Standard ▾		-5.100	1.200	SF5		4.000	U	0.000	4.000	0.0...	-
4	(aper)	Standard ▾		-12.460	7.035			4.000	U	0.000	4.000	0.0...	0.000
5	IMAGE	Standard ▾		Infinity	-			0.000	U	0.000	0.000	0.0...	0.000

Table 3. Archer Optics Lens APD-8-10A prescription.

To illuminate the eye, another COTS part was chosen. LED780L is manufactured by Thorlabs and is readily available. The LED's spectrum and angular distribution are shown in Figure 4. To model this LED in ZEMAX, We use the 'source radial' object type and input the relative intensity for 7 points between 0 and 40 deg. The curve was then fit by Zemax as a cubic spline. The LED is assumed to be monochromatic at 780nm to satisfy the desired NIR illumination. The size of the emitter is provided by Thorlabs and is rectangularly 4mm by 4mm.

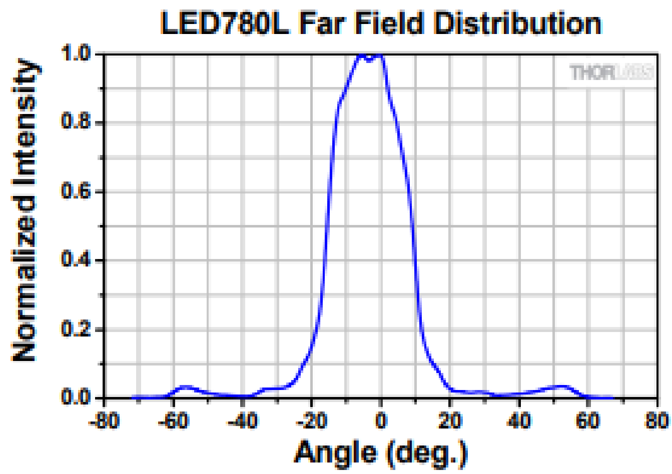
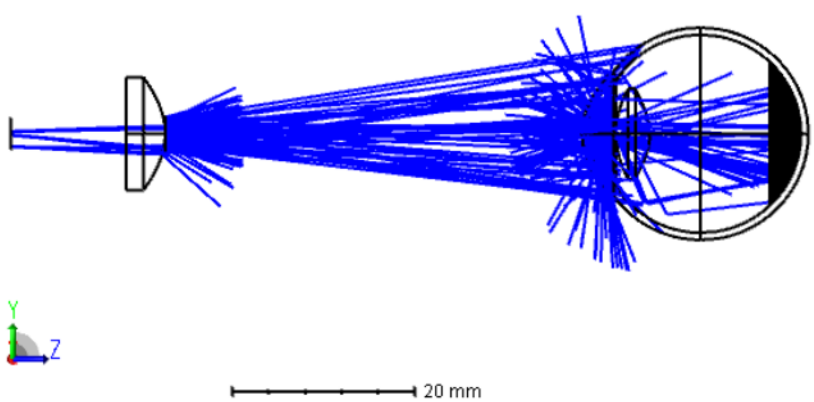
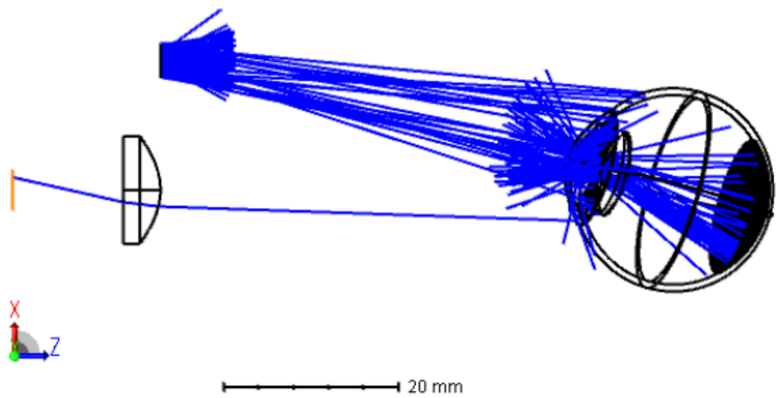


Figure 12. The angular profile of the LED used to illuminate the eye, Thorlabs LED780L.

When on axis, the LED coincides with the principal plane of the lens. When off axis, it is displaced in the +X direction. Figure 13 shows the complete setup of the simulation. It is worth noting that the number of rays traced in Figure 13 is reduced for display clarity purposes. Figure 13 does not reflect the actual number of rays incident on the detector in analysis.



a)

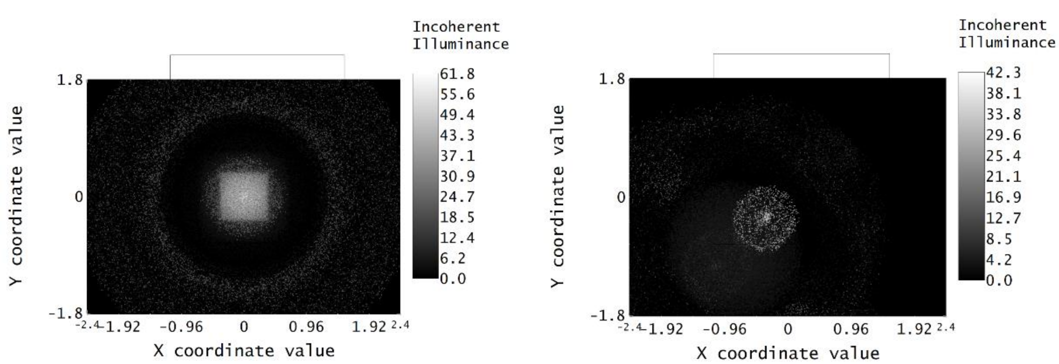


b)

