Wireless Extended Reality (XR): Challenges and New Research Directions

Ian F. Akyildiz[1] **and Hongzhi Guo**[2]

[1]Truva Inc., Alpharetta, GA 30022, USA
[2]Engineering Department, Norfolk State University, VA, 23504, USA

Emails: ian@truvainc.com, hguo@nsu.edu

***Abstract*—Extended Reality (XR) is the umbrella term that includes Augmented Reality (AR), Mixed Reality (MR), and Virtual Reality (VR). XR has a tremendous market size and will profoundly transform our lives by changing the way we interact with the physical world. However, existing XR devices are mainly tethered by cables which limit users' mobility and Quality-of-Experience (QoE). Wireless XR leverages existing and future wireless technologies, such as 5G, 6G, and Wi-Fi, to remove cables that are tethered to the Head-Mounted Devices. Such changes can free users and enable a plethora of applications. High-quality ultimate XR requires an uncompressed data rate up to 2.3 Tbps with a latency lower than 1 ms. Although 5G has significantly improved data rates and reduced latency, it still cannot meet such high requirements. This article provides a roadmap towards wireless ultimate XR. The basics, existing products, and use cases of AR, MR, and VR are reviewed, upon which technical requirements and bottleneck of realizing ultimate XR using wireless technologies are identified. Challenges of utilizing 6G wireless systems and the next-generation Wi-Fi systems and future research directions are provided.**

***Index Terms*—6G, augmented reality, extended reality, mixed reality, virtual reality, Wi-Fi, wireless communication, wireless networks.**

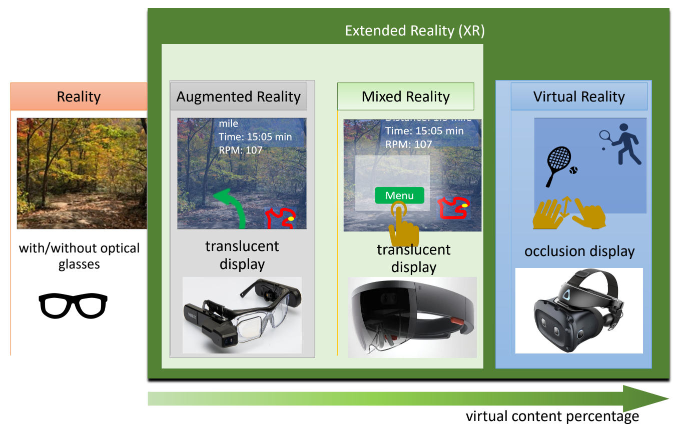

Fig. 1. Augmented Reality (AR), Mixed Reality (MR), Virtual Reality (VR), and Extended Reality (XR) in the Reality-Virtuality continuum. The AR device is VUZIX M4000, the MR device is Microsoft HoloLens 2, and the VR device is HTC Vive Cosmos Elite. The illustration pictures are used to show the concepts which are not created by these devices.

I. Introduction

In the 1960s, various Augmented Reality (AR) and Virtual Reality (VR) devices were invented, such as the Sensorama VR machine (1962) [1] and the Sword of Damocles AR machine (1968) [2]. Since then, AR and VR have been evolving with the development of sensors, displays, and computers. Recently, MR is emerging as we have the capability to interact with virtual/digital objects in real environments. AR, MR, and VR are all spatial computing technologies [3] which are encompassed by Extended Reality (XR). Their differences mainly reside in the rendering format and percentage of virtual content, as shown in Fig. 1. Today's XR technologies are mainly used for immersive gaming, remote assistance, and professional training [4], [5]. Customers have a wide variety of options which cost from around $300 to $5,000.

Existing XR devices use Head-Mounted Displays (HMDs) which have strict constraints on power consumption and weight. HMDs have to be made thin and light to meet the requirements of Quality-of-Experience (QoE). Thus, most computing and storage tasks are offloaded to a computer or a server to reduce the overall power consumption and weight of HMDs. Most existing XR devices use cables to connect HMDs with computers. This significantly limits users' mobility and QoE. Wi-Fi (802.11b/g/n/ac) and Bluetooth are adopted by mainstream XR devices to provide wireless services. Due to limited data rates of Wi-Fi and Bluetooth, they can only support entry-level low-quality XR.

