

# Contents

<i>Preface</i>	<i>xi</i>
<i>Acknowledgments</i>	<i>xiii</i>
<i>Acronyms</i>	<i>xv</i>
<b>1 Introduction</b>	<b>1</b>
Word of Caution for the Rigorous Optical Engineer	4
<b>2 Maturity Levels of the AR/VR/MR/Smart-Glasses Markets</b>	<b>5</b>
<b>3 The Emergence of MR as the Next Computing Platform</b>	<b>13</b>
3.1 Today’s Mixed-Reality Check	15
<b>4 Keys to the Ultimate MR Experience</b>	<b>19</b>
4.1 Wearable, Vestibular, Visual, and Social Comfort	19
4.2 Display Immersion	22
4.3 Presence	23
<b>5 Human Factors</b>	<b>25</b>
5.1 The Human Visual System	25
5.1.1 Line of sight and optical axis	25
5.1.2 Lateral and longitudinal chromatic aberrations	26
5.1.3 Visual acuity	27
5.1.4 Stereo acuity and stereo disparity	29
5.1.5 Eye model	29
5.1.6 Specifics of the human-vision FOV	30
5.2 Adapting Display Hardware to the Human Visual System	32
5.3 Perceived Angular Resolution, FOV, and Color Uniformity	34
<b>6 Optical Specifications Driving AR/VR Architecture and Technology Choices</b>	<b>37</b>
6.1 Display System	37
6.2 Eyebox	37
6.3 Eye Relief and Vertex Distance	40

6.4	Reconciling the Eyebox and Eye Relief	41
6.5	Field of View	42
6.6	Pupil Swim	43
6.7	Display Immersion	44
6.8	Stereo Overlap	44
6.9	Brightness: Luminance and Illuminance	46
6.10	Eye Safety Regulations	49
6.11	Angular Resolution	50
6.12	Foveated Rendering and Optical Foveation	52
<b>7</b>	<b>Functional Optical Building Blocks of an MR Headset</b>	<b>57</b>
7.1	Display Engine	58
7.1.1	Panel display systems	59
7.1.2	Increasing the angular resolution in the time domain	63
7.1.3	Parasitic display effects: screen door, aliasing, motion blur, and Mura effects	65
7.1.4	Scanning display systems	67
7.1.5	Diffraction display systems	69
7.2	Display Illumination Architectures	71
7.3	Display Engine Optical Architectures	74
7.4	Combiner Optics and Exit Pupil Expansion	75
<b>8</b>	<b>Invariants in HMD Optical Systems, and Strategies to Overcome Them</b>	<b>77</b>
8.1	Mechanical IPD Adjustment	81
8.2	Pupil Expansion	82
8.3	Exit Pupil Replication	83
8.4	Gaze-Contingent Exit Pupil Steering	84
8.5	Exit Pupil Tiling	84
8.6	Gaze-Contingent Collimation Lens Movement	85
8.7	Exit Pupil Switching	86
<b>9</b>	<b>Roadmap for VR Headset Optics</b>	<b>89</b>
9.1	Hardware Architecture Migration	89
9.2	Display Technology Migration	91
9.3	Optical Technology Migration	92
<b>10</b>	<b>Digital See-Through VR Headsets</b>	<b>97</b>
<b>11</b>	<b>Free-Space Combiners</b>	<b>101</b>
11.1	Flat Half-Tone Combiners	101
11.2	Single Large Curved-Visor Combiners	102

