Journal of Physics: Photonics

TOPICAL REVIEW

Virtual reality and augmented reality displays: advances and future perspectives

Kun Yin[1], Ziqian He[1], Jianghao Xiong[1], Junyu Zou[1], Kun Li[2] and Shin-Tson Wu[1,*]

[1] College of Optics and Photonics, University of Central Florida, Orlando, FL 32816, United States of America
[2] Goertek Electronics, 5451 Great America Parkway, Suite 301, Santa Clara, CA 95054, United States of America
* Author to whom any correspondence should be addressed.

E-mail: swu@creol.ucf.edu

Abstract

Virtual reality (VR) and augmented reality (AR) are revolutionizing the ways we perceive and interact with various types of digital information. These near-eye displays have attracted significant attention and efforts due to their ability to reconstruct the interactions between computer-generated images and the real world. With rapid advances in optical elements, display technologies, and digital processing, some VR and AR products are emerging. In this review paper, we start with a brief development history and then define the system requirements based on visual and wearable comfort. Afterward, various VR and AR display architectures are analyzed and evaluated case by case, including some of the latest research progress and future perspectives.

1. Introduction

As a promising next-generation display, virtual reality (VR) and augmented reality (AR) provide an attractive new way for people to perceive the world. Unlike conventional display technologies, such as TVs, computers, and smartphones that place a panel in front of the viewer, VR and AR displays are designed to revolutionize the interactions between the viewer, display, and surrounding environment. As a kind of information acquisition medium, VR and AR displays bridge the gap between computer-generated (CG) images and the real world. On the one hand, VR displays generate a fully immersive virtual environment based on CG images, which has a sufficient field of view (FOV) to provide refreshing virtual experience without relying on the viewer's real environment. On the other hand, AR display offers see-through capability with an enriched surrounding environment. By overlapping virtual images with the real world, viewers can immerse in an imaginative world that combines fiction and reality.

Although some commercial VR and AR displays have emerged in recent years, the origin of this technology can be traced back to the last century [1]. With the introduction of head-mounted display (HMD) and the virtual environment in the 1960s [2, 3], such a novel display concept was once considered as state-of-the-art. However, due to the lack of flat panel displays, image rendering capabilities, related sensors, wireless data transfer, and well-designed optical components, this display technology, which was ahead of its time then, came to an end. Fortunately, with the rapid development of optics [4–6], high resolution displays [7], and information technologies [8] in recent years, VR and AR are blooming again. Because of the impressive visual experience and high degrees of interaction between viewers and CG images, VR and AR are promising for widespread applications, including, but not limited to, healthcare, education, engineering design, manufacturing, and entertainment.

The goals of VR and AR displays are to provide reality-like clear images that can simulate, merge into, or rebuild the surrounding environment without wearer discomfort [9, 10]. Specifically, visual comfort has to meet the requirements of the human visual system based on the eye-to-brain imaging process, otherwise the viewer will feel unreal, unclear, or even dizzy and nauseous. Usually, the human eye has a large FOV: about 160° in the horizontal and 130° in the vertical directions for each eye (monocular vision). The overlapped binocular vision still has 120° FOV in the horizontal direction [11]. In parallel, the dioptre and rotation of
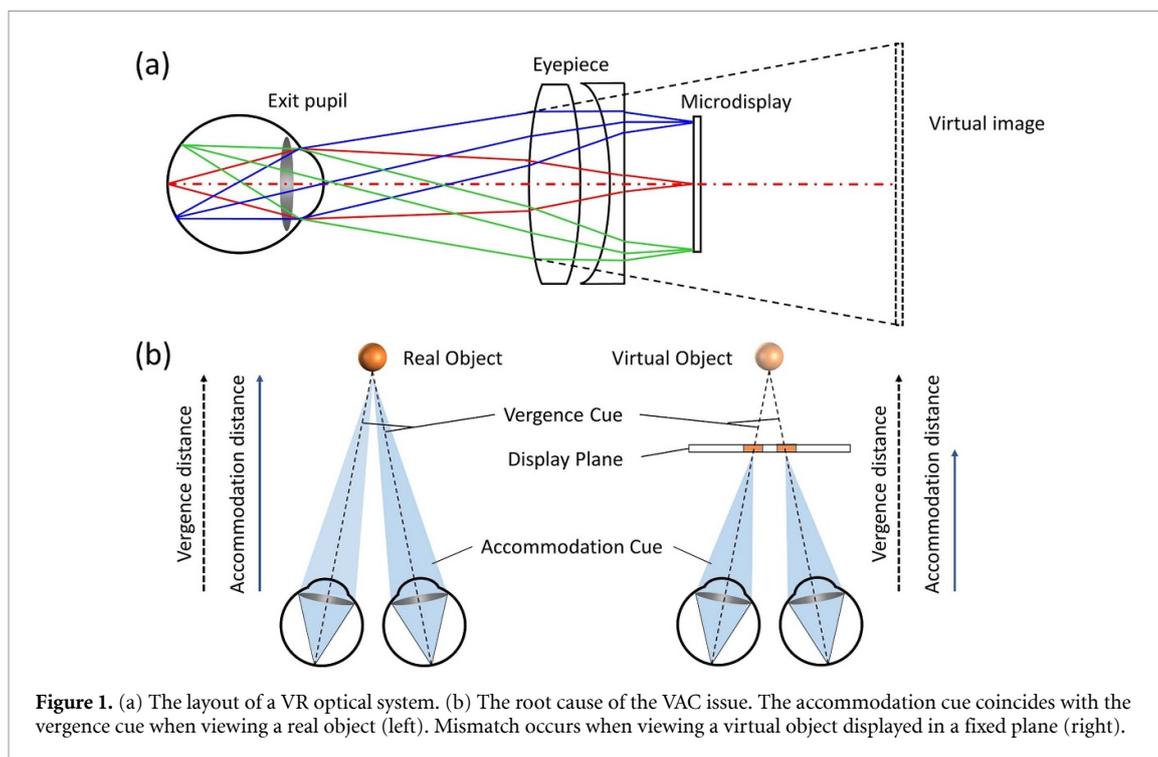
Figure 1. (a) The layout of a VR optical system. (b) The root cause of the VAC issue. The accommodation cue coincides with the vergence cue when viewing a real object (left). Mismatch occurs when viewing a virtual object displayed in a fixed plane (right).

the human eye lens can collaborate to focus on different positions of a real object with the correct depth of field and blur the other portions [12]. Therefore, to achieve visual comfort, the optical system should provide an adequate FOV, generate 3D images with matched depth and high resolution, and offer sufficient contrast and brightness, to name just a few examples. Regarding wearer comfort, a compact and lightweight structure is desired for long-time use. At present, due to the pros and cons between different optical components and system designs, it is still challenging for VR and AR to meet these goals. Therefore, in this paper, we focus on advanced VR and AR architectures aiming at visual and wearer comfort, and a more comprehensive understanding of the status quo.
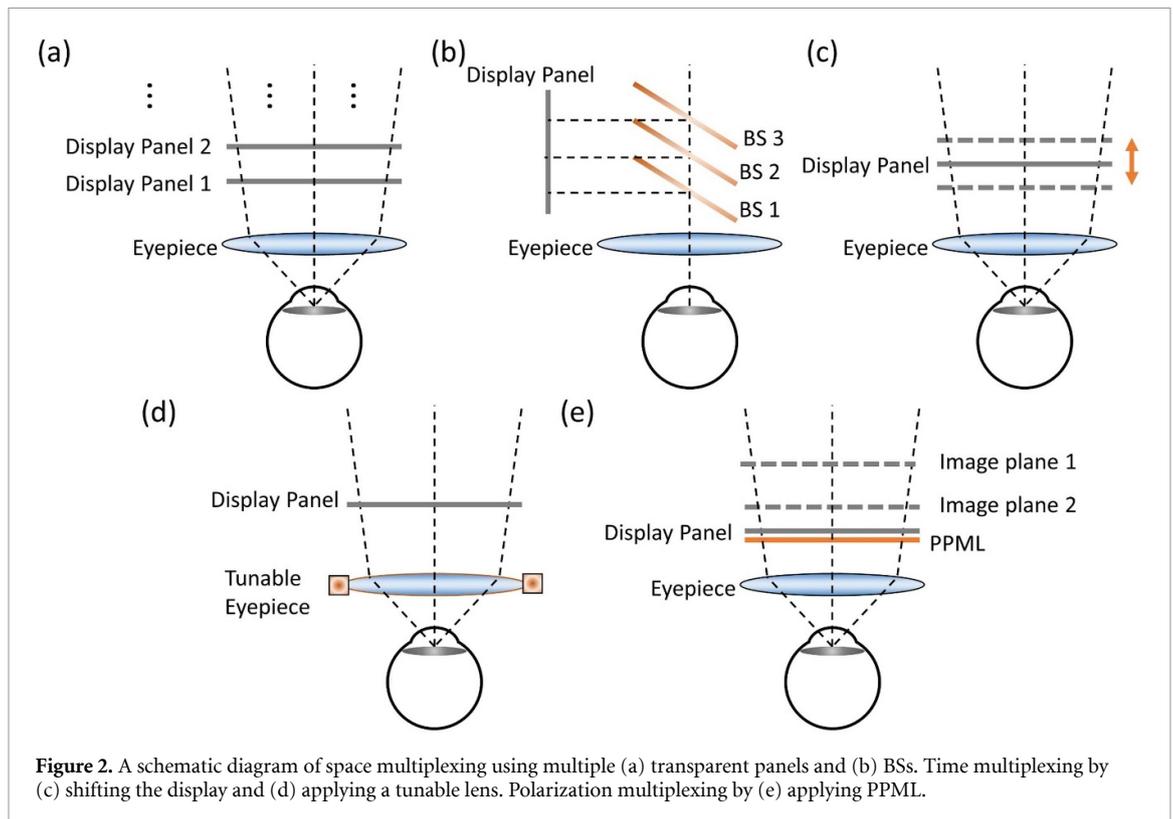
2. Advanced architectures for VR displays