5G cellular networks and Wi-Fi wireless systems have demonstrated to achieve peak data rates of several Gbps [6]–[9]. Using these networks, better wireless connections for XR can be realized. However, recent studies have shown that the ultimate XR requires uncompressed data rates of 2.3 Tbps with a latency lower than 1 ms which cannot be supported by existing 5G cellular networks and Wi-Fi [10].

On the long run, XR has the potential to replace computers and portable devices and become general computing platforms. As mentioned above, wireless communication based on 5G/6G and latest Wi-Fi systems will play an important role in realizing mobile XR. Here we point out that XR will not only be used for entertainment as it is the case currently, but also can transform the way we interact with the physical world in the future. Our objective in this paper is to fill the gap between XR and wireless communications

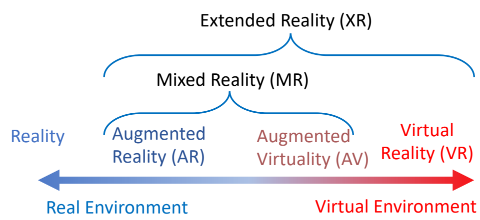

Fig. 2. Extended Reality (XR), Augmented Reality (AR), Mixed Reality (MR), and Virtual Reality (VR) in Miligram and Kishino's Reality-Virtuality Continuum.

research and highlight the according research challenges. The paper is organized as follows. In Section II, we introduce the basics of XR, including AR, MR, and VR, and highlight their differences. In Section III, we review existing representative products and point out the gap that need to be filled to realize wireless XR. Then, the current and potential XR use cases are presented in Section IV. After that, in Section V, we identify the technical design requirements that are related to wireless communication. Lastly, we introduce the grand research challenges in realizing wireless ultimate XR. Research problems and potential solutions are provided.

Note that, in the literature, some excellent works cover wireless mobile AR and/or VR [4], [5], [11]–[14]. This paper drastically differs from existing related work because we give a broad overview of XR, including AR, MR, and VR, in both indoor and outdoor environments. Also, we provide a research roadmap towards ultimate XR using 6G wireless systems and the next-generation Wi-Fi systems.

II. Basics of XR

Human perception of real objects is based on five basic senses: *sight, hearing, touch, smell, and taste*. If a virtual object can deliver the same synthesized senses as a real object, it seems that the virtual object does exist. The virtual content is created using digital technologies, which is also called digital reality [15]. Based on the format and the percentage of virtual content, we can divide XR into different categories, as shown in Fig. 2. According to Miligram and Kishino's Reality-Virtuality Continuum [16], the environment includes:

- *Reality:* The surrounding environments and objects are real;
- *Augmented Reality (AR):* the surrounding environments are *real* but enriched with virtual augmentations;
- *Augmented Virtuality (AV):* the surrounding environments are virtual but enriched with real augmentations;
- *Virtual Reality (VR):* the surrounding environments are fully virtual.

Note that, in AR, users can see virtual objects or information in real environments, whereas in AV, users can see real objects in virtual environments. The Mixed Reality (MR) includes AR and AV. In MR, the virtual objects can interact with the user and the real environment. In other words, MR has richer and more interactive virtual contents than AR. AR, MR, and VR share some common features and requirements, which are encompassed by XR [4], as shown in Fig. 1. The "X" in XR may represent any spatial computing technologies [3]. Although XR may include more technologies in the future, we focus on AR, MR, and VR in this paper.

A. Augmented Reality (AR)

As shown in Fig. 1, the reality is the physical world that we observe without any virtual content. AR overlays a virtual layer on top of reality. Next, we introduce two aspects of AR, namely, the content that AR provides and the device that can realize AR.

Contents: Virtual contents in AR are presented in two formats.