---

11.3	Air Birdbath Combiners	104
11.4	Cemented Birdbath–Prism Combiners	105
11.5	See-Around Prim Combiners	106
11.6	Single Reflector Combiners for Smart Glasses	107
11.7	Off-Axis Multiple Reflectors Combiners	107
11.8	Hybrid Optical Element Combiners	108
11.9	Pupil Expansion Schemes in MEMS-Based Free-Space Combiners	109
11.10	Summary of Free-Space Combiner Architectures	112
11.11	Compact, Wide-FOV See-Through Shell Displays	112
<b>12</b>	<b>Freeform TIR Prism Combiners</b>	<b>115</b>
12.1	Single-TIR-Bounce Prism Combiners	115
12.2	Multiple-TIR-Bounce Prism Combiners	116
<b>13</b>	<b>Manufacturing Techniques for Free-Space Combiner Optics</b>	<b>119</b>
13.1	Ophthalmic Lens Manufacturing	119
13.2	Freeform Diamond Turning and Injection Molding	119
13.3	UV Casting Process	123
13.4	Additive Manufacturing of Optical Elements	124
13.5	Surface Figures for Lens Parts Used in AR Imaging	125
<b>14</b>	<b>Waveguide Combiners</b>	<b>127</b>
14.1	Curved Waveguide Combiners and Single Exit Pupil	128
14.2	Continuum from Flat to Curved Waveguides and Extractor Mirrors	129
14.3	One-Dimensional Eyebox Expansion	131
14.4	Two-Dimensional Eyebox Expansion	133
14.5	Display Engine Requirements for 1D or 2D EPE Waveguides	134
14.6	Choosing the Right Waveguide Coupler Technology	136
14.6.1	Refractive/reflective coupler elements	136
14.6.2	Diffraction/holographic coupler elements	137
14.6.3	Achromatic coupler technologies	146
14.6.4	Summary of waveguide coupler technologies	148
<b>15</b>	<b>Design and Modeling of Optical Waveguide Combiners</b>	<b>151</b>
15.1	Waveguide Coupler Design, Optimization, and Modeling	151
15.1.1	Coupler/light interaction model	151

15.1.2	Increasing FOV by using the illumination spectrum	154
15.1.3	Increasing FOV by optimizing grating coupler parameters	156
15.1.4	Using dynamic couplers to increase waveguide combiner functionality	157
15.2	High-Level Waveguide-Combiner Design	158
15.2.1	Choosing the waveguide coupler layout architecture	158
15.2.2	Building a uniform eyebox	159
15.2.3	Spectral spread compensation in diffractive waveguide combiners	160
15.2.4	Field spread in waveguide combiners	161
15.2.5	Focus spread in waveguide combiners	162
15.2.6	Polarization conversion in diffractive waveguide combiners	165
15.2.7	Propagating full-color images in the waveguide combiner over a maximum FOV	165
15.2.8	Waveguide-coupler lateral geometries	167
15.2.9	Reducing the number of plates for full-color display over the maximum allowed FOV	170
<b>16</b>	<b>Manufacturing Techniques for Waveguide Combiners</b>	<b>175</b>
16.1	Wafer-Scale Micro- and Nano-Optics Origination	175
16.1.1	Interference lithography	177
16.1.2	Multilevel, direct-write, and grayscale optical lithography	178
16.1.3	Proportional ion beam etching	180
16.2	Wafer-Scale Optics Mass Replication	180
<b>17</b>	<b>Smart Contact Lenses and Beyond</b>	<b>185</b>
17.1	From VR Headsets to Smart Eyewear and Intra-ocular Lenses	185
17.2	Contact Lens Sensor Architectures	185
17.3	Contact Lens Display Architectures	187
17.4	Smart Contact Lens Fabrication Techniques	189
17.5	Smart Contact Lens Challenges	189
<b>18</b>	<b>Vergence–Accommodation Conflict Mitigation</b>	<b>191</b>
18.1	VAC Mismatch in Fixed-Focus Immersive Displays	192
18.1.1	Focus rivalry and VAC	193