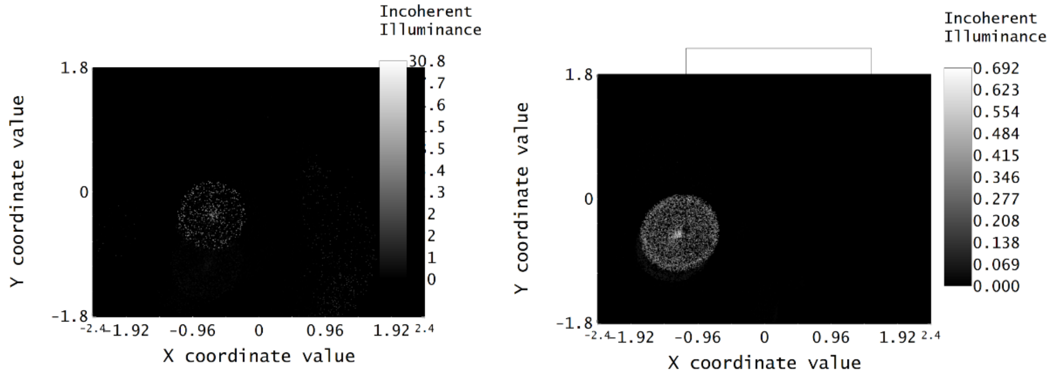
Figure 13. Complete setup of simulation. a) LED on axis b) LED off axis displaced by 10mm.

3.5 Simulation Procedure & Method for Tracking Glint.

In both cases, when the LED is on axis and off axis, the angle of eye rotation is varied diagonally from -30 deg to +30 deg in 5 deg increments. The glint location is then chosen as the (x,y) coordinate on the detector that corresponds to the center of the brightest spot. This (x,y) location is measured in mm and is easily convertible to pixels due to the known dimensions and pixel density of the detector. The brightest spot is inherently the first Purkinje image, or the glint. There is other noise on the detector from other Purkinje images, but the relative intensity of these images is very low and could be thrown away algorithmically in a physical implementation of this simulation. When the LED is off axis, the location of the illumination source is varied between 5 mm and 30 mm in the positive X direction and the angles of rotation are varied in the same manner as for the on-axis case. Figure 14 shows the CCD detector reading for several LED displacements and angles of eye rotation. In each case, the first Purkinje reflection is clearly imaged, and its center can be easily determined. In the on-axis case, for zero eye rotation, we confirm the quality of the system by noticing the dark pupil, the obvious central glint, and the narcissus image of the rectangular source with slight barrel distortion due to being reflected off of the round cornea. The relative illuminances between the detector images shown should not be considered, as the contrast enhancement has been adjusted between the collection of images for better display here. Again, this contrast adjustment could easily be performed algorithmically in the case of an actual experiment replicating this simulation.



a) b)



c) d)

Figure 14. Detector collections for various LED locations and angles of diagonal eye rotation. a) LED on axis with no eye rotation. b) LED on axis with -15deg of diagonal eye rotation. c) LED displaced 15mm off axis with -15deg of diagonal eye rotation. d) LED displaced 30mm off axis with -30deg of diagonal eye rotation.

3.6 Characterization of Glint Movement

3.6.1 LED on axis

Tracking of the first Purkinje image is nearly linear out to 30 degrees of diagonal eye rotation for an LED on axis. It is intuitive that the displacement begins to level out as the rotation of the eye increases due to the increasing obliquity of the reflections. The velocity is ‘jerky’ due to the small sample size and large angular increments. The velocity intuitively decreases off axis for the same reason that the linearity of the displacement decreases. From the data shown, the system has a glint tracking resolution of ~.5-2 deg/px for the entire range of motion. This resolution is lower than that reported in Hua’s mathematical model. The decrease in resolution can be attributed to the physical model of the imaging system, the difference in eye modeling, and the use of manual central glint detection. More specifically, the modelling of a real LED, the transmittance and reflectance of ocular elements, the aspheric nature of the eye, and consideration of phenomena such as scattering have provided degraded results compared to the tracking accuracy of the mathematical simulation. Still, we observe in the results the expected trends of linearity of displacement and decreasing velocity as the angle of eye rotation increases. The linearity of displacement is strong, with linear fit correlations of .987 and .98 for horizontal and vertical displacement respectively. Figures 15 and 16 report the results.