- *Virtual objects* are placed in real environments, and users cannot easily distinguish them from real objects. The widely used Pokémon GO is an example.
- *Virtual information*, such as real-time maps, notations, and sensory data, is provided to help users understand the real environment and provide the desired assistance. For example, AR navigation information can be displayed in real-time to assist drivers [17].

Devices: AR users can observe the real environment and virtual contents simultaneously. Currently, there are mainly two approaches to view AR contents:

- *Non-immersive AR* using phones, tablets, or any other handheld smart devices with cameras;
- *Immersive AR* using smart glasses or other Optical Head-Mounted Devices (OHMDs)

The non-immersive AR allows users to watch virtual contents through cameras on smart devices. For example, consider smartphone-based applications, cameras capture real-time real environments, and smartphones augment virtual contents and display mixed environments to users. The immersive AR presents mixed environments directly in users' sight. Even users turn their heads around, they can still observe virtual content. An example of AR OHMD is given in Fig. 3 with VUZIX M4000 AR glasses. MR has similar glasses and system architectures. The glasses send sensing information, such as head and eye tracking, and real-time videos of real environments to a server via wired or wireless communication. The server can be a computer, an edge server, or the core cloud, which performs environment and human understanding using Artificial Intelligence (AI), particularly Machine Learning (ML) algorithms, and renders virtual contents. The rendered video is sent back to the glasses for display.

Note that, depending on the display technique, the Augmented Reality (AR) can be divided into [18]

1) Optical See-Through (OST) and
2) Video See-Through (VST).

The OST glasses are translucent, as the one shown in Fig. 3. The VST glasses or headsets use a camera to capture real-time videos and then augment virtual contents onto them. A simple example of VST AR is the smartphone-based application. More complicated VST AR can use HMDs. This article mainly focuses on the OST AR, which is extensively used in high-end AR devices.