---

18.2	Management of VAC for Comfortable 3D Visual Experience	193
18.2.1	Stereo disparity and the horopter circle	195
18.3	Arm's-Length Display Interactions	196
18.4	Focus Tuning through Display or Lens Movement	197
18.5	Focus Tuning with Micro-Lens Arrays	198
18.6	Binary Focus Switch	200
18.7	Varifocal and Multifocal Display Architectures	201
18.8	Pin Light Arrays for NTE Display	205
18.9	Retinal Scan Displays for NTE Display	206
18.10	Light Field Displays	207
18.11	Digital Holographic Displays for NTE Display	210
<b>19</b>	<b>Occlusions</b>	<b>215</b>
19.1	Hologram Occlusion	215
19.2	Pixel Occlusion, or "Hard-Edge Occlusion"	215
19.3	Pixelated Dimming, or "Soft-Edge Occlusion"	216
<b>20</b>	<b>Peripheral Display Architectures</b>	<b>217</b>
<b>21</b>	<b>Vision Prescription Integration</b>	<b>221</b>
21.1	Refraction Correction for Audio-Only Smart Glasses	222
21.2	Refraction Correction in VR Headsets	223
21.3	Refraction Correction in Monocular Smart Eyewear	223
21.4	Refraction Correction in Binocular AR Headsets	225
21.5	Super Vision in See-Through Mode	226
<b>22</b>	<b>Sensor Fusion in MR Headsets</b>	<b>227</b>
22.1	Sensors for Spatial Mapping	229
22.1.1	Stereo cameras	229
22.1.2	Structured-light sensors	230
22.1.3	Time-of-flight sensors	230
22.2	Head Trackers and 6DOF	231
22.3	Motion-to-Photon Latency and Late-Stage Reprojection	232
22.4	SLAM and Spatial Anchors	233
22.5	Eye, Gaze, Pupil, and Vergence Trackers	234
22.6	Hand-Gesture Sensors	239
22.7	Other Critical Hardware Requirements	240
	<b>Conclusion</b>	<b>241</b>
	<i>References</i>	<i>243</i>
	<i>Index</i>	<i>255</i>

# Chapter 1

## Introduction

Defense was the first application sector for augmented reality (AR) and virtual reality (VR), as far back as the 1950s.<sup>1</sup> Based on such early developments, the first consumer AR/VR boom expanded in the early 1990s and contracted considerably throughout that decade, a poster child of a technology ahead of its time and also ahead of its markets.<sup>2</sup> However, due to the lack of available consumer display technologies and related sensors, novel optical display concepts were introduced throughout the 90s<sup>3,4</sup> that are still considered as state of the art, such as the “Private Eye” smart glass from Reflection Technology (1989) and the “Virtual Boy” from Nintendo (1995)—both based on scanning displays rather than flat-panel displays. Although such display technologies were well ahead of their time,<sup>5-7</sup> the lack of consumer-grade IMU sensors, low-power 3D-rendering GPUs, and wireless data transfer technologies contributed to the end of this first VR boom. The other reason was the lack of digital content, or rather the lack of a clear vision of adapted AR/VR content for enterprise or consumer spaces.<sup>8,9</sup>

The only AR/VR sector that saw sustained efforts and developments throughout the next decade was the defense industry (flight simulation and training, helmet-mounted displays (HMDs) for rotary-wing aircrafts, and head-up displays (HUDs) for fixed-wing aircrafts).<sup>10</sup> The only effective consumer efforts during the 2000s was in the field of automotive HUDs and personal binocular headset video players.

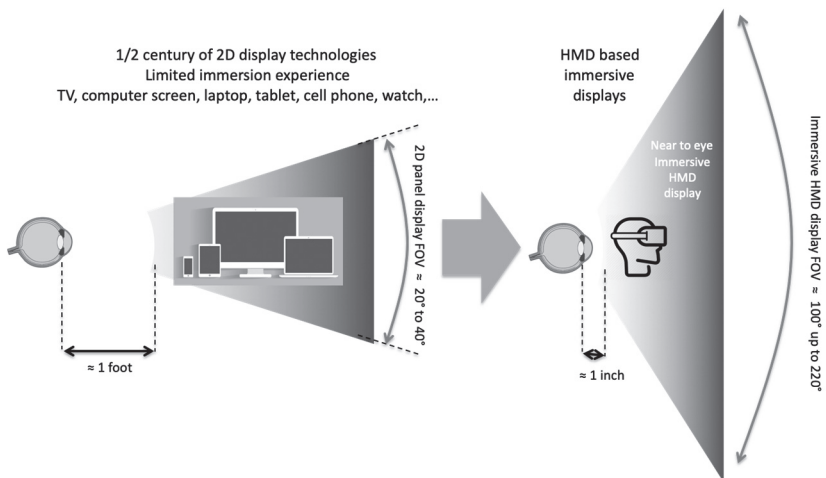
Today’s engineers, exposed at an early age to ever-present flat-panel display technologies, tend to act as creatures of habit much more than their peers 20 years ago, who had to invent novel immersive display technologies from scratch. We have therefore seen since 2012 the initial implementations of immersive AR/VR HMDs based on readily available smartphone display panels (LTPS-LCD, IPS-LCD, AMOLED) and pico-projector micro-display panels (HTPS-LCD, mu-OLED, DLP, LCoS), IMUs, and camera and depth map sensors (structured light or time of flight (TOF)). Currently, HMD architectures are evolving slowly to more specific technologies, which might be a