Figure 1(a) depicts a schematic diagram of a VR optical system. For visual comfort, a broad FOV covering the human vision range can be achieved by designing a compact eyepiece with a low f-number ($f/\#$) [13]. However, due to the immersive experience with a completely virtual environment, the main issue is with the CG-3D image generation. When evaluating the capability of generating 3D images in VR, an important aspect of the human visual system is stereo perception. The real observation of a 3D object induces an accommodation cue (the focus of the eyes) and a vergence cue (relative rotation of the eyes), that match with each other (figure 1(b): left) [14, 15]. However, in most of the current VR systems, there is only one fixed display plane with different rendered contents. To capture the image information, the viewer's eyes will focus on the display plane, but the position of the CG-3D object is usually not in the display plane. As a result, the visual system in the viewer's brain will force the eyeball to focus on the virtual 3D object, while the eye lens focuses on the display plane, which leads to mismatched accommodation distance and vergence distance (figure 1(b): right). This phenomenon is called vergence–accommodation conflict (VAC) [16], which causes dizziness and nausea. Besides visual comfort, the overall weight and volume of the system will also limit the usage time and applications. To achieve wearer comfort, the system should be as light as possible while keeping a broad FOV in the virtual space. In this section, we will focus on advanced VR architectures that address 3D image generation to mitigate VAC and reduce the headset volume.

2.1. VAC mitigation

2.1.1. Multi-focal system
The multi-focal display was proposed to solve the VAC problem of HMDs in the late 1990s [17]. The basic principle of a multi-focal system is to generate multiple image planes or shift the position of image planes to match the vergence distance and accommodation distance, thereby overcoming the VAC issue. Based on

Figure 2. A schematic diagram of space multiplexing using multiple (a) transparent panels and (b) BSs. Time multiplexing by (c) shifting the display and (d) applying a tunable lens. Polarization multiplexing by (e) applying PPML.

different architectures and principles, multi-focal VR systems can be categorized into space multiplexing, time multiplexing, and polarization multiplexing systems.

Space multiplexing can simultaneously generate multiple image planes with different depths. To achieve this goal, Rolland *et al* [18] proposed a very straightforward method to physically place multiple screens based on transparent panels, as illustrated in figure 2(a). However, the transparent panels will not only increase the cost but also exhibit obvious moiré patterns after stacking multiple panels together [19]. To avoid this problem, beam splitters (BSs) can be utilized to help establish the space multiplexing system, as figure 2(b) shows [20]. In this design, the display panel is placed on one side, while the BSs reflect different parts of the display. Since the distance between each BS and the human eye is different, the image is displayed at different depths. Space multiplexing provides a direct solution to address VAC in VR displays and maintains image quality and frame rate. However, this architecture requires multiple display panels or BSs, which leads to dramatically increased weight and volume. Recently, a focal plane display with a phase-only spatial light modulator (SLM) has been demonstrated [21]. This architecture can achieve multi-focal planes with reduced system size and weight, but it requires an expensive SLM, and the image quality is not ready for commercial products yet.

The time multiplexing method relies on dynamic components and can timely change the panel distance (figure 2(c)) or the effective focal length (figure 2(d)) [22, 23]. The panel distance is usually changed by a mechanical motor, which leads to a lack of stability and modulation rate. For time multiplexing, the modulation rate of the dynamic components should be at least N times (N is the number of image planes) the display frame rate to avoid motion blurs. Therefore, compared with mechanically tuning the panel position, tuning the effective focal length through an electrically driven eyepiece is more favourable. Although it is still challenging to fabricate an adaptive lens with a wide tuning range and fast response time, this method can reduce the number of physical elements, so the system volume is much more compact than that of spatial multiplexing.

Polarization multiplexing generates multiple image planes based on different polarization states. To distinguish different polarization states, the most critical optical component is a polarization-dependent lens with different focal lengths for two orthogonal polarization states. Two such examples are: (a) the Pancharatnam–Berry phase lens, based on left-handed circularly polarized light (LCP)/right-handed circularly polarized light (RCP), and (b) the birefringent lens based on horizontal/vertical linearly polarized light [24, 25]. Figure 2(e) depicts the basic polarization multiplexing system. The light emitted from the display panel transmits through a pixelated polarization modulation layer (PPML), which can modulate the ratio of two orthogonal polarization states, so the light intensity of each pixel in the corresponding focal
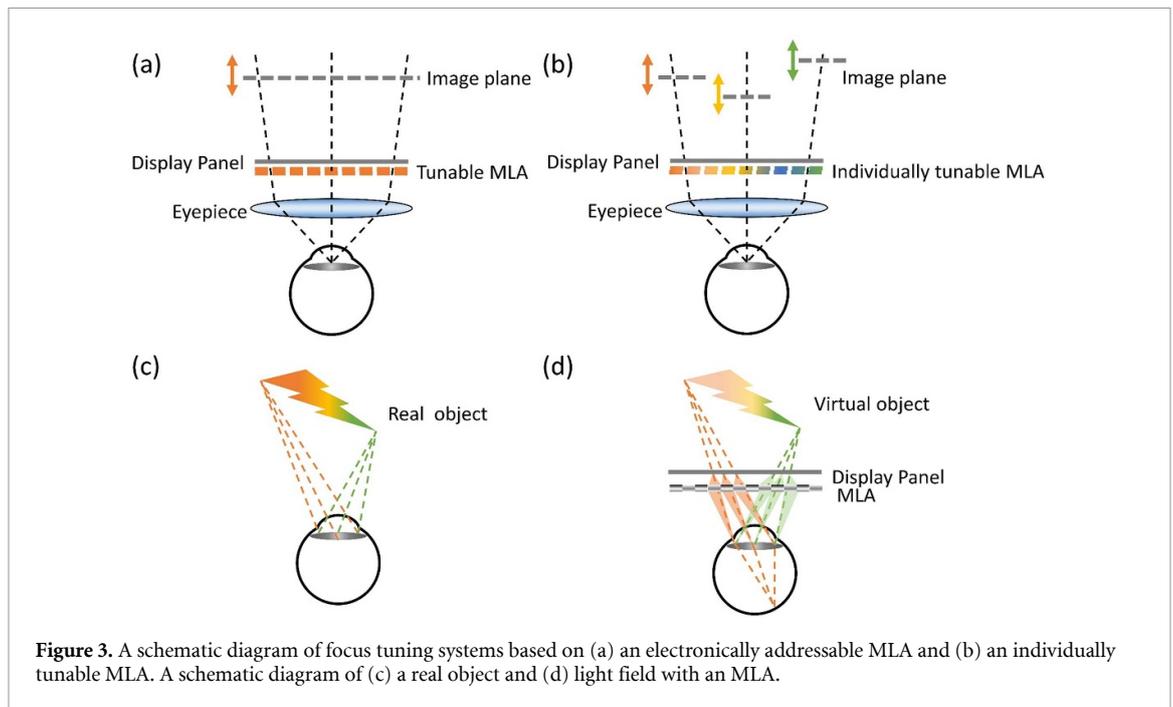
Figure 3. A schematic diagram of focus tuning systems based on (a) an electronically addressable MLA and (b) an individually tunable MLA. A schematic diagram of (c) a real object and (d) light field with an MLA.
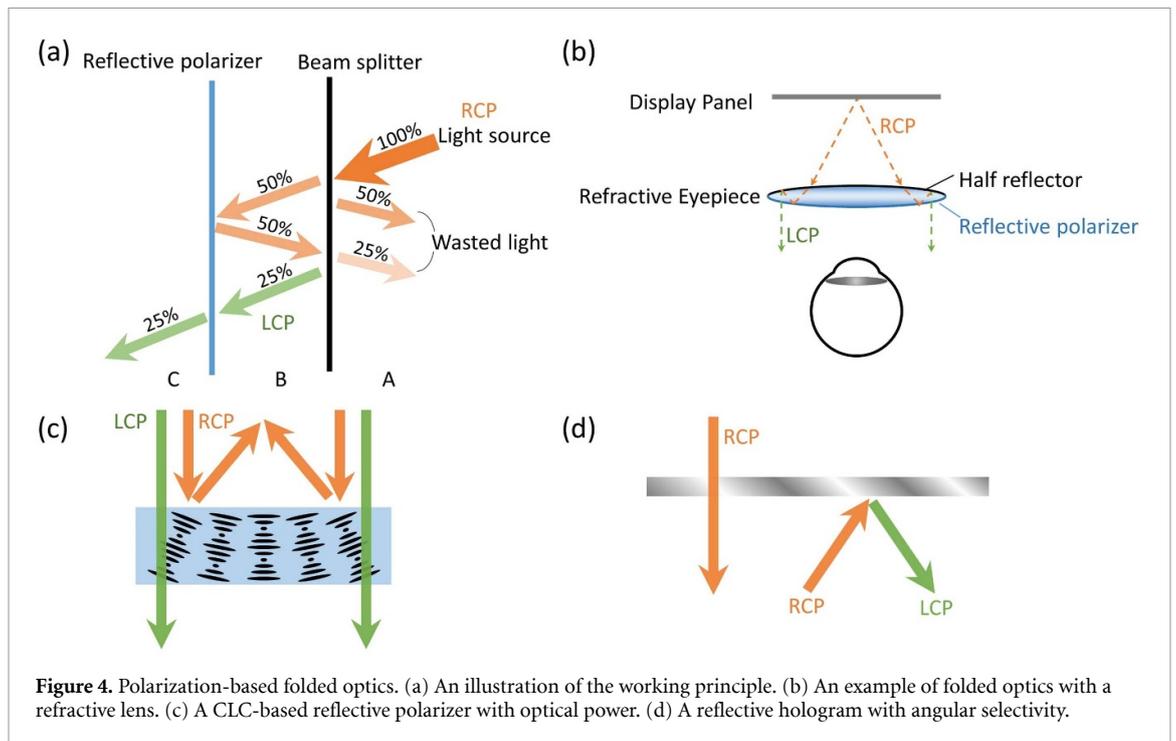
plane can be adjusted independently. PPML can be a polarization rotator for a linear polarization system [26] or an integrated polarization rotator and quarter-wave plate for a circularly polarized system [27]. The advantage of polarization multiplexing is that it can generate multiple image planes without sacrificing the frame rate or an enlarged system volume. However, the major limitation of polarization multiplexing is that only two orthogonal polarization states can be utilized. It should be mentioned that these multiplexing approaches can be combined. For example, time multiplexing or space multiplexing can be combined with polarization multiplexing to increase the number of focal planes [27, 28].