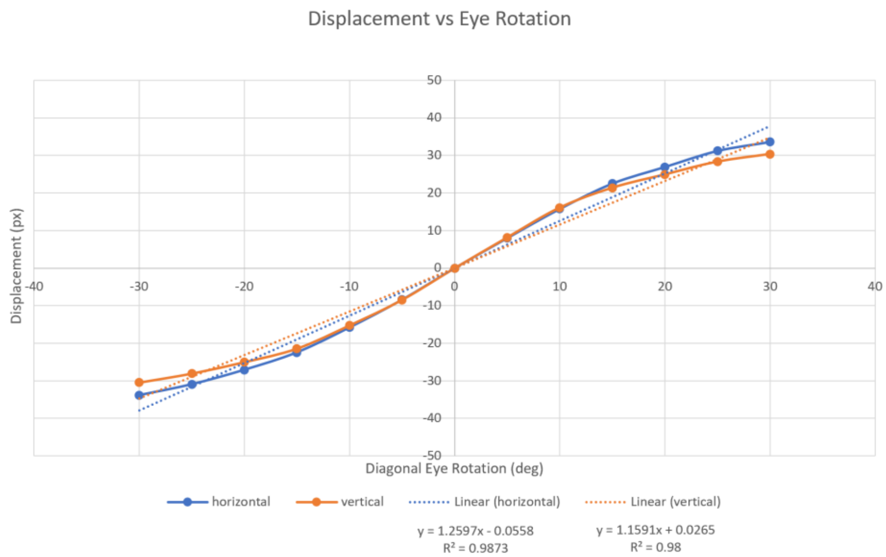


Figure 15. Glint displacement vs diagonal eye rotation for an LED on axis.

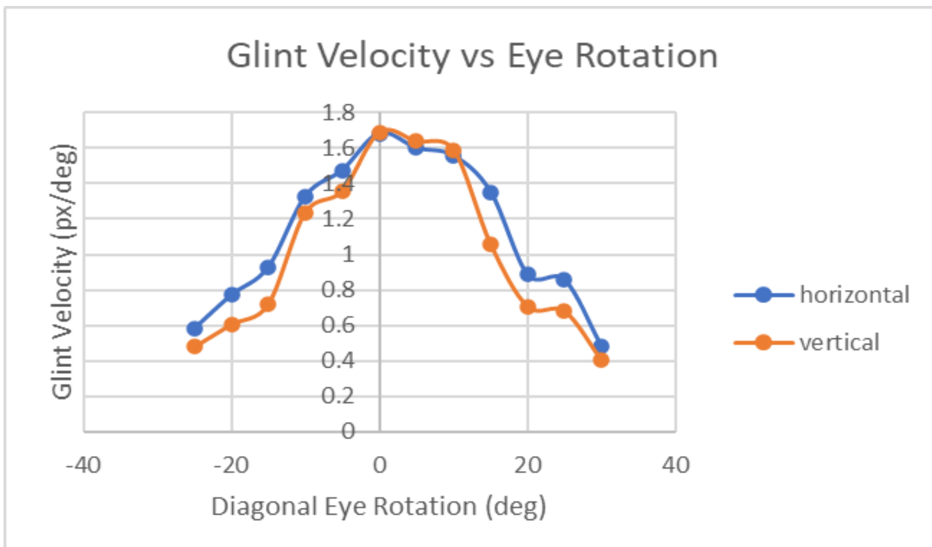
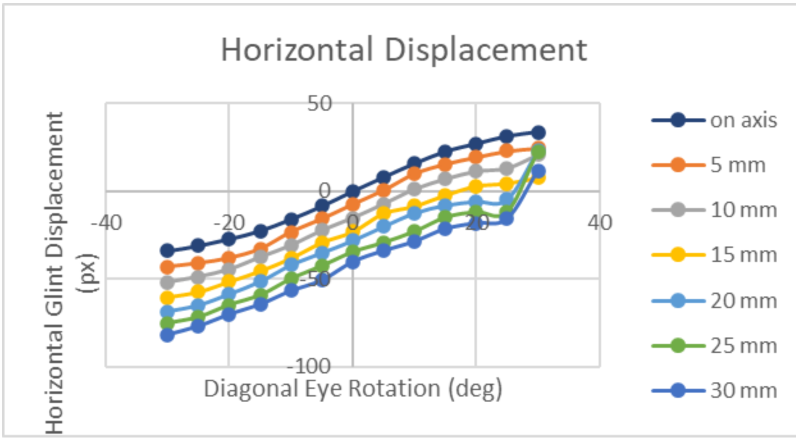


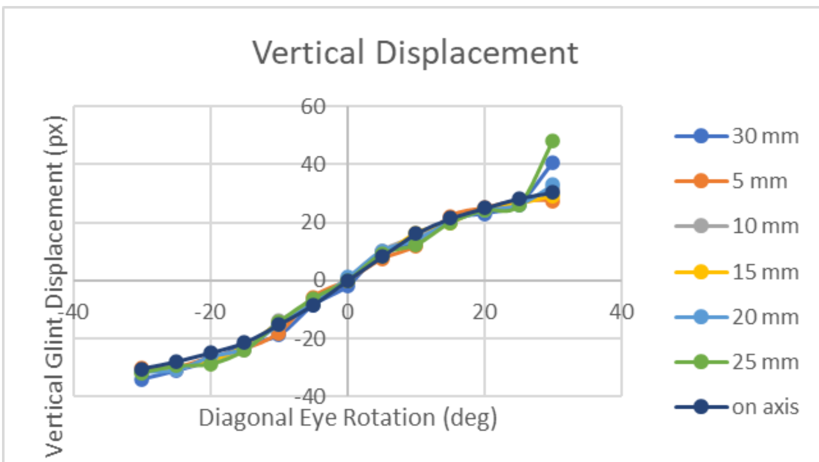
Figure 16. Glint velocity vs diatonal eye rotation for an LED on axis.