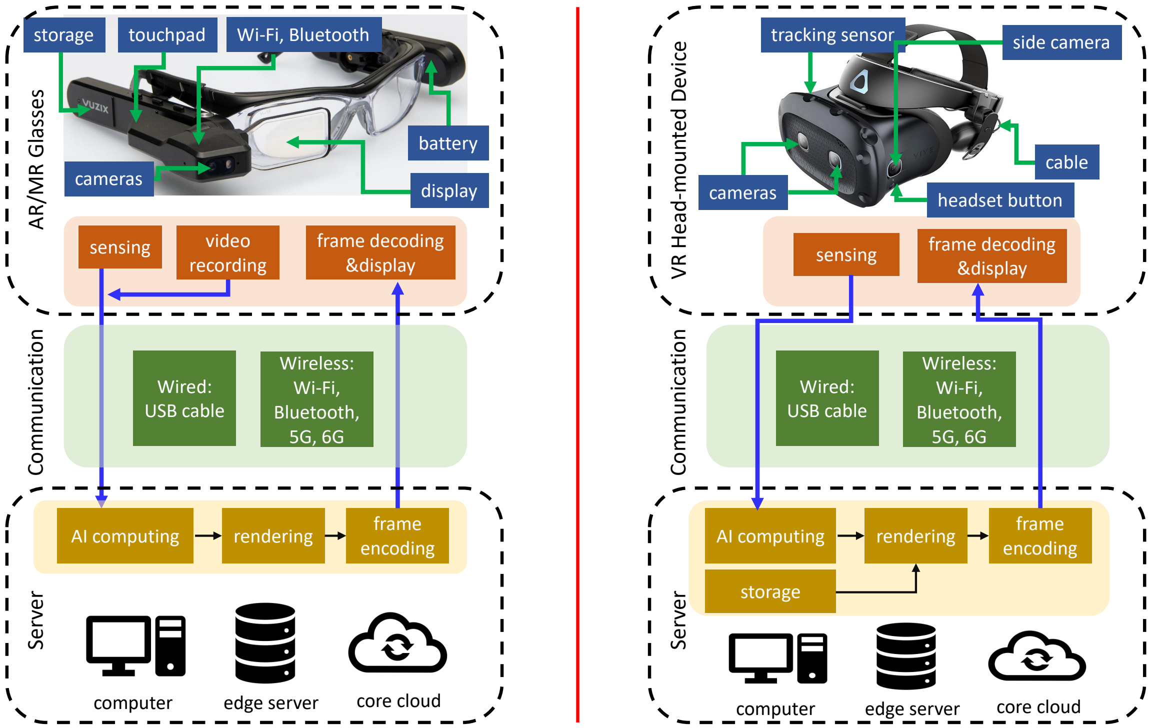

Fig. 3. Illustration of AR/MR (left) and VR (right) system diagrams. The AR glasses is VUZIX M4000 and the VR HMD is HTC Vive Cosmos Elite.

B. Mixed Reality (MR)

As shown in Fig. 1, MR is a broad concept containing all the technologies that mix real and virtual environments, including AR. AR and MR are used interchangeably in the literature due to the lack of a clear boundary. In this paper, we consider MR as an advanced version of AR. The MR contents and devices are summarized below.

Contents: MR presents richer and more capable virtual contents than AR. MR allows users to interact with active virtual contents, while AR only displays passive virtual contents.

Devices: Similar to AR, MR can also use smartphones and tablets. However, the Quality-of-Experiences (QoE) may not be acceptable. OHMDs are used extensively for MR because they can provide more immersive experiences compared with smartphones and tablets. MR can share the same system architecture as AR, as shown in Fig. 3. However, the AI computing of MR for the environment and human understanding is more complex than that of AR since it provides more interactions with virtual contents.

C. Virtual Reality (VR)

Below, we summarize VR contents and devices.

Contents: All the contents presented by VR are virtual which are not related to the user's real surrounding environment. VR synthesizes an immersive environment that can be isolated from the real world.

Devices: Occluded HMDs are used to block the real surrounding environment and provide the user with immersive experiences. Although VR can also be presented in smartphones and tablets, due to the non-immersive low-quality experience, we do not consider them in this paper. An illustration of VR system architecture is shown in Fig. 3. VR HMD sends sensing information, such as head and eye tracking, to the server which performs AI computing to understand users' behaviors. The connection between the HMD and the server can also be wired or wireless. VR does not require videos of the real environment since all the presented contents are virtual. Similar to AR and MR, the server can be a computer, an edge server, or the core cloud. Pre-created videos are saved in the storage which is rendered based on sensing information. The rendered video is sent back to the HMD for decoding and display.

D. Technical Differences

Although AR, MR, and VR have common features, we highlight the following major technical differences which affect the wireless system design.

- **Display:** AR and MR use OST-based translucent display installed on OHMDs. Users can observe the real environment. VR uses an occlusion display installed on HMDs; the real environment is not visible.
- **Human Understanding:** Interactions with virtual contents highly rely on external inputs. Voice control, touchpad, head tracking, eye tracking, and hand tracking are widely used in AR, MR, and VR. Currently, AR and MR are mainly used for assistance and training, while VR is mainly used for gaming. Therefore, AR and MR prefer an all-in-one format, which can be worn conveniently. Inputs, such as touchpad and buttons, are integrated into

the glasses. On the contrary, VR uses external controllers to provide more immersive gaming experiences.