2.1.2. Micro-lens array system
Unlike using a large single lens as an eyepiece, another advanced architecture involves adding a micro-lens array (MLA) in front of the display panel to globally or individually change the position of virtual images in a VR system [10]. When the MLA is precisely aligned with the display panel, a small movement of the MLA can lead to a large focus change for the virtual image. As a result, instead of moving a thicker display panel or bulkier lens over a longer range, pushing or pulling the MLA plate a small distance can significantly mitigate the VAC. It is worth mentioning that the focus of an MLA based on liquid crystal materials can be switched dynamically for several microns, which means the movement of virtual images can be obtained without any mechanical motion, as shown in figure 3(a). Furthermore, as figure 3(b) shows, if each MLA element can be precisely controlled independently, then we can produce a specific focus for each lenslet and generate pixelated depth. These techniques are suitable for VR displays as well as for free-space-based couplers in AR displays. It is worth mentioning that in the MLA system, the resolution is usually an important issue, which needs to be further improved.

2.1.3. Light field system
To mitigate VAC, both temporally and spatially changed displays have been proposed. However, due to a limited or discrete tuning range, these methods can only partially recreate the 3D object with the correct depth. Rather than changing the image focus, light field displays ideally recreate a physical wavefront similar to that created by a real object. The light field capture (e.g. integral imaging) [29–31] can be achieved by a lens array to convert the light from display pixels to rays with arbitrary spatial angles. As depicted in figure 3(d), the spatial points correspond to the pixels on the display panel. To display a virtual 3D object, we trace the points on the object and light the corresponding pixels on the display panel. Then, the light field on those points can be approximated with discrete emitting rays. Although this method can provide correct depth information and retinal blur, the resolution is sacrificed. If the amount of information is taken into consideration, it is not surprising that these approaches that aim to show true 3D information cannot offer sufficient resolution due to the limited bandwidth of the current devices. Generally, the resolution is limited by the display and the individual lens. Although a high-resolution display has been proposed, the pixel pitch is still determined by the diffraction limit of the employed lens [29]. These approaches should gradually

Figure 4. Polarization-based folded optics. (a) An illustration of the working principle. (b) An example of folded optics with a refractive lens. (c) A CLC-based reflective polarizer with optical power. (d) A reflective hologram with angular selectivity.
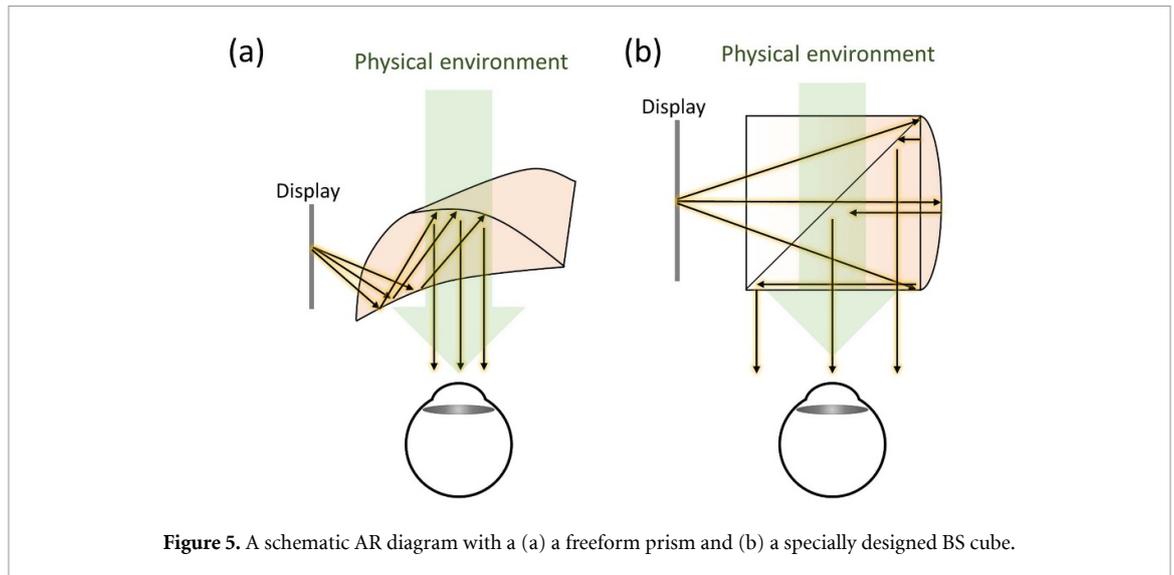
mature in the long run and eventually reach a satisfactory level for viewers. But at the current stage, the main drawbacks of this architecture are resolution loss, refresh rate increase, and/or redundancy of display panels.

2.2. Pancake VR

As discussed above, aside from visual comfort, wearable comfort is another important consideration. To reduce the volume and weight of a VR system, thereby improving its wearable comfort, a compact optical design, while taking the headset's central gravity into consideration, is urgently needed. Recently, polarization-based folded optics (or pancake optics) with further reduced form factors have attracted increasing attention. The system was originally proposed for use in flight simulators [32] and it has gained renewed interest due to rapid development of VR [33, 34]. The basic concept is to create a cavity to fold the optical path into a smaller space. The working mechanism is illustrated in figure 4(a). The cavity lies between a BS and a reflective polarizer. The BS (including a metallic or dielectric half mirror) has 50% transmittance and it flips the handedness of incident polarized light upon reflection. The reflective polarizer selectively transmits light with one polarization state and reflects the orthogonal one, which can be achieved by a wire-grid polarizer, birefringent multi-layer film, or cholesteric liquid crystal (CLC). The former two respond to a linear polarization, while the latter respond to a circular polarization. To explain the working principle, we use a circularly polarized light as an example. As shown in figure 4(a), the incoming RCP light in region A firstly passes through the BS (50%) and gets reflected by the reflective polarizer. Then, it is reflected by the BS again (25%), while flipped to the LCP state. Finally, the LCP light passes the reflective polarizer and enters region C. Because of the BS, only 25% of the total energy is delivered to the viewer's side. Therefore, system efficiency is an important issue in the pancake VR system. Practical systems often involve one or more refractive elements, which can be placed in any of the specified ABC regions. The surfaces of the reflective polarizer and BS can also be curved according to design requirements. An example with a refractive lens placed in region B is plotted in figure 4(b). The BS (half reflector) in this case is coated on the curved surface of the lens.