3.6.2 LED off axis

In the off-axis case, for LED displacement varying between 5mm and 30mm, the vertical glint displacement is unaffected until we observe high angles of rotation at high LED displacements. For 20mm, 25mm, and 30mm displacements, at 30deg of angular rotation, the displacement spikes outside of our on-axis model. This is likely due to off axis aberrations in the imaging lens and a sharply changing intensity profile at oblique viewing angles of the LED. For smaller horizontal LED displacements, however, measures of vertical glint are unaffected. This is intuitive as the displacement of the glint is perpendicular to the displacement of the LED. For horizontal glint measurement, we observe the same linearity as for an on-axis LED but with a displacement proportional to the displacement of the LED. Again, for larger displacements and large angles of eye rotation, the glint tracking is nonlinear and unreliable. These irregularities cannot be solved algorithmically but may be resolved with hardware changes. To improve the tracking at high angles of obliquity, one could choose an LED with a wider angular dispersion or increase the robustness of the imaging optics. If one were to increase the angular dispersion of the LED, more light would be lost and the system would be less efficient photonically. If the robustness of the imaging optics were to be improved, one would sacrifice compactness of the system. The likeness of the velocity graphs to the on-axis case for the well-behaved displacements and rotations equate to essentially no change in resolution. The non-linearity of glints observed in Hua's mathematical model is not observed here due to the method of selection of glint location detailed above. It is reasonable that this method could be implemented algorithmically in this setup if the proper LED displacement and angle of eye rotation range is chosen to keep us within the linear bounds of the data. Figures 17 and 18 show the off-axis simulation results.

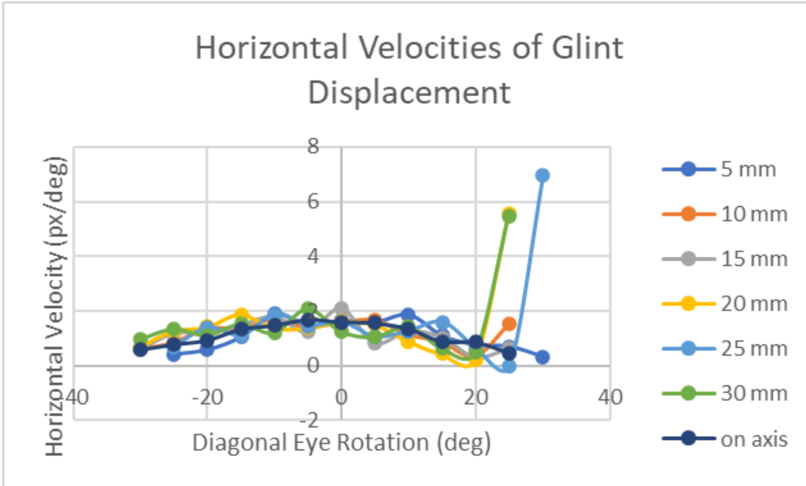


a)

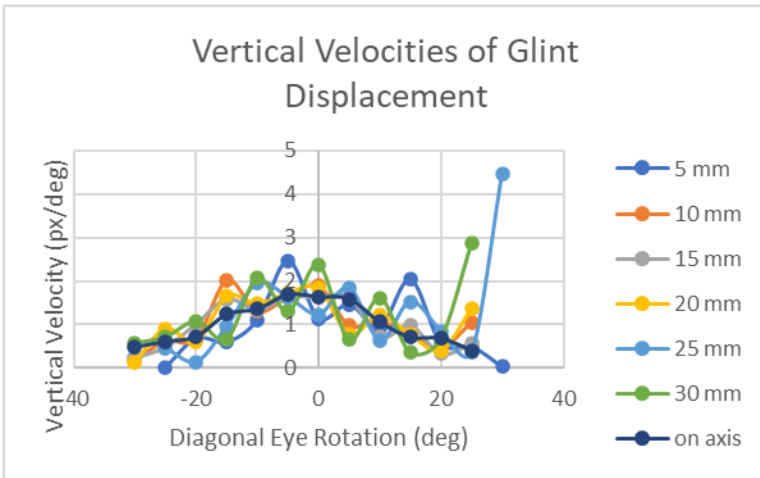


b)

Figure 17. Glint displacement for an LED off axis. a) horizontal glint displacement vs diagonal eye rotation. b) vertical glint displacement vs diagonal eye rotation.



a)



b)

Figure 18. Glint velocities for LED's off axis. a) horizontal glint velocity vs diagonal eye rotation. b) vertical glint velocity vs diagonal eye rotation.

3.7 Final Results of the Simulation

The results of this simulation validate the eye-tracking capabilities for one LED on or off axis proposed by Hua while proposing a new method of algorithmically selecting the glint in the off-axis case. The new method of selecting the glint as the centroid of the brightest spot on the detector alleviates the off-axis nonlinearity of the mathematical model. The extension of the mathematical model to a physics-based model showed little degradation in range and resolution and provided a COTS-based apparatus for the system to be tested in a real-world environment. Future work in this area should include an analysis of the effects of different COTS parts in the current configuration, an analysis of the effect of axial distance from the source and camera to the eye and comparison to Hua's mathematical model. In fine, this simulation details evidence that there is room in the market to decrease both size and cost.