better fit for immersive requirements than flat panels were, sometimes resembling the display technologies invented throughout the first AR/VR boom two decades earlier (inorganic mu-iLED panels, 1D scanned arrays, 2D laser/VCSEL MEMS scanners, etc.).

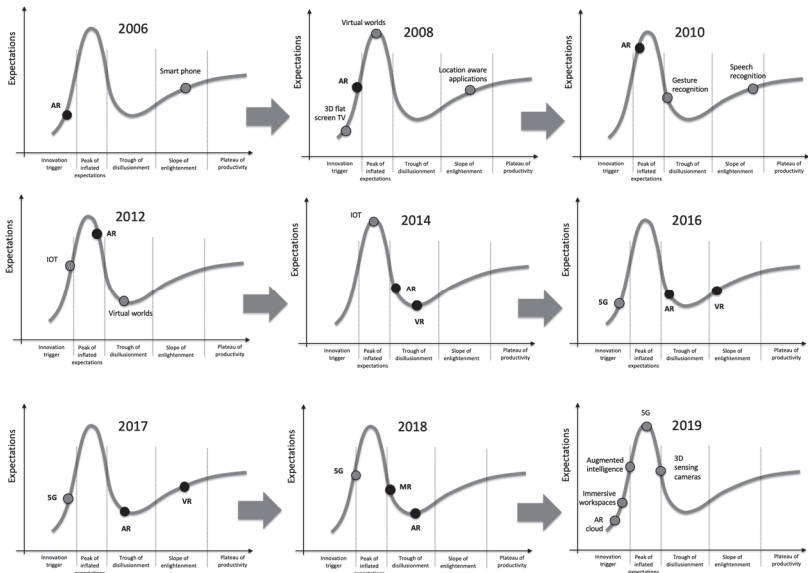
The smartphone technology ecosystem, including the associated display, connectivity, and sensor systems, shaped the emergence of the second AR/VR boom and formed the first building blocks used by early product integrators. Such traditional display technologies will serve as an initial catalyst for what is coming next.

The immersive display experience in AR/VR is, however, a paradigm shift from the traditional panel display experiences that have existed for more than half a century, going from CRT TVs, to LCD computer monitors and laptop screens, to OLED tablets and smartphones, to LCoS, DLP, and MEMS scanner digital projectors, to iLED smartwatches (see Fig. 1.1).

When flat-panel display technologies and architectures (smartphone or micro-display panels) are used to implement immersive near-to-eye (NTE) display devices, factors such as etendue, static focus, low contrast, and low brightness become severe limitations. Alternative display technologies are required to address the needs of NTE immersive displays to match the specifics of the human visual system.