All the above discussions only consider traditional geometric optics, where the optical power is provided by reflection or refraction of the curved surfaces. Recent advances in holographic optics, however, offer an even wider range of choices for optical elements. Both the reflective polarizer and BS can be flat holographic films [35]. As figure 4(c) shows, the reflective polarizer can have a focusing power by patterning the CLC molecules. The polarization selectivity of CLC leads to an optical power for one circular polarization and total transparency for the other. The BS can also be replaced by a phase hologram. Such a phase hologram is often fabricated by holographic exposure of a photopolymer [36]. Its index modulation is usually small, resulting in narrow angular and spectral responses. This angular selectivity can be utilized to boost the overall system efficiency. As depicted in figure 4(d), for a certain reflective hologram, light within the angular response is reflected with flipped handedness. Other incident lights that do not meet the Bragg conditions

Figure 5. A schematic AR diagram with a (a) a freeform prism and (b) a specially designed BS cube.

will traverse the hologram. With this feature, the BS efficiency can potentially reach 100% because both the transmission and reflection efficiencies can reach 100%. This means the overall system efficiency can be improved from 25% to nearly 100%. However, the narrow angular and spectral selectivity also indicates the requirement for a directional backlight with narrow spectral linewidth, which could be challenging for practical implementation.

3. Advanced architectures for AR displays

In contrast to the immersive experience provided by VR displays, AR displays aim for see-through systems that overlap CG images with physical environments. To obtain this unique visual experience with wearable comfort, the near-eye systems need to possess high transmittance with sufficient FOV and a compact form factor. Therefore, freeform optics with broad FOV and high transmittance are essential for AR displays. However, due to the prism-shape, this architecture presents a relatively large volume and heavy weight. To reduce the system size while keeping a sufficient FOV, a lightguide-based structure and free-space coupler are commonly used to create a delicate balance between visual comfort and wearable comfort.

3.1. Freeform prisms and BS architectures
Freeform prisms have been extensively investigated due to the development of diamond-turning machines. Typically, the freeform prisms used in an AR system need a partially reflective surface and a total internal reflection (TIR) surface to overlap the CG images and transmit the surrounding environments. As shown in figure 5(a), this configuration sophisticatedly incorporates two refraction surfaces, a TIR surface, and a partial reflection surface into one element, and therefore allows extra design freedom [37, 38]. This design provides high-quality images with a wide FOV, but due to its volume limitation the entire system will be bulky and heavy. Another common example of a freeform-based AR device uses a designed BS cube as the coupler. In figure 5(b), the magnifying optics is a reflective concave mirror disposed directly on the BS cube, which has more freedom to be further optimized. This device architecture provides the simplest solution to AR display with a broad FOV but a larger form factor. Moreover, there is another trade-off between the FOV and eyebox (or exit pupil) due to the conservation of étendue, which is the product of the FOV and eyebox. Therefore, the larger the FOV, the smaller the eyebox [39].

3.2. Lightguide-based architectures
Compared to the freeform design, the lightguide-based structure has a more balanced performance between visual comfort and wearable comfort, especially in the compact and thin form factor [40, 41]. Over the past decade, lightguide-based near-eye display (LNED) has become one of the most widely used architectures for AR displays and is applied in many commercial products, such as HoloLens 2, Magic Leap 1, and Lumus. For an LNED, input and output couplers are pivotal optical elements affecting the system's performance. Typically, the input coupler has a high efficiency enabling it to fully utilize the light emitted from the optical engine. In contrast, the output coupler has low and gradient efficiency across the exit pupil to ensure an expanded and uniform eyebox. According to different coupler designs, LNEDs can be categorized into grating-based lightguides (figure 6(a)) and geometrical lightguides (figure 6(b)).
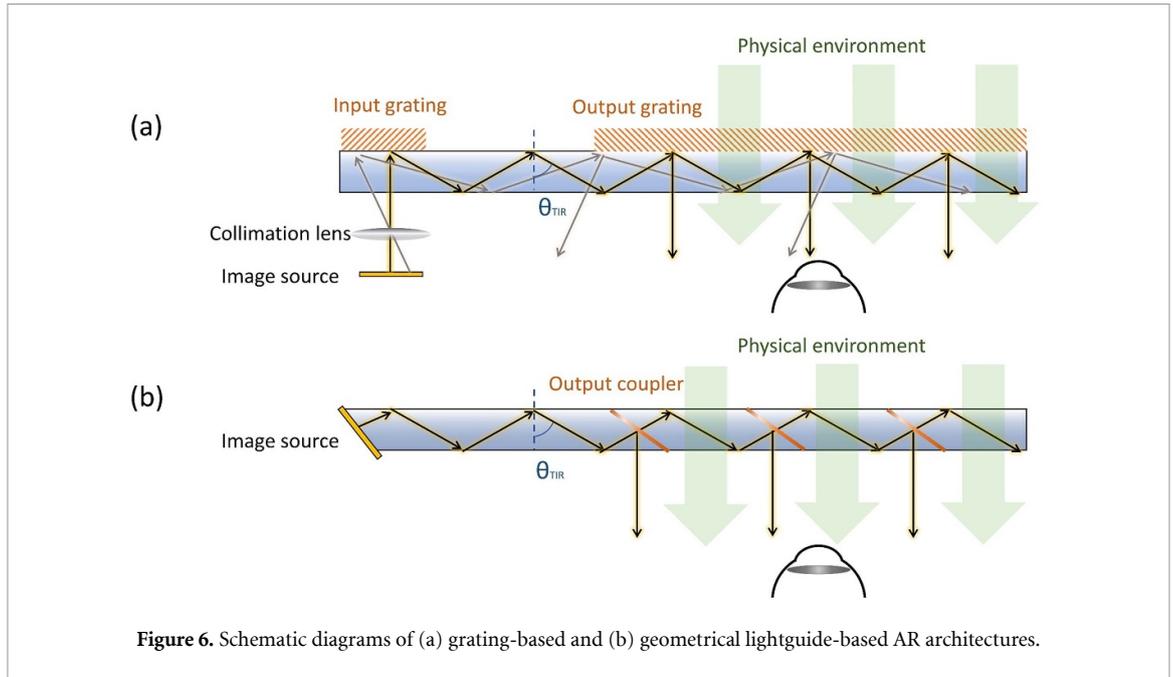
Figure 6. Schematic diagrams of (a) grating-based and (b) geometrical lightguide-based AR architectures.

3.2.1. Grating-based lightguide

As shown in figure 6(a), the display light is coupled into the lightguide by an input grating and then propagates inside the lightguide through TIR. When it encounters the output grating, the light is replicated and diffracted into the viewer's eyes. To provide a comprehensive understanding, we will theoretically analyze the FOV limit and discuss the commonly used grating couplers. For a diffractive grating, the first-order grating equation can be stated as:

$$n_{in} \sin \theta_{in} - n_{out} \sin \theta_{out} = \frac{\lambda}{\Lambda}, \tag{1}$$
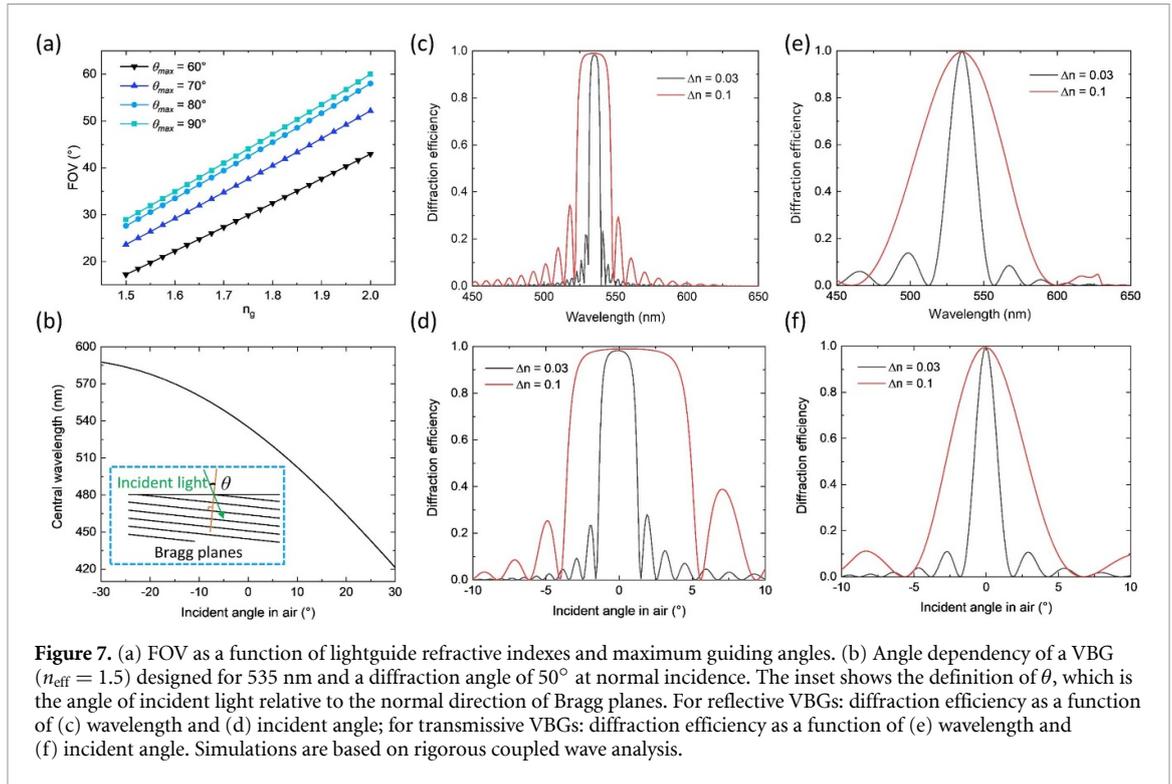
where θ_{in} and θ_{out} represent the incident angle and diffracted angle, respectively, n_{in} and n_{out} are the refractive index of the incident medium and output medium, λ is the wavelength in vacuum, and Λ is the grating period. With this simple grating equation, the maximum system FOV can be calculated. If we assume the FOV in air is centrosymmetric, then the viewing angle in air (θ_{air}) is related to the minimum/maximum guiding angles ($\theta_{min}/\theta_{max}$) in the lightguide as:

$$n_g \sin \theta_{min} - n_{air} \sin(-\theta_{air}) = \frac{\lambda}{\Lambda}, \\ n_g \sin \theta_{max} - n_{air} \sin \theta_{air} = \frac{\lambda}{\Lambda}, \tag{2}$$

where n_g is the refractive index of the lightguide, n_{air} is the refractive index of air, θ_{min} can be set to the TIR angle in the lightguide, and θ_{max} should be less than 90°. Thus, the maximum horizontal FOV is [42]:

$$\text{FOV} = 2\theta_{air} = 2 \arcsin \left[\frac{n_g}{2n_{air}} (\sin \theta_{max} - \sin \theta_{min}) \right]. \tag{3}$$

Figure 7(a) shows the FOV as a function of n_g and θ_{max}. In an ideal case where $\theta_{max} = 90°$ and $n_g = 2$, the maximum system FOV is only 60°. In practical designs, such a high index lightguide substrate is still challenging to achieve, and θ_{max} cannot approach 90° due to image quality considerations. This FOV limit is generally true for most grating-based lightguide AR. However, some methods can be employed to circumvent this limit. For instance, using a different system configuration [42], FOV can be expanded to 100°, or by leveraging polarization-dependent optical elements the FOV can be nearly doubled [43]. In equation (3), it seems that the FOV is independent of the wavelength, but the wavelength dependency is implicitly embedded in equation (2). For the extreme case with $\theta_{max} = 90°$ and $n_g = 2$, if the waveguide is designed at 535 nm, then the grating period is calculated to be 357 nm and the horizontal FOV is [−30°, 30°]. Utilizing such a grating period for blue (e.g. 450 nm) and red (e.g. 630 nm) with the assumption that the angle ranges in the lightguide are the same will lead to an FOV of [−15°, 48°] and [−50°, 14°], respectively. Thus, more than one grating is needed to obtain the same FOV for RGB colours. Although implementation of three gratings with narrow spectral bandwidths for R, G, and B in one lightguide is possible, it is still hard to eliminate colour crosstalk among different gratings. A more common

Figure 7. (a) FOV as a function of lightguide refractive indexes and maximum guiding angles. (b) Angle dependency of a VBG ($n_{eff} = 1.5$) designed for 535 nm and a diffraction angle of 50° at normal incidence. The inset shows the definition of θ, which is the angle of incident light relative to the normal direction of Bragg planes. For reflective VBGs: diffraction efficiency as a function of (c) wavelength and (d) incident angle; for transmissive VBGs: diffraction efficiency as a function of (e) wavelength and (f) incident angle. Simulations are based on rigorous coupled wave analysis.

choice is to have two (e.g. one for R, and one for G and B) or three (e.g. R, G, and B) lightguides [44], where the system's compactness is slightly sacrificed. Another important aspect is that the spectral response of most gratings depends on the incident angle. This can be well illustrated using a volume Bragg grating (VBG) as an example. For a VBG, the central wavelength is defined by the Bragg condition as:
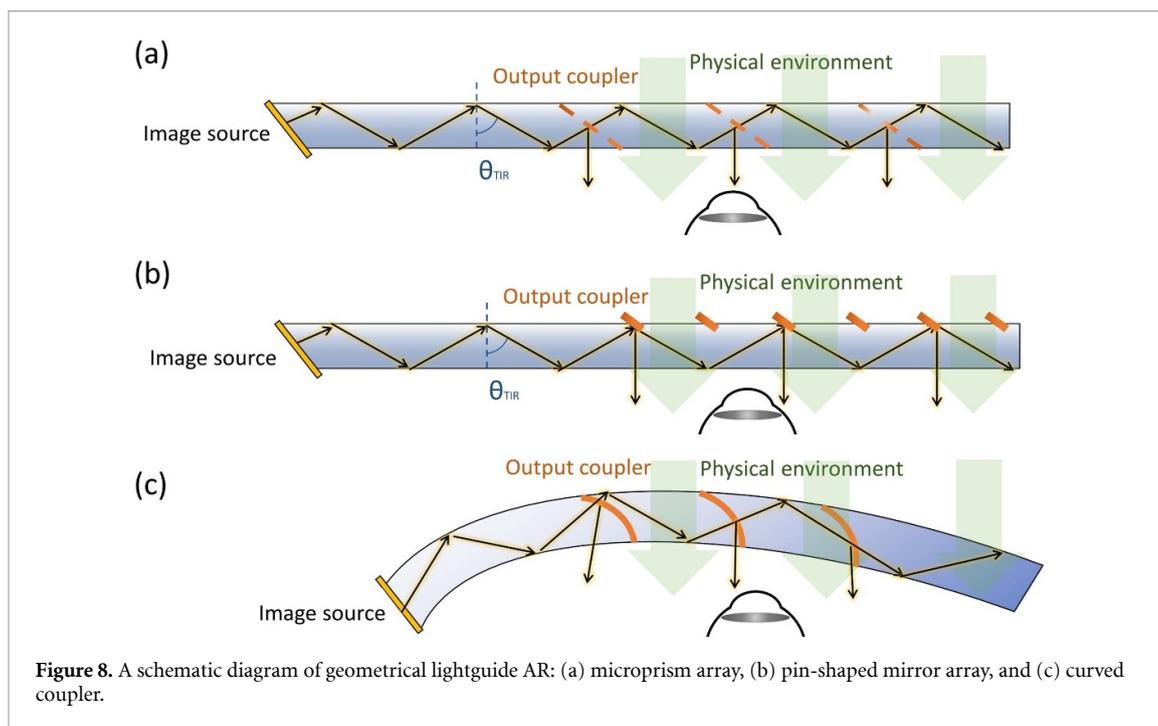
$$\lambda_B = 2n_{eff}\Lambda_B \cos\theta, \tag{4}$$

where θ represents the incident light angle with respect to the normal direction of Bragg planes (see the inset in figure 7(b)), and n_{eff} is the effective refractive index of the VBG. If a VBG (e.g. $n_{eff} = 1.5$) is designed for a normally incident green light ($\lambda = 535$ nm) with 50° diffraction angle in a lightguide, then the angle-dependent central wavelength can be calculated, as figure 7(b) depicts. For such a VBG, the central wavelength would shift from green to blue as the incident angle increases. Therefore, when designing a VBG-based lightguide AR for full-colour operation, such a colour crosstalk should be carefully analyzed.

In terms of selecting grating couplers, two types of gratings are commonly used in lightguide AR: holographic VBGs and surface relief gratings (SRGs). In holographic VBGs, sinusoidal refractive index modulation in the volume is introduced by interference exposure of holographic photopolymers. The refractive index modulation can be described by [45]:

$$n_{VBG} = n_{ave} + \Delta n \cos(\vec{K} \cdot \vec{r}), \tag{5}$$

where n_{ave} is the average refractive index, Δn is the refractive index modulation, $\vec{K}$ is the grating vector, and $\vec{r}$ denotes the vector of the spatial coordinate. Both transmissive and reflective VBGs can be obtained by designing different grating vectors. Generally, diffraction efficiency depends on the Δn as well as the VBG thickness. For both reflective and transmissive types to achieve a nearly 100% diffraction efficiency, the minimum hologram thickness increases monotonically as Δn decreases [46]. Besides diffraction efficiency, Δn also strongly influences the spectral and angular responses of VBGs. Figure 7 shows two examples of reflective (figures 7(c) and (d)) and transmissive (figures 7(e) and (f)) VBGs with $\Delta n = 0.03$ and $\Delta n = 0.1$. A higher Δn provides wider spectral and angular bandwidths for both reflective and transmissive VBGs, and the required thickness to achieve high diffraction efficiency is also decreased dramatically. The traditional stable holographic photopolymers usually have $\Delta n < 0.04$ [47], which will result in narrow spectral and angular bandwidths and a thicker film for large-angle diffraction gratings. Interestingly, DigiLens employed relatively large Δn holographic photopolymers based on holographic polymer-dispersed liquid crystals (HPDLC) to widen the bandwidths and reduce the thickness. The intrinsic electrical switching ability of HPDLC makes this design even more intriguing.