THE MODERN AR MARKET

4.1 Marketspace Overview

According to a Mordor Intelligence market report, the AR market was valued at 1.98 billion USD in 2020 [37]. The largest sector of this market lies in North America, but the Pacific Asian market is expanding rapidly. The study forecasts a 151% compound annual growth rate (CAGR) for AR in 2025. According to Equation 1, this would put the market value in 2025 at 197 billion USD.

$$\boldsymbol{CAGR = \left(\frac{Final\,Value}{Starting\,Value}\right)^{\frac{1}{N}} - 1}$$

Equation 1. Given the known CAGR of 1.51, N of 5, and starting value of 1.98 billion, we can calculate the projected market value of AR in 2025.

CitiBank forecasts that the majority of the growth will come from commerce, communication, and hardware. [37] This is due to the dependence on hardware to provide the size and affordability for commerce and communication success.

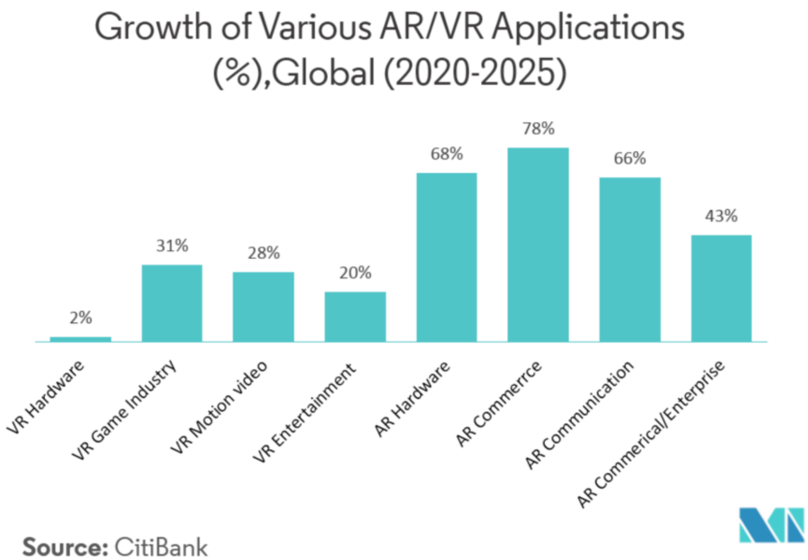


Figure 19. Growth of various XR applications from 2020-2025

Such a prediction falls precisely in line with the technological trends previously analyzed in this paper: the size and cost of hardware are the primary drivers inhibiting widespread adoption of AR technologies. Still, with advances coming, Mordor calculates that over 800 million smartphones will be equipped with AR hardware by 2025. This represents 21% of the 3.8 smartphone users worldwide in just three years’ time. As the market continues to mature, businesses and consumers will soon find exciting new ways to engage with the world around them.

4.2 Microsoft HoloLens


Figure 20. A Microsoft HoloLens user.

Microsoft's HoloLens [38] is the most robust system available for purchase today. As such it will be used as the reference point when discussing its competitors. It takes the "start with everything" approach to design, compromising on price, power consumption, size, weight, and form factor to achieve what is potentially the first truly MR system on the market. HoloLens features hand tracking, voice interaction, eye-tracking, iris-based bio-authentication, real world spatial mapping to anchor images, and software partnerships that enable a wide range of industrial applications. To achieve all of these capabilities, the environment around the user is captured and mapped using four visible-spectrum cameras, a time-of-flight sensor, an accelerometer, a gyroscope and a magnetometer. The hallmark capability of HoloLens that makes it the industry leader is its ability to anchor and continuously interact with images in the user's analog surroundings. Meanwhile, the user's line of sight is tracked with 2 IR cameras. The display is a combination of MEMS and holographic elements, making use of the technology discussed earlier in this paper to provide >2.5k light points per radian holographic density. The HMD runs on an Android OS, making application-specific collaboration and integration accessible. On the other hand, the battery life is published at 2-3 hours, immediately disqualifying it from an all-day-use device. HoloLens weighs 556g, by far the heaviest product detailed in this paper. Currently in its second generation, the HoloLens 2 retails for between 3,500 and 5,000 USD, depending on the application it's intended for: the price increases come with clean room and hard-hat safety certifications. This combination of battery life, price point, and form factor make the current iteration of Microsoft's XR system all but unusable in day-to-day life. Still, the technology and capabilities seen here are unmatched by the rest of the market.