**Figure 1.1** Immersive NTE displays: a paradigm shift in personal information display.



**Figure 2.1** Gartner Hype Cycles for Emerging Technologies (2006–2019) for AR/VR/MR.

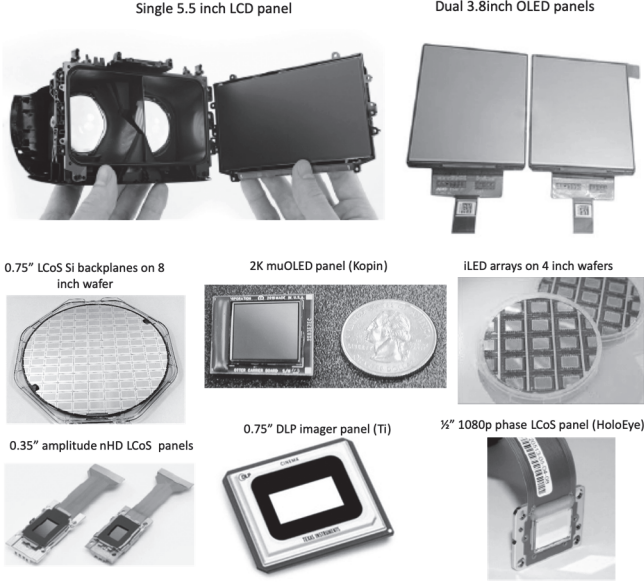
The 2008–2010 span introduced several technologies to the cycle that are now critical pillars to the AR experience, such as location-aware applications, gesture recognition, and speech recognition. Gesture recognition has had a tremendous boost with the Kinect technology development for the Xbox through 2009–2015 (structured illumination and then TOF), as well as speech recognition for personal assistants in smartphones.

IOT technologies appeared on the graph in 2012, culminated in hype in 2014, and were dropped promptly the next year, becoming a real product used by millions in consumer and enterprise fields. Many IOT core technologies share functionality with AR hardware.

AR peaked in its hype from 2010–2012, the years when Google Glass was introduced, along with many other smart glasses (Lumus, Optinvent, Reconjet, Epson Moverio, Sony, ODG, etc.).

VR appeared on the graph in 2014, the year Oculus was bought by Facebook for \$3B and coincided with the first large round of investment by Magic Leap Corp. (\$1/2B by Google and Qualcomm), which was followed by many similar rounds (a round E continues this trend today, 7 years after its creation and 2 years after its first product



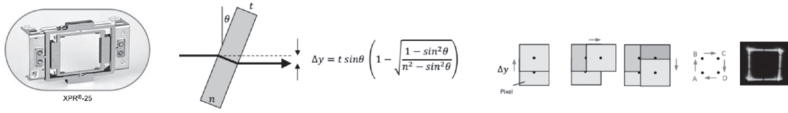


**Figure 7.4** Various display panels used in AR/VR products today.

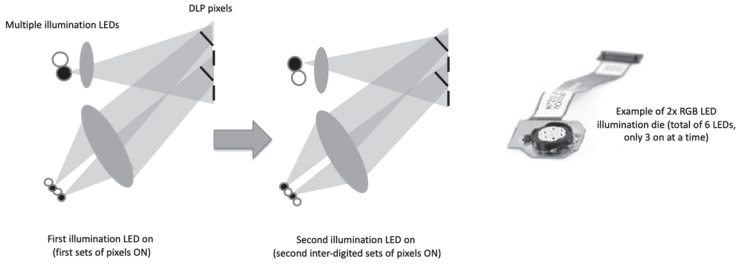
Polarization and emission cone are also important features of any micro-display-panel system (emissive or non-emissive), as they can considerably affect both the brightness of the immersive image at the eye as well as the perceived eyebox size. For example, LCoS-based and LC-based phase panels are polarized display panels (and thus require single polarized illumination), whereas LED (mini-LED or micro-iLED), mu-OLED or DLP panels and MEMS scanners are unpolarized displays and can therefore use all illumination polarization states. Using a single polarization state (linear or circular) does not necessarily mean reducing the illumination brightness by a factor of  $2\times$ , since polarization recovery schemes can be quite efficient and convert 20–30% of the wrong polarization, bringing it up to 70–80% (especially in free-space illumination architectures used in pico-projector illumination engines). Figure 7.4 shows some of these panels used in many AR/VR products today.

Finally, the efficiency of micro-display panels is paramount when it comes to wearable displays. Color-sequential LCoS displays are nearly 50% efficient, whereas color-filter LCoS displays are only about 15% efficient, and LTPS LCD micro-display panels (Kopin) are

Wobulation using a switchable window position (Optotune AG)



Wobulation using a switchable illumination source



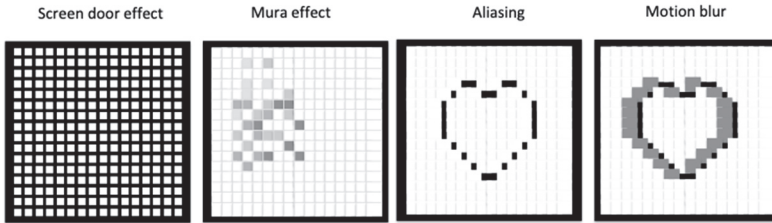
**Figure 7.5** Optical wobulation using mechanical movement or multiple illumination.