Figure 8. A schematic diagram of geometrical lightguide AR: (a) microprism array, (b) pin-shaped mirror array, and (c) curved coupler.
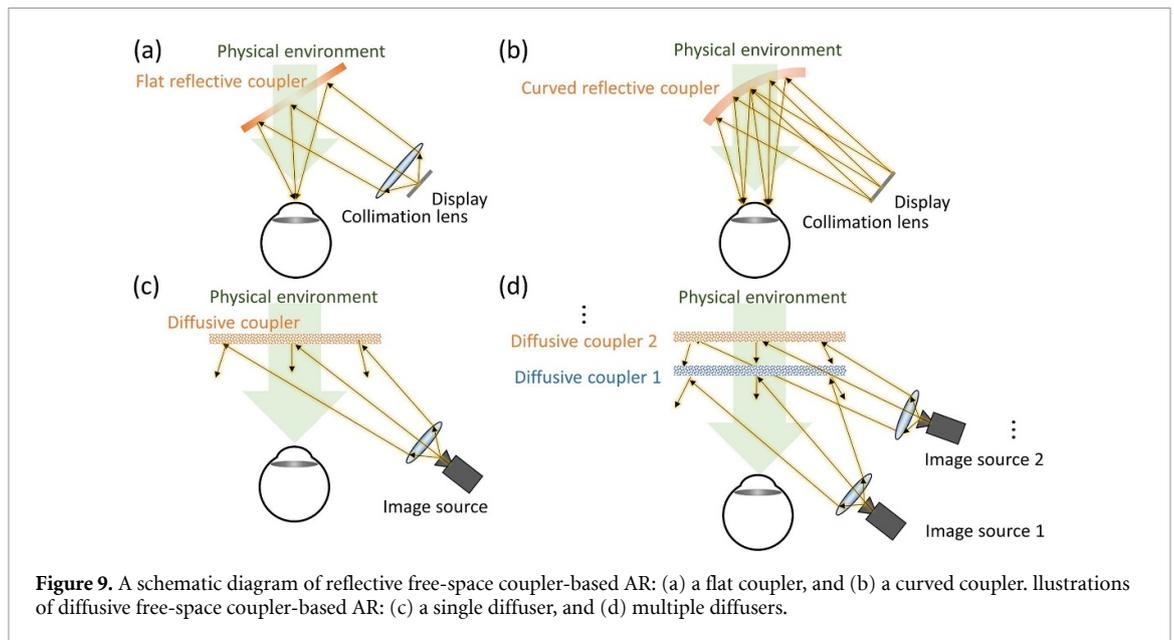
Unlike holographic VBGs that have refractive index modulation in the bulk, SRGs have specially designed microstructures on the surface, which can be massively produced by nanoimprinting [48]. The surface structures have a large design degree of freedom. The shapes of grating structures can be blazed, slanted, binary, and even analogue, according to different needs [10]. The spectral and angular responses of SRGs strongly depend on the shape of surface structures. Due to high refractive index contrast between the substrate and air, the structure height can be submicron to achieve high diffraction efficiency.

Besides holographic VBG and SRG, CLC-based polarization volume grating (PVG) is also a strong contender [49, 50]. Due to their volume grating nature, PVGs can be treated as a branch of holographic VBGs and their spectral and angular responses are very similar. However, PVGs exhibit some unique properties. First, PVGs are strongly circular-polarization dependent originating from CLCs [51], while VBGs and SRGs have weak polarization dependency on linear polarizations. For example, for a left-handed reflective PVG, it only diffracts the LCP light within the bandwidth into the first order, while transmitting the RCP light. This feature is useful for designing polarization-dependent optical elements. Second, if we use equation (5) to approximately describe the behaviour of PVGs (in fact, to describe PVGs the refractive indices in equation (5) should be replaced by dielectric constants), Δn can be very large. For instance, if the host liquid crystal has a birefringence of 0.2, the effective Δn can be as large as $0.5 \sim 0.6$ for a VBG. As a result, its spectral and angular bandwidths can be much larger than those of holographic VBGs. Moreover, recent studies show that multi-layer PVGs or gradient-pitch PVGs can be easily achieved to further enlarge the angular bandwidth [52, 53].

3.2.2. Geometrical lightguide
Compared to grating-based lightguides, geometrical lightguides need more complex designs (e.g. spatial variant coatings) to achieve gradient efficiency, and it is relatively hard to add a lens power to the output. However, the working principle is very simple, and all the designs are based on surface reflection. Generally, geometrical lightguides use embedded reflective surfaces as the exit pupil expander to reflect and replicate the light [54, 55].

As figure 6(b) shows, a series of cascaded, embedded, and partially reflective surfaces can be used as output couplers in the geometrical lightguide architecture. As the embedded surface is reflective, it yields good colour uniformity over the entire FOV. However, this cascaded design produces the Louver effect [10], which is unfavourable for see-through devices. Recently, this effect has been reduced due to better cutting, polishing, coating, and design, but it is still a limitation. In addition, these complicated fabricating processes put more burdens on manufacturers. As an extension, the embedded partially reflective surfaces can be designed as flat surfaces (figure 6(b)), pin-shaped mirror arrays (figure 8(a)), microprism arrays (figure 8(b)), or a curved lightguide with curved surfaces (figure 8(c)) [56].

Figure 9. A schematic diagram of reflective free-space coupler-based AR: (a) a flat coupler, and (b) a curved coupler. llustrations of diffusive free-space coupler-based AR: (c) a single diffuser, and (d) multiple diffusers.

3.3. Free-space coupler-based architectures

Unlike freeform optical devices or LNEDs, free-space couplers have greater freedom in the architecture, and there are no special restrictions on volume or TIR. Undoubtedly, due to large degrees of freedom, numerous architectures based on free-space couplers have been proposed, but each design has its pros and cons. These systems can be classified into three categories based on the working principles: reflective coupler, diffusive coupler, and diffractive coupler.

3.3.1. Reflective coupler

A reflective free-space coupler is based on the surface reflection of a flat or curved surface. Due to the high transmittance requirement, these surfaces should be partially reflective with sufficient reflection and transmittance. Figure 9(a) depicts the most straightforward architecture with a flat coupler, which is a tilted partial reflective surface. The CG images emitted from the display are collimated by the lens and then reflected into the viewer's eye through the flat coupler. To further simplify the system, such a flat coupler can be replaced by a partially reflective curved or freeform surface with a specially designed profile, as shown in figure 9(b). This design is aimed at smartphone displays rather than complex off-axis imaging and micro-display. This architecture has been successfully applied to Meta 2 by Meta Vison, DreamGlass by Dream World, and NorthStar by LeapMotion. Due to a large display panel and curved reflective surface, such a reflective coupler exhibits a relatively broad FOV but also a large system volume.

3.3.2. Diffusive coupler

A diffusive free-space coupler is based on the light scattering of optical elements [57]. In such a system, the displayed images are directly projected onto the coupler, which is usually a diffuser with a flat or curved surface. As illustrated in figure 9(c), the light is scattered by the coupler and then the image is displayed on the diffuser surface. Usually, the image source is a liquid-crystal-on-silicon (LCoS) or digital micro-mirror device, and the image resolution is controlled by the display and projection lens. To keep see-through capability, the diffuser should have angular selectivity to scatter the off-axis incident image and transmit the environment light in front of the eye. Therefore, the system can accommodate more than one diffuser, and thereby has the space to construct a 3D image with multiple planes [58], similar to the multiplane design in a VR system. As depicted in figure 9(d), each diffuser scatters the incoming light with the corresponding incident angle and do not interfere with each other.