4.3 Magic Leap 1

Figure 21. Left to right: the Magic Leap 1 headset, Lightpack, and controller

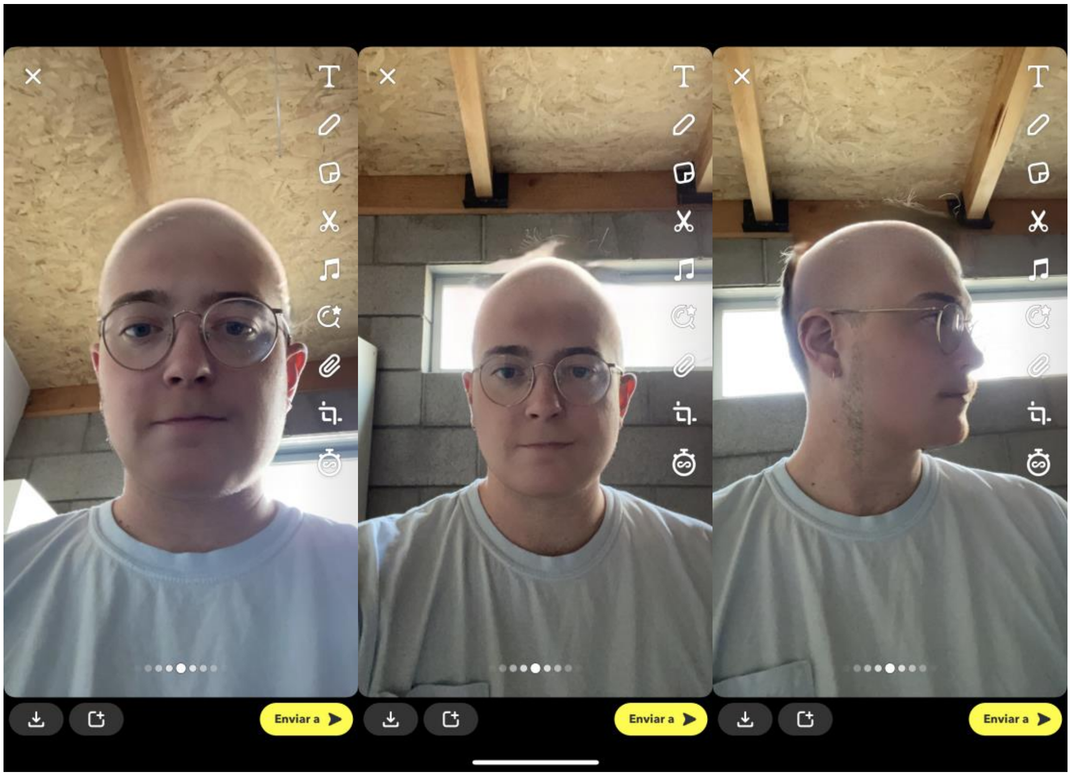
The Magic Leap 1 [39] is the most direct competitor to HoloLens, as it is intended for the same enterprise, healthcare, and manufacturing environments. The Magic Leap has a number of advantages over the benchmark. First, this system wins in price, retailing for between $2300 and $3000. This system is also lighter than its Microsoft competitor, weighing 316 grams. The weight margin is suspect, though, due to the placement of a “Lightpack” (electrical hardware) in the user’s pocket. Capabilities are similar to that of HoloLens, with head and hand tracking, voice control, and eye tracking to enable the user to interact with the digital scene. It achieves all of these features using the same technology: a mix of visible and IR cameras and a slew of local positioning devices. A major obstacle for the Magic Leap is its OS. This XR system runs on Lumin OS, making customization more difficult than with the HoloLens’s ubiquitous Android platform. Most notably, though, is the lack of spatial mapping and image anchoring. While this feature is certainly on the horizon given the outward facing cameras, accelerometer, gyroscope and magnetometer, it is not available in the Magic Leap 1. This is the key capability feature that secures HoloLens’ position at the top of the industry.

4.4 Snap Inc. Spectacles

a. b.

Figure 22. Snap Inc’s a) Spectacles 3 , b) New Spectacles

Snap Inc. is a company known for their social appeal, mainly due to their mobile photo messaging app, Snapchat. It’s no surprise that their foray into XR hardware follows the “form first” path. Already a leader in the personal amusement sector of AR, Snapchat’s mobile app allows users to apply a variety of live filters to their photos and videos, as exemplified below in Figure 23.



a. b. c.

Figure 23. Examples of a Snapchat filter that makes the user appear bald a)under normal use, b) with varied background texture, c) at a high angle of incidence

Notice above how the filter is rather robust; it blocks out the user's head and facial hair while recognizing and maintaining the presence of eyebrows. The AR works rather effectivley when the background texture is varyied and when the facial angle of incidence is near 90 degrees. This is just one example of hundreds of live AR filters Snap Inc. has already deployed. Though not demonstrable here, the filters also work in videos, which is essential for the Spectacle's performance. The Spectacles 3, an earlier version of the "New Spectacles," offer incredible form factor but not much functionality. [40] The glasses, which are avialable for consumer purchase at a designer sunglasses-friendly price of $380, are light enough to be worn all day. They weight just 56 grams. The two front-facing camers provide an 86 degree 3D full field of view, and can capture 70 videos per charge. The camera's capture photo's at 1728 x 1728px and video's at 1216 x 1216 px. This reduction suggests robust video-stabilization. The real intent here is to capture what the user is seeing in first person and add some of Snap Inc.'s AR filters after the fact using their mobile app.

The New Spectacles, however, feature a binocular 3D waveguide display, meaning they actually inhabit the live XR world. The display is capable of producing 2000 nits of brightness, sufficient for sunny outdoor activity. As seen above, this comes at the cost of form factor. Weight has also increased, but not dramatically: up to 134 grams. Environment anchoring capabilities include hand, marker, and surface tracking. The user interaction is not nearly as sophisticated as the HoloLens, utilizing only a touchpad on the side. This means that once media is overlayed into the analog scene it can no longer be interacted with in a meaningful way. Current applications are limited, as the New Spectacles are still only available in a creative partnership capacity. Still, this is the most exciting wearble XR product in the industry right now.