- **Environment Understanding:** VR does not need to understand the user's real surrounding environment. Differently, AR and MR mix virtual contents with the real surrounding environment. Especially, when a virtual object is placed in a real setting, it has to be put at a suitable location, e.g., a virtual cup should be placed on a table rather than in the air. Thus, AR and MR must understand the environment. This is achieved by sending real-time videos by HMD cameras to the server, where AI algorithms are used. The computation in a VR server is to render videos based on users' Field-of-View (FoV) and inputs, whereas AR and MR servers recognize objects and create extra useful information using AI.
- **Uplink vs Downlink:** We define that the uplink channel (UL) is from the HMD/OHMD to the server and the downlink channel (DL) is in the opposite direction. For VR, the UL is used to send sensing information such as head moving and eye-tracking. The DL is used to send rendered videos. Thus, the DL requires larger bandwidth compared to the UL. AR and MR OHMDs also send sensing information through the UL. In addition, they stream real-time videos of the real surrounding environment, which require a large bandwidth. Generally, AR and MR ULs require higher data rates than VR.
- **Latency Tolerance:** AR and MR mix virtual contents with the real environment. The real environment is highly dynamic, e.g., the user is looking at a moving vehicle, AR and MR have to respond to these dynamics. Thus, they require ultra-low latency of video rendering. Moreover, MR is more challenging than AR since it interacts with virtual contents; it has more strict requirements on latency. The latency tolerance of VR depends on the specific application. For high-interactive VR, such as gaming, the latency tolerance is low and it requires ultra-low latency. For low-interactive VR, such as virtual movie theaters, the latency tolerance is high since the user barely interacts with the movie.

III. Existing XR Devices

There are various XR products on the market. We cannot enumerate all of them due to the limited space here. A few representative products are given in Table I with their technical specifications obtained in January 2022. We apologize in advance if some important products are missed here.

1) **Augmented Reality Devices:** These are used for controlling UAVs (unmanned aerial vehicles), remote support for field technicians, operations, and telemedicine. By wearing AR glasses, operators and technicians do not need to hold or look at smart devices, which improve their efficiency. Various AR products are available or under development, such as Epson Moverio BT300, VUZIX M4000, Apple Glasses, and Google Glass.
2) **Mixed Reality Devices:** Microsoft HoloLens is a representative product. Similar to AR, MR devices also aim to boost productivity for manufacturing, engineering and construction, healthcare, and education. The vendor also provides software support, and users can develop their own applications. Generally, MR glasses are more expensive than AR glasses due to their complicated functions.
3) **Virtual Reality Devices:** These are mainly used for gaming and movies. The Sony Playstation, Oculus Quest series, and HTC Vive series are popular products. Users can purchase VR games and movies in online stores.

From Table I we notice that the current products need to be improved in the following aspects to realize wireless XR in the future.

1) The existing XR devices' connectivity relies on Wi-Fis and cables. Most devices have the option of USB cables, which provide data communication and power. This significantly affects the user's mobility and user QoE. Moreover, the wireless options are available using Wi-Fi 5 but its peak data rate is not sufficient to support high-quality ultimate XR applications. Intel WiGig will be used for Vive Cosmos Elite which is based on millimeter-wave (mmWave) radios at 60 GHz. It can support 3 players with a range of 7 m. This is not widely used for other XR products.
2) The existing XR devices have limited computing and storage capabilities which makes it challenging to perform complex computing tasks, such as machine learning-based motion prediction and content caching. A computer or a server is necessary to run XR applications.
3) The QoE is limited by power consumption and headset weight. High power consumption of wireless communication and computation not only drains the battery fast but also generates heating which affects the user experience. Also, different from computers that are placed on desks, XR devices are wearable and their weight should be minimized. Today's XR devices can only support 2 to 3 hours of operation which is not sufficient for persistent applications. Their weight is around 500 g which is much higher than wearable optical glasses — the weight of standard optical glasses is around 20 g.

IV. Use Cases

In this section, we introduce representative use cases of XR, including AR, MR, and VR. With the 6G wireless systems and the next-generation Wi-Fis, we anticipate that these applications can be fully supported by wireless technologies.

A. Augmented Reality (AR)

1) Sports: AR can improve the performance of athletes and the QoE of the audience.

- With AR glasses, athletes can receive real-time AI support. For example, the optimal pass route can be displayed in soccer players' glasses, so that the success pass rate can be improved.
- The audience can observe the information that are not available before, such as players' names, ratings, and

TABLE I
EXISTING AR, MR, AND VR DEVICES.