While mechanical moving and steerable wobulation can act on any display type (provided their refresh rate is high enough), illumination switch wobulation is limited to non-emissive displays such as DLP, HTPS, LCD, and LCoS displays.

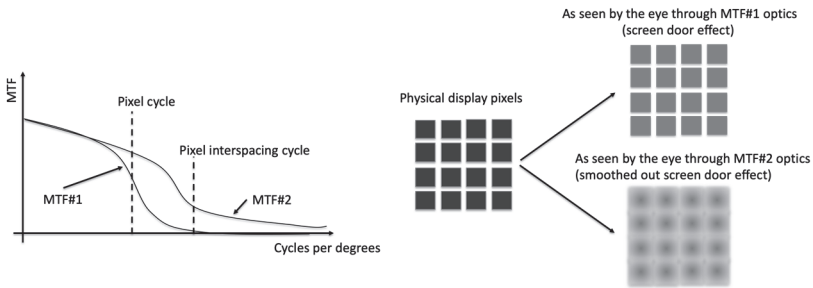
Optical wobulation can effectively increase the angular resolution (PPD) without increasing the number of pixels in the display panel. This technique is however limited to display architecture that have potential high refresh rates such as DLPs, and fast LCoS displays.

Another very compact wobulation technique would use multiple mirror pointing angles in a single DLP array, but that considerably increases the difficulty in designing and fabricating the MEMS DLP array. It would yield the most compact optical wobulation architecture.

Optical foveation and optical wobulation both can synthetically increase the number of pixels to yield a high-resolution perception for the viewer without increasing the physical number of pixels in the display. However, optical wobulation is not necessarily a form of optical foveation. It can morph into an optical foveation architecture if the wobulation is gaze contingent and can be steered by fractions of pixels over larger parts of the immersive FOV (see also the wobulation section on LBS display architectures).



**Figure 7.6** Screen-door effect, Mura effect, motion blur, and display aliasing.



**Figure 7.7** Tuning the MTF of the optics to smooth out the screen-door effect.

Other parasitic display effects from direct-view or micro-display panels are crepuscular rays. Crepuscular rays are streaks of light that can come from various sources, such as diffusion, diffraction, or even Fresnel lens rings. In a VR system, they can be prominent due to bright white text over a dark field.

The screen-door effect might be reduced by tuning either the MTF of the collimation lens in a VR system or the MTF of the display engine in an AR system (see Fig. 7.7). Although the physical display pixels still show a screen door, as will the virtual image through a high quality lens (MTF#1), an imaging system with a reduced MTF (MTF#2) can smooth out the virtual image in the angular space if the optics cannot resolve the pixel interspacing cycles further.

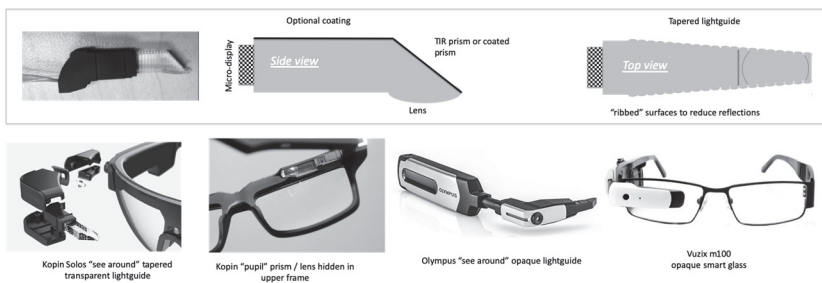
The human eye, with its impressive visual acuity allowing one to resolve features well below the arcmin scale, can image whatever the display engine can provide, at least with today's limited-pixel-density display technologies. Thus, acting on the display engine's MTF can

backlight in V1 and Kopin's LCD with a backlight in V2) is temple-mounted and collimated by a 100% reflective metal-coated lens located on the nasal side. The collimated field is then redirected to the user's eye by a 50/50 beam splitter. The use of a PBS to redirect the field into the user's eye would have been much more effective, but the lack of available low-birefringence plastic to make the rod led to the optimal choice of a 50/50 beam splitter, as losing brightness is a better option than producing ghost images from unwanted polarization states.