4.5 Google Glass

Figure 24. Google Glass

After revolutionizing the AR sector in 2012, Google glass is perhaps the most uninspired piece of tech in the industry [41]. In the balance between form and function, Glass somehow misses out on both. User interaction is limited to voice control and a touchpad on the arm of the system; there is no spatial mapping, image anchoring, or pupil tracking. The display is monocular with a 640 x 360 px resolution. The saving grace of Glass is the light weight (46 grams), making it optimal for industrial employees working long shifts.

4.6 Niche-application products

The rest of the marketspace is comprised of application-specific HMD's. The Everysight Raptor [42] is designed for cyclists. It can project speed, route, mileage, and location onto the lenses via its OLED waveguided. Front facing cameras capture the ride for film study or social sharing. The device is interfaced with via a handlebar-mounted controller and voice commands. Everysight's Raptor weighs in at 98 grams, provides 8 hours of battery life, and starts at $600.

Figure 25. a) Everysight Raptor, b) Epson Moverio, c) Lenovo ThinkReality A3

Two similar design's are Epson's Moverio [43] smart glasses and Lenovo's Think-Reality A3 [44]. Both circumvent the demands of form, contrast ratio, and dynamic interaction ability by applying themselves to the in-home market. In the home, weight and heat are of concern but social presence is not. Because they can be used at a desktop, controls and battery life are already at the user's fingertips. As such, they do not require the same all-encompassing features that a HoloLens competitor does. The Moverio glasses start at $700 while Lenovo's version is double that at $1500.

Overall, this sector of the market chooses neither of the preconceived design paths we have thus far followed. By tightly constraining design parameters to fit within only a small subset of possible uses for an XR system, next generation XR products already exist in mature forms. The niche-application market's largest obstacle is cost; while the average consumer may be willing to spend ~$1,000 on an everyday-use smartphone replacement, it can be difficult to justify a similar expense on a product with dramatically fewer capabilities.

5. CONCLUSION

The evolution of MEMS-based laser projection, waveguide combiners, CGH's, and pupil tracking technologies are the critical factors needed to progress MR systems from the lab to heads of consumers: MEMS-based laser projection excels with speed. It trumps pixelated display's inherent size limitations, allowing for wide FOV and full color displays. Waveguide combiners are the basis of all HMD's, as they provide a true heads up viewing experience. Leveraging the Fourier domain applications of CGH's can solve a variety of issues including aberrations, contrast ratio, and VAC, with the computational trade off being the main downside. Lastly, touch-free and voice-free system interactions are going to be made possible via robust LOS tracking capabilities. Here, we have shown the effectiveness of a single LED placed on or off the optical axis in tracking both eye and head movement for large angles of eye rotation. At the time of publication, though, size and cost factors limit the market's ability to be successful at scale. Currently, functional designs are available only for single-application markets such as biking or watching movies. These designs, while achieving technical relevance with acceptable form factor, are still too expensive to be widely adopted. While companies like Microsoft and Magic Leap are taking the "start with everything" approach to design, others, like Snapchat, are making considerable progress from the other, form-oriented direction. Still, the industry is expected to grow astronomically in the coming years as technology matures and miniaturizes. In short, the XR industry, while not yet a dominant subset of the tech industry at the consumer level, is already integrating the analog and digital worlds in ways never seen before.