	Vendor	Model	Weight (g)	Display (per eye)	Refresh rate (Hz)	Human understanding	Storage (GB)	Memory (GB)	Connectivity	Power (Hour)
AR	Epson	Moverio BT300	69	1280×720	30	controller	16	2	Wi-Fi, Bluetooth, cable	∼6
	VUZIX	M4000	∼246	854×480	–	touchpad, voice,buttons	64	6	Wi-Fi, Bluetooth, cable	2 to 12
MR	Microsoft	HoloLens2	566	2K	–	head/eye/hand tracking	64	4	Wi-Fi , Bluetooth	2 to 3
VR	Oculus	Quest 2	503	1832×1920	72	controller	256	6	Air Link (wireless)	2 to 3
	HTC	Vive Cosmos Elite	–	1440×1700	90	controller	–	–	cable, wireless adapter (60GHz)	2.5 (wire-less)
	Sony	Playstation	∼600	960×1080	120	controller	–	–	cable	–
	HP	Reverb G2	550	2160×2160	90	controller	–	–	Bluetooth, cable	–

The human understanding considers major input or tracking methods. Most devices have eye and head tracking which are not shown if they also use external controllers. The power of HTC Vive Cosmos Elite is estimated based on the power of its wireless adapter.

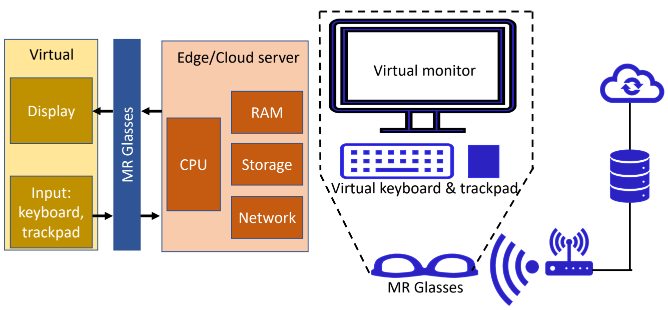

Fig. 4. Computer virtualization with MR virtual inputs and virtual display.

background information. For instance, sports fans sitting in the back of a big stadium cannot clearly see players. Their view can be augmented with players' names and performances close to the corresponding player.

2) Automotive Industry: AR plays an import role in the life cycle of motor vehicles, including design, manufacturing, sales, driving, and maintenance. AR can help designers choose different options by virtually manufacturing the car. Also, it can virtually change the car model for customers to select. AR has been used to provide collision warning, driver assistance, navigation, and lane departure warning for drivers with OHMDs [19]. Moreover, it is usually challenging for a driver to identify problems of a car due to the lack of professional training. With AR devices, AI can identify the problems and display them. Additional information such as nearby car repair shops or contact information can also be shown.

B. Mixed Reality (MR)

1) Computing Platform: Today's computers and portable devices have integrated inputs, computing units, storage, and outputs. Laptops and smart devices have significantly changed our lives thanks to their small size and lightweight. MR will provide a completely different computing platform by offloading computing, storage, and many other functions to the edge and cloud. The inputs, such as keyboard and trackpad, and outputs will become virtual using MR, as shown in Fig. 4. The computer manufacturers will not physically produce computers, instead, they will provide virtual computers in the edge and cloud. Users can utilize MR glasses with network access to create virtual inputs and displays. As 5G and 6G will provide ubiquitous wireless services, users can have access to their computers anytime anywhere. Besides computers, smartwatches and cell phones can also be accessed using MR devices.

2) Healthcare: MR will provide better medical services to the board community, especially for those without convenient access to medical providers. Doctors will leverage different medical examination results which can be projected in their MR glasses. AI-enabled identification and classification can also help doctors evaluate patients' health status and perform surgeries in real-time.

3) 3D Design: Existing 3D Computer-Aided Design (CAD) is conducted in a computer with 2D displays, which physically limits productivity. With MR, 3D CAD can be performed in a 3D space. 3D printing is a useful tool in various contexts, but the 3D model design is challenging for ordinary people without training. MR will reduce the complexity of 3D CAD design and make 3D printing more accessible to general users.