3.3.3. Diffractive coupler

A diffractive free-space coupler is based on flat diffraction optical elements with designed phase profiles, such as lens or freeform optics [59, 60]. More specifically, the architectures based on diffractive couplers can be divided into free-space systems, Maxwellian systems, and integral imaging systems. The free-space-based diffractive couplers, as illustrated in figure 10(a), are utilized in a pupil-forming system, which means it uses relay optics to first image the object and then deliver the relayed image to the viewer's eye with the diffractive coupler [61, 62]. The image source includes but is not limited to a conventional 2D display and laser light
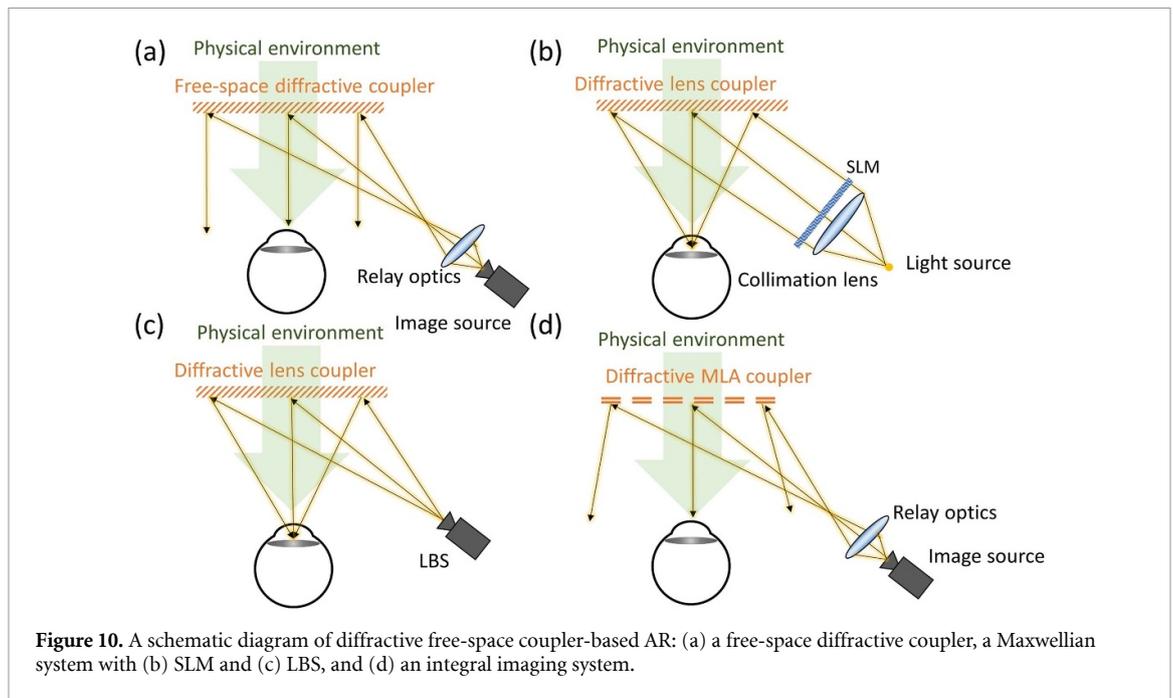
Figure 10. A schematic diagram of diffractive free-space coupler-based AR: (a) a free-space diffractive coupler, a Maxwellian system with (b) SLM and (c) LBS, and (d) an integral imaging system.

Table 1. Performance comparison of various AR architectures.

Type	Architecture		Form factor[a]	FOV diagonal[b]
Freeform	Prism		Large	~90°
	Beam splitter		Large	~90°
Lightguide	Grating coupler	VBG	Medium	~50°
		SRG	Medium	~50°
		HPDLC	Medium	~50°
	Geometrical coupler	Partially reflective surface	Medium	~40°
		Pin-shaped mirror array	Medium	~60°
		Microprism array	Medium	~30°
Free space	Reflective surface	Flat surface	Large	~90°
		Curved surface	Large	~90°
	Diffusive surface		Medium	~20°
	Diffractive surface	Off-axis lens	Small	~50°
		Maxwellian	Small	~90°
		Integral imaging	Medium	~20°

[a] These not only depend on the FOV and eyebox design but also include an optical engine part.

[b] These typical values come from products and prototypes.

source. However, due to the nature of diffractive flat optics and off-axis system configuration, aberrations like coma and astigmatism are large and need to be tackled with sophisticated optical design or image pre-processing. The Maxwellian system adopts the principle of a Maxwellian view [63], which directly forms a focus-free image on the retina. The diffractive couple can be a reflective off-axis lens with a designed focal length [64, 65]. It is worth mentioning that because the light needs to be focused on the pupil, the eyebox in the Maxwellian system is relatively small. To expand the eyebox, an exit pupil shifting can be applied to increase the area covered by the focal point [66]. Generally, the image light is focused by the coupler and the focal spot is located at the eye lens. As a result, the image on the retina stays in-focus no matter how much the optical power of the eye lens changes. Depending on the image source, the system can be achieved by an LCoS (figure 10(a)) or a laser beam scanner (LBS) (figure 10(c)) for a simpler design. The light field system with an MLA can also be applied to the AR system, such as the light field in a VR display [67, 68]. As depicted in figure 10(d), a typical configuration is the projection system which is used to relay the original image from the image source to near the focus of the diffractive coupler, similar to the free-space combination system. The relayed image then works in the same way as depicted in figure 3(d) and produces the light field to display 3D virtual objects. Similar to the multiplexing method in VR displays, these different AR architectures with fictional optics are not independent. On the contrary, they can be combined with each other to balance their respective advantages and trade-offs, and even enable new features [69].

To quantitatively summarize the performance of AR architectures based on visual comfort and wearable comfort, table 1 compares the form factor and FOV among different coupling methods. It should be mentioned that for each architecture, the performance can be further improved based on the current value but at the cost of other parameters. Therefore, the contents listed in table 1 are the general conditions rather than strict restrictions.

4. Conclusions

In this review, we summarize the advanced architectures with different optical components in the rapidly evolving VR and AR systems, including the most recent optical research and products, and analyze the systems based on the visual and wearable comforts case by case. Because of the various advanced architectures with unique features, such as reducing VAC through adjustable lenses, solving compact size issues using polarizing films, and providing a large FOV through freeform optics, VR and AR displays present both scientific significance and broad application prospects. Although, at the current stage, it is still challenging for these architectures to meet all the requirements for visual and wearable comfort, learning about and reviewing advanced systems will certainly help us focus on unresolved issues and inspire more elegant solutions.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

ORCID iD

Shin-Tson Wu ⓘ https://orcid.org/0000-0002-0943-0440

References

[1] Sutherland I E 1968 A head-mounted three dimensional display *Proc. FJCC* vol 1 pp 757–64
[2] Fisher S S, McGreevy M, Humphries J and Robinett W 1987 Virtual environment display system *Proc. 1986 Workshop on Interactive 3D Graphics* pp 77–87
[3] Amitai Y, Friesem A A and Weiss V 1989 Holographic elements with high efficiency and low aberrations for helmet displays *Appl. Opt.* **28** 3405–16
[4] Lee B 2003 Review of the present status of optical fiber sensors *Opt. Fiber Technol.* **9** 57–79
[5] Lin D, Fan P, Hasman E and Brongersma M L 2014 Dielectric gradient metasurface optical elements *Science* **345** 298–302
[6] Chen W T, Zhu A Y, Sanjeev V, Khorasaninejad M, Shi Z, Lee E and Capasso F 2018 A broadband achromatic metalens for focusing and imaging in the visible *Nat. Nanotechnol.* **13** 220–6
[7] Chen H W, Lee J H, Lin B Y, Chen S and Wu S T 2018 Liquid crystal display and organic light-emitting diode display: present status and future perspectives *Light Sci. Appl.* **7** 17168
[8] Schank R C 2014 *Conceptual Information Processing* (New York: Elsevier) (https://doi.org/10.1016/C2013-0-12220-9)
[9] Zhan T, Yin K, Xiong J, He Z and Wu S T 2020 Augmented reality and virtual reality displays: perspectives and challenges *iScience* **23** 101397
[10] Kress B C 2020 Optical architectures for augmented-, virtual-, and mixed-reality headsets *Proc. SPIE* **PM316** 270
[11] Wheelwright B, Sulai Y, Geng Y, Luanava S, Choi S, Gao W and Gollier J 2018 Field of view: not just a number *Proc. SPIE* **10676** 1067604
[12] Curcio C A, Sloan K R, Kalina R E and Hendrickson A E 1990 Human photoreceptor topography *J. Comp. Neurol.* **292** 497–523
[13] Geng Y *et al* 2018 Viewing optics for immersive near-eye displays: pupil swim/size and weight/stray light *Proc. SPIE* **10676** 1067606
[14] Tovée M J 1996 *An Introduction to the Visual System* (Cambridge: Cambridge University Press) (https://doi.org/10.1017/cbo9780511801556.004)
[15] Reichelt S, Häussler R, Fütterer G and Leister N 2010 Depth cues in human visual perception and their realization in 3D displays *Proc. SPIE* **7690** 76900B
[16] Hoffman D M, Girshick A R, Akeley K and Banks M S 2008 Vergence–accommodation conflicts hinder visual performance and cause visual fatigue *J. Vis.* **8** 33
[17] Wann J P, Rushton S and Mon-Williams M 1995 Natural problems for stereoscopic depth perception in virtual environments *Vis. Res.* **35** 2731–6
[18] Rolland J P, Krueger M W and Goon A 2000 Multifocal planes head-mounted displays *Appl. Opt.* **39** 3209–15
[19] Saveljev V V, Son J Y, Javidi B, Kim S K and Kim D S 2005 Moiré minimization condition in three-dimensional image displays *J. Disp. Technol.* **1** 347–53
[20] Akeley K, Watt S J, Girshick A R and Banks M S 2004 A stereo display prototype with multiple focal distances *ACM Trans. Graph.* **23** 804–13
[21] Matsuda N, Fix A and Lanman D 2017 Focal surface displays *ACM Trans. Graph.* **36** 86