REFERENCES

[1] T. Widjanarko, M. El Guendy, A. O. Spiess, D. M. Sullivan, T. J. Durrant, O. A. Tastemur, A. J. Newman, D. F. Milne, A. Kaczorowski, "Clearing key barriers to mass adoption of augmented reality with computer-generated holography," Proc. SPIE 11310, Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR), 113100B (2020)
[2]Grayson C. "From Apple to OSRAM; From Displays to Devices: a Pre-CES Year-End Review of Smart Glasses.", http://www.giganti.co/SmartGlassesEOYroundup (2020)
[3] Dmitry V. Shmunk, "Near-eye display optic deficiencies and ways to overcome them," Proc. SPIE 11765, Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) II, 117650N (2021)
[4] Jörg Reitterer, Zhe Chen, Anna Balbekova, Gerhard Schmid, Gregor Schestak, Faraj Nassar, Manuel Dorfmeister, Matthias Ley, "Ultra-compact micro-electro-mechanical laser beam scanner for augmented reality applications," Proc. SPIE 11765 (2021)
[5] Haas, G., "40-2: Invited Paper: Microdisplays for Augmented and Virtual Reality," SID Symp. Dig. Tech. Pap. 49(1), 506–509 (2018).
[6] Jang, H. J., Lee, J. Y., Kim, J., Kwak, J. and Park, J.-H., "Progress of display performances: AR, VR, QLED, and OLED," J. Inf. Disp. 21(1), 1–9 (2020).
[7] Oleg Petrak, Fabian Schwarz, Leon Pohl, Marcel Reher, Christian Janicke, Jan Przytarski, Frank Senger, Jörg Albers, Thorsten Giese, Lars Ratzmann, Peter Blicharski, Stephan Marauska, Thomas von Wantoch, Ulrich Hofmann, "Laser beam scanning based AR-display applying resonant 2D MEMS mirrors," Proc. SPIE 11765, Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) II, 1176503 (2021)
[8] Fidler, F., Balbekova, A., Noui, L., Anjou, S., Werner, T. and Reitterer, J., "Laser Beam Scanning in XR – benefits and challenges," Proc. SPIE (2021).
[9] Peillard, E., Itoh, Y., Moreau, G., Normand, J.-M., Lecuyer, A. and Argelaguet, F., "Can Retinal Projection Displays Improve Spatial Perception in Augmented Reality?," 2020 IEEE Int. Symp. Mix. Augment. Real. ISMAR, 80–89, IEEE, Porto de Galinhas, Brazil (2020).
[10] Hedili, M. K., Ulusoy, E., Kazempour, S., Soomro, S. and Urey, H., "Next Generation Augmented Reality Displays," 2018 IEEE Sens., 1–3 (2018).
[11] Lantian Mi, Chao Ping Chen, Wenbo Zhang, Jie Chen, Yuan Liu, and Changzhao Zhu., "A retinal-scanningbased near-eye display with diffractive optical element," presented at Proc.SPIE, (2020)
[12] Hofmann, U., Janes, J. and Quenzer, H.-J., "High-Q MEMS Resonators for Laser Beam Scanning Displays," 2, Micromachines 3(2), 509–528 (2012).
[13] Urey, H., Wine, D. W. and Osborn, T. D., "Optical performance requirements for MEMS-scanner-based microdisplays," MOEMS Miniaturized Syst. 4178, 176–185, International Society for Optics and Photonics (2000).
[14] J. W. Goodman, "Introduction to Fourier optics," in Introduction to Fourier Optics, J. W. Goodman, Ed., 3rd ed., Vol. 1, Roberts & Co. Publishers, Englewood, Colorado (2005).
[15] B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary masks," Appl. Opt. 5(6), 967–969 (1966).
[16] Youchao Wang, Daoming Dong, Peter J. Christopher, Andrew Kadis, Ralf Mouthaan, Fan Yang, Timothy D. Wilkinson, "Hardware implementations of computer-generated holography: a review," Opt. Eng. 59(10), 102413 (2020)
[17] M. Lucente, "The first 20 years of holographic video—and the next 20," in 2nd Annu. Int. Conf. Stereoscopic 3D Media and Entertainment, SMPTE, pp. 1–16 (2011).
[18] 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)
[19] Haas, G., "40-2: Invited Paper: Microdisplays for Augmented and Virtual Reality," SID Symp. Dig. Tech. Pap. 49(1), 506–509 (2018)
[20] 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).
[21] 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).
[22] 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).

[23] 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).
[24] Kuze, J. and Ukai, K., “Subjective evaluation of visual fatigue caused by motion images,” Displays 29(2), 159–166 (2008).
[25] Schwiegerling, Jim. “Field Guide to Visual and Opthalmic Optics” (2004).
[26] 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)
[27] 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).
[28] Bernard C. Kress, "Optical waveguide combiners for AR headsets: features and limitations," Proc. SPIE 11062, Digital Optical Technologies 2019, 110620J (2019)
[29] Khaled Sarayeddine, Pascale Benoit, Gilhem Dubroca and Xavier Hugel “Monolithic Low-Cost plastic light guide for full colour see through personal video glasses”, in ISSN-L 1883-2490/17/1433 ITE and SID (IDW 10), pp1433-1435. (2010)
[30] Herwig Kogelnik, “Coupled Wave theory for Thick Hologram Gratings”, The Bell System Technical Journal, Vol 48, No 9 (1969).
[31] Hong Hua, Prasanna Krishnaswamy, and Jannick P. Rolland, "Video-based eyetracking methods and algorithms in head-mounted displays," Opt. Express 14, 4328-4350 (2006)
[32] C. W. Hess, R. Muri, O. Meienberg, “Recording of horizontal saccadic eye movements: methodological comparison between electro-oculography and infrared reflection oculography,” Neuro-Ophthalmology 6, (1986).
[33] D. A. Robinson, “A method of measuring eye movements using a scleral search coil in a magnetic field,” IEEE Trans. Biomed. Electron. BME 10, 137-145 (1963).
[34] S. W. Shih, and Jin Liu, “A novel approach to 3-D gaze tracking using stereo cameras,” IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics 34, 234-245 (2004).
[35] K. Ryoung Park, “Gaze detection by wide and narrow view stereo camera,” Lecture Notes in Computer Science (CIARP’2004) 3287, 140-147 (2004).
[36] Edward Boettner and J Reimer Wolter. "Transmission of the Ocular Media," Investigative Opthalmology (1962)
[37] Mordor Intelligence, “Augmented Reality Market Overview.” https://www.mordorintelligence.com/ (2022)
[38] https://www.microsoft.com/en-us/hololens
[39] https://www.magicleap.com/
[40] https://www.spectacles.com/
[41] https://www.google.com/glass/start/
[42] https://everysight.com/product/raptor/
[43] https://moverio.epson.com/
[44] https://www.lenovo.com/us/en/p/tablets/vr-smartdevices/virtual-reality/thinkreality-a3/