C. Virtual Reality (VR)

1) Personal Movie Theater: VR allows users to virtually sit in a movie theater with a wide virtual screen. Multiple users can watch a movie together just like they were in a real movie theater. This is a low-interactive VR application. Users do not interact with the virtual world. Therefore, it is relatively simple to implement. Most existing VR devices support this application.

2) Gaming: VR can place users into an immersive virtual world. Different from existing computer games using a keyboard and a mouse to control a character in the game, users play VR games as if they were in the real world. This is a high-interactive VR application, which has strict requirements on latency. Otherwise, users may feel nausea while they are playing.

Although not listed here, XR can support a plethora of other applications, such as remote education, holographic teleportation, retail, tourism, fitness and many more [5], [20].

V. Technical Requirements

In this section, we first review the Key Performance Indicators (KPIs) of XR. Then, we study the KPIs of the existing XR and the ultimate XR. We focus on the parameters that are related to wireless communication and networking.

A. Basics of XR Parameters

As discussed in Section II, the human perception is based on five senses: sight, hearing, touch, smell, and taste. To create a fully or partially virtual environment, we need to synthesize all of these senses. However, the sight is the most challenging sense because it requires a large amount of multimedia data. Thus, the existing XR KPIs are mainly related to videos.

The *Field-of-View (FoV)* is the angle of the maximum area that we can observe. Each human eye can cover nearly 130°. With two eyes, we can observe nearly 180°. In VR, a widely used term is 360° video which records every direction at the same time. This can be achieved by using special cameras or multiple regular cameras. Human eyes cannot observe such a wide-angle without turning heads.

The *Pixels-per-Degree (PPD)* is the number of pixels that are in the view for each degree. A large PPD means there are more pixels and the video is sharp and clear, vice versa.

The *Resolution* is the measurement of a video frame's width and height in pixels. Note that, resolution, FoV, and PPD are related, i.e., resolution = horizontal degree$\times$PPD$\times$vertical degree$\times$PPD. Thus, for a given resolution, a large FoV results in a small PPD. Existing AR and MR glasses have a small FoV, i.e., 20° to 50°, but their PPD is above 30. This provides the user with clear virtual contents in presence of real environments. On the contrary, VR displays have a large FoV, i.e., 100°- 150°, to provide users with immersive experiences. As a result, the PPD of VR displays is relatively small which is around 10 to 15.

The *Refresh Rate* is the number of video frames that can be displayed in one second. A low refresh rate for XR may result in headaches or nausea. Usually, a 90 Hz refresh rate is suggested for XR devices.

The *Data Rate* can be obtained based on the KPIs of XR videos, including the refresh rate, the resolution, and the number of bits of color.

TABLE II
Typical Existing VR, AR, and MR System Specifications and Technical Requirements.

Specification	AR	MR	VR
Screen	translucent	translucent	occlusion
Display	OHMD	OHMD	HMD
Environment	passive virtual & real	passive virtual, active virtual, & real	virtual
Uplink Data Rate	0.02 - 1.0 Gbps	0.02 - 1.0 Gbps	150 kbps
Downlink Data Rate	0.02 - 1.0 Gbps	0.02 - 1.0 Gbps	0.02 - 1.0 Gbps
Latency	20 ms	10 ms	20 - 1000 ms
Refresh Rate	∼90 Hz	∼90 Hz	∼90 Hz
Pixels-per-Degree	30 - 60	30 - 60	10 - 15
Field-of-View	20° - 50°	20° - 50°	100°- 150°

The *Latency* is the response time of XR devices when there is a change caused by the real environment or user. It is determined by the specific XR application. For example, the display currently shows frame A and, meanwhile, the HMD/OHMD camera captures a new frame B or sensors receive new inputs from the real environment or user. Now, frame B needs to be rendered with virtual content and this should be reflected in the next frame that is displayed. The latency that can be tolerated is the time from displaying frame A to displaying rendered frame B. Depending on the refresh rate, the latency tolerance can be as low as several milliseconds.