## 11.5 See-Around Prim Combiners

Another declination of the lateral linear lightguide combiner for small-FOV "see-around" smart glasses is shown in Fig. 11.7. These are not see-through combiners, but they are instead opaque. However, as the tip of the lateral lightguide combiner is tapered to a height that is smaller than the human pupil size (typically 3 mm or less), the combiner can be considered as "see-around" for the user, at least in the far-field domain.

As the lightguide combiner is not see-through here, the best adapted architecture might not be the double-pass birdbath Google Glass architecture shown in Fig. 11.5 but rather a more efficient single-pass version based on a prism ending and a collimation lens at the exit surface of the lightguide (see combiner tip with prism and lens in Fig. 11.6). This yields also a larger eyebox since the collimation lens is closer to the eye. Examples of such see-around smart glasses are the Kopin Solis (designed for cycling sports) or the Olympus smart glasses. The lateral surfaces can be structured (or ribbed) to reduce



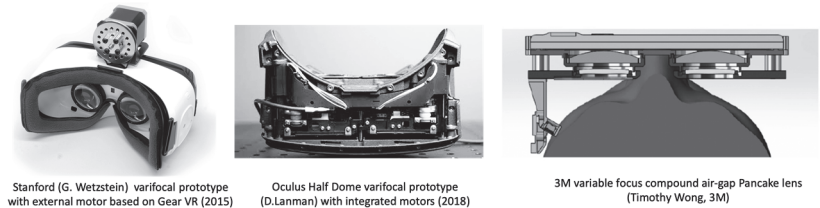
**Figure 11.7** Opaque tapered see-around lightguides (left and center) and wider opaque lightguide (right) for small-FOV monocular smart glasses.

## 18.4 Focus Tuning through Display or Lens Movement

The most straightforward way to change the location of the virtual image in front of the user, especially in a VR system, is to mechanically change the distance between the display panel and the collimation lens. This has been investigated by the computational display group at Stanford University<sup>100</sup> and in the Oculus Half Dome prototype unveiled in 2018 by Facebook Reality Labs (FRL); see Fig. 18.6.

FRL produced another VR varifocal system by using a non-moving optical architecture based on stacked liquid crystal lenses, as presented at the Oculus Connect 6 event in 2019 (see Fig. 18.7).

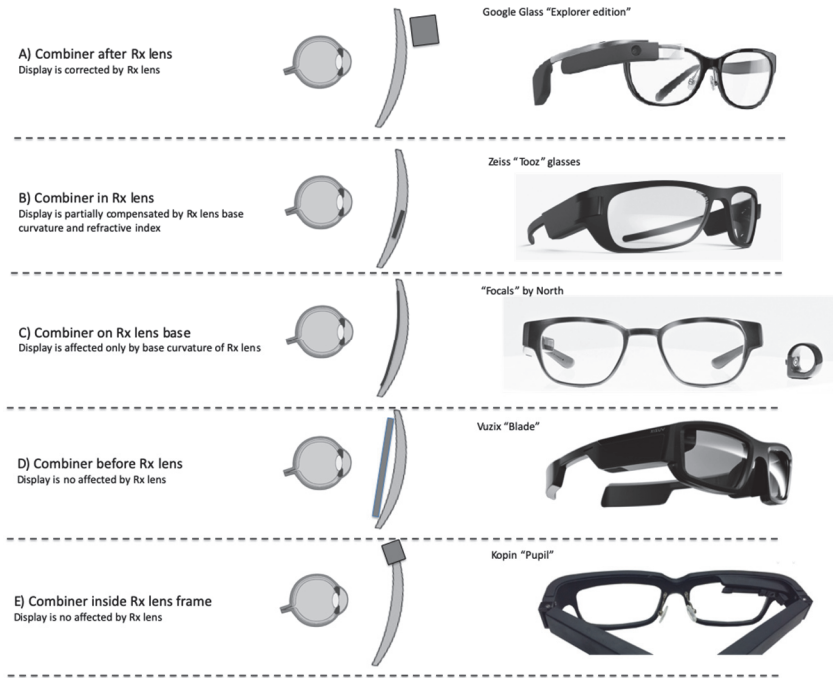
In the FRL architecture, six stacked LC lenses provide up to  $2^6$  (128) different foci to the VR content. This is a smart way to alleviate the downside of LC lenses, which cannot produce enough optical power due to the limited LC birefringence and LC thickness that can be reached. By stacking low-power LC lenses, relatively large LC lens apertures can be used without moving towards the use of Fresnel LC