[22] Shibata T, Kawai T, Ohta K, Otsuki M, Miyake N, Yoshihara Y and Iwasaki T 2005 Stereoscopic 3D display with optical correction for the reduction of the discrepancy between accommodation and convergence *J. Soc. Inf. Disp.* **13** 665–71

[23] Liu S, Cheng D and Hua H 2008 An optical see-through head mounted display with addressable focal planes *IEEE/ACM ISMAR* 33–42

[24] Gao K, McGinty C, Payson H, Berry S, Vornehm J, Finnemeyer V, Roberts B and Bos P 2017 High-efficiency large-angle Pancharatnam phase deflector based on dual-twist design *Opt. Express* **25** 6283–93

[25] Yin K, Xiong J, He Z and Wu S T 2020 Patterning liquid-crystal alignment for ultrathin flat optics *ACS Omega* **5** 31485–9

[26] Zhu S, Jin P, Qiao W and Gao L 2018 High-resolution head mounted display using stacked LCDs and birefringent lens *Proc. SPIE* **10676** 106761B

[27] Tan G, Zhan T, Lee Y H, Xiong J and Wu S T 2018 Polarization-multiplexed multiplane display *Opt. Lett.* **43** 5651–4

[28] Lee Y H, Peng F and Wu S T 2016 Fast-response switchable lens for 3D and wearable displays *Opt. Express* **24** 1668–75

[29] Lanman D and Luebke D 2013 Near-eye light field displays *ACM Trans. Graph.* **32** 1–10

[30] Martinez-Cuenca R, Saavedra G, Martinez-Corral M and Javidi B 2009 Progress in 3D multiperspective display by integral imaging *Proc. IEEE* **97** 1067–77

[31] Xiao X, Javidi B, Martinez-Corral M and Stern A 2013 Advances in three-dimensional integral imaging: sensing, display, and applications *Appl. Opt.* **52** 546–60

[32] LaRussa J A and Gill A T 1978 The holographic pancake window TM *0162* pp 120–9

[33] Narasimhan B A 2018 Ultra-compact pancake optics based on thineyes super-resolution technology for virtual reality headsets *Proc. SPIE* **10676** 106761G

[34] Wong T, Yun Z, Ambur G and Etter J 2017 Folded optics with birefringent reflective polarizers *Proc. SPIE* **10335** 103350E

[35] Maimone A and Wang J 2020 Holographic optics for thin and lightweight virtual reality *ACM Trans. Graph.* **39** 67

[36] Xiong J, Yin K, Li K and Wu S T 2021 Holographic optical elements for augmented reality: principles, present status, and future perspectives *Adv. Photon. Res.* **2** 2000049

[37] Hua H, Hu X and Gao C 2013 A high-resolution optical see-through head-mounted display with eyetracking capability *Opt. Express* **21** 30993–8

[38] Wang J, Liang Y and Xu M 2015 Design of a see-through head-mounted display with a freeform surface *J. Opt. Soc.* **19** 614–8

[39] Lee Y H, Zhan T and Wu S T 2019 Prospects and challenges in augmented reality displays *Virtual Real. Intell. Hardw.* **1** 10–20

[40] Levola T 2006 Diffractive optics for virtual reality displays *J. Soc. Inf. Disp.* **14** 467–75

[41] Mukawa H, Akutsu K, Matsumura I, Nakano S, Yoshida T, Kuwahara M and Aiki K 2009 A full-color eyewear display using planar waveguides with reflection volume holograms *J. Soc. Inf. Disp.* **17** 185–93

[42] Xiong J, Tan G, Zhan T and Wu S T 2020 Wide-view augmented reality display with diffractive cholesteric liquid crystal lens array *J. Soc. Inf. Disp.* **28** 450–6

[43] Yoo C, Bang K, Chae M and Lee B 2020 Extended-viewing-angle waveguide near-eye display with a polarization-dependent steering combiner *Opt. Lett.* **45** 2870–3

[44] Weng Y, Zhang Y, Cui J, Liu A, Shen Z, Li X and Wang B 2018 Liquid-crystal-based polarization volume grating applied for full-color waveguide displays *Opt. Lett.* **43** 5773–6

[45] David B R 2013 *Understanding Diffraction in Volume Gratings and Holograms* (London: IntechOpen)

[46] Bruder F K, Fäcke T, Hagen R, Hönel D, Orselli E, Rewitz C, Rölle T and Walze G 2015 Diffractive optics with high Bragg selectivity: volume holographic optical elements in Bayfol® HX photopolymer film *Proc. SPIE* **9626** 96260T

[47] Bruder F-K, Thomas F and Thomas R 2017 The chemistry and physics of Bayfol® HX film holographic photopolymer *Polymers* **9** 472

[48] Levola T and Laakkonen P 2007 Replicated slanted gratings with a high refractive index material for in and outcoupling of light *Opt. Express* **15** 2067–74

[49] Yin K, Zhan T, Xiong J, He Z and Wu S T 2020 Polarization volume gratings for near-eye displays and novel photonic devices *Crystals* **10** 561

[50] Yin K, Lee Y H, He Z and Wu S T 2019 Stretchable, flexible, rollable, and adherable polarization volume grating film *Opt. Express* **27** 5814–23

[51] Lee Y H, He Z and Wu S T 2019 Optical properties of reflective liquid crystal polarization volume gratings *J. Opt. Soc. Am.* B **36** D9–D12

[52] Xiang X, Kim J and Escuti M J 2018 Bragg polarization gratings for wide angular bandwidth and high efficiency at steep deflection angles *Sci. Rep.* **8** 1–6

[53] Yin K, Lin H Y and Wu S T 2019 Chirped polarization volume grating with ultra-wide angular bandwidth and high efficiency for see-through near-eye displays *Opt. Express* **27** 35895–902

[54] Frommer A 2018 Lumus optical technology for AR *SID Symp. Digest of Technical Papers* vol 48 pp 134–5

[55] Xu M and Hua H 2019 Methods of optimizing and evaluating geometrical lightguides with microstructure mirrors for augmented reality displays *Opt. Express* **27** 5523–43

[56] Maimone A, Lanman D, Rathinavel K, Keller K, Luebke D and Fuchs H 2014 Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources *ACM Trans. Graph.* **33** 89

[57] Yeom J, Jeong J, Jang C, Li G, Hong K and Lee B 2015 Three-dimensional/two-dimensional convertible projection screen using see-through integral imaging based on holographic optical element *Appl. Opt.* **54** 8856–62

[58] Lee S, Jang C, Moon S, Cho J and Lee B 2016 Additive light field displays: realization of augmented reality with holographic optical elements *ACM Trans. Graph.* **35** 1–13

[59] Yin K, He Z and Wu S T 2020 Reflective polarization volume lens with small f-number and large diffraction angle *Adv. Opt. Mater.* **8** 2000170

[60] Khorasaninejad M, Chen W T, Devlin R C, Oh J, Zhu A Y and Capasso F 2016 Metalenses at visible wavelengths: diffraction-limited focusing and subwavelength resolution imaging *Science* **352** 1190–4

[61] Li G, Lee D, Jeong Y, Cho J and Lee B 2016 Holographic display for see-through augmented reality using mirror-lens holographic optical element *Opt. Lett.* **41** 2486–9

[62] Peng H, Cheng D, Han J, Xu C, Song W, Ha L, Yang J, Hu Q and Wang Y 2014 Design and fabrication of a holographic head-up display with asymmetric field of view *Appl. Opt.* **53** H177–H185

[63] Westheimer G 1966 The maxwellian view *Vis. Res.* **6** 669

[64] Takahashi H and Hirooka S 2008 Stereoscopic see-through retinal projection head-mounted display *Proc. SPIE* **6803** 68031N

[65] Kim J *et al* 2019 Foveated AR: dynamically-foveated augmented reality display *ACM Trans. Graph.* **38** 1–15

[66] Jang C, Bang K, Li G and Lee B 2018 Holographic near-eye display with expanded eye-box *ACM Trans. Graph.* **37** 1–14

[67] Hong K, Yeom J, Jang C, Li G, Hong J and Lee B 2014 Two-dimensional and three-dimensional transparent screens based on lens-array holographic optical elements *Opt. Express* **22** 14363–74

[68] Hong K, Yeom J, Jang C, Hong J and Lee B 2014 Full-color lens-array holographic optical element for three-dimensional optical see-through augmented reality *Opt. Lett.* **39** 127–30

[69] Lee S, Wang M, Li G, Lu L, Sulai Y, Jang C and Silverstein B 2020 Foveated near-eye display for mixed reality using liquid crystal photonics *Sci. Rep.* **10** 1–11