**Figure 18.6** Focus tuning of an entire scheme by moving the display panel to the lens in VR systems.



**Figure 18.7** Facebook Reality Labs VR varifocal lens based on stacked LC lenses, and operation example.

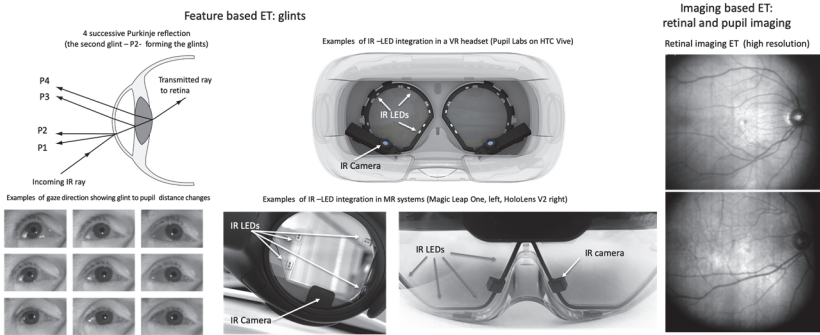


**Figure 21.2** Prescription correction lenses in various monocular smart glass architectures available today, and their effect on the digital display.

Figure 21.2 summarizes various prescription (Rx) compensations integration in some of the monocular smart glass architectures available today in industry. Five different cases are depicted in this figure:

- A: combiner after the Rx lens,
- B: combiner inside the Rx lens,
- C: combiner on the base surface of the Rx lens,
- D: combiner before the Rx lens, and
- E: combiner set inside the Rx lens frames.

There is no example of an architecture where the combiner would be located on the outer surface of the Rx lens since this would produce a complex “Mangin mirror” effect that would be difficult to manage optically. In conventional ophthalmic lenses, the base curvature is usually the generic curvature (cast within the ophthalmic puck) while



**Figure 22.7** Feature- and imaging-based ET techniques and implementation examples.

There are two main ET techniques used today (see Fig. 22.7):

- image-based ET (pupil position, size, and orientation, as well as more complex retinal imaging), and
- feature-based ET (using glints produced by IR LEDs).

Both techniques require IR cameras and some sort of IR lighting (single or multiple LEDs, flood or structured illumination), all in proximity to the eye. Glint-based ET that uses sets of IR LEDs around the combiner, pointing to the eye, are the most popular today (SMI/Apple, EyeFluence, Tobii, Pupil Labs, Magic Leap One, HoloLens V2, etc.). Both image- and feature-based architectures rely on IR illumination (850 nm up to 920 nm, depending on the IR sources used) to be most effective with B&W silicon photodetector arrays and also to be insensitive to display- or world-illumination changes (regardless of intensity and orientation).

Retinal imaging is a well-known pupil pursuit technique that has been used in ophthalmology for decades but very seldom for ET. When retinal scanning is combined with pupil center tracking, the technique can be made insensitive to slippage (movement of the headset due to sweat, shocks, etc.). Glint-based ET techniques are less forgiving for slippage. Retinal scanning can be easily combined with iris recognition. Retinal scanning is also a good technique to investigate diabetes-induced blood-vessel degeneration, glaucoma, and age-related macular degeneration (AMD).