Next, we mainly focus on the data rate and latency which are two key parameters that affect wireless system design. As shown in Fig. 5, wireless XR will evolve with the development of wireless technologies. Currently, we are at the stage 1, where XR is moving from wired connections to wireless connections. The ultimate XR at the stage 3 will require several Tbps throughputs and ultra-low latency which can be lower than 1 ms. The stage 2 is a transition stage between stage 1 and stage 3. A summary of existing typical XR KPIs is given in Table II.

Note that, AR can be considered as a simple version of MR. Finally, the advanced AR technologies may be merged into MR. Thus, the ultimate XR only consists of MR and VR in Fig. 5.

B. Data Rates

Due to the environment and human understanding, AR and MR require similar UL and DL data rates. The OHMD has to send real-time videos to the server and the server sends back rendered videos. The UL of VR requires very low data rates, e.g., less than 150 kbps [21], since it only transmits sensing information. The DL of VR requires similar data rates as AR and MR.

Existing XR: Although current AR, MR, and VR displays have different FoV, their resolution and refresh rates are comparable, which require similar data rates. The resolution of the XR display is determined by the FoV and Pixels-Per-Degree (PPD) [10]. We use HTC Vive Cosmos Elite as an example to evaluate current requirements for wireless XR. Consider the 1440×1700 resolution (per eye) with a refresh rate of 90 Hz and 8 bits of color. The required data rate without compression is 10.6 Gbps. The data rate is obtained using $1440\times1700\times3\times8$(bits of color)$\times2$(2 eyes)$\times90$ (refresh rate). Using standard video lossy compression techniques with a 300:1 rate, the required data rate can be reduced to 35.3 Mbps. Intel WiGig wireless adapter can support 8 Gbps data rates which are sufficient to provide reliable wireless connections. However, this is only an entry-level VR that has relatively low PPD, refresh rate, and bit of color [22].

Ultimate XR: The Ultimate (or Extreme) XR, which is the stage 3 in Fig. 5, requires $360^\circ\times180^\circ$ full-view with 120 Hz refresh rate, 64 PPD, and 12 bits of color [10], [22]. Although a refresh rate higher than 120 Hz can improve the video quality, most users may not be able to distinguish the difference [10], [22]. Thus, the required data rate without compression is 2.3 Tbps. Using video lossy compression at the rate of 300:1, the reduced data rate is 7.7 Gbps. To reduce the required data rate, an FoV of $110^\circ\times110^\circ$ can be used. The updated data rates are 428.2 Gbps without compression and 1.4 Gbps with a compression rate of 300:1.

C. Latency

Existing XR: We can divide XR into two categories based on latencies, namely,

1) AR, MR, and high-interactive VR, and
2) low-interactive VR.

The former has strict requirements on latency and the latter can tolerate a certain latency. The latency is affected by the refresh rate. Currently, a 90 Hz refresh rate is widely used, which requires a latency smaller than 11 ms. As shown in Table II, some low-interactive VR can tolerate around 1000 ms latency. This is because the HMD can use a buffer to save multiple frames and play them with a certain delay, as shown in Fig. 5. This can effectively address network jitter. The buffer size and delay can be determined by the specific application's latency tolerance.

Ultimate XR: Based on the refresh rate of ultimate XR, the latency should be smaller than 8.3 ms. Note that, the latency consists of wireless communication, sensing data fusion, computing, access to edge or cloud servers, and display response time. Since the latency caused by each party is highly stochastic, the communication and networking latency should be much smaller than 8.3 ms to provide a high QoE.

VI. Challenges and Future Research Directions

Depending on the environment, we divide XR applications into two categories, namely, Local Area VR, AR, and MR (LAVAR) and Wide Area VR, AR, and MR (WAVAR). LAVAR supports applications in small areas, such as apartments, offices, and retail stores, whereas WAVAR supports applications in large areas, such as sports stadiums and autonomous vehicles, as shown in Fig. 5. Note that WAVAR is a broader concept than the Mobile AR (MAR) [4], [23], and it aims to provide ubiquitous wireless services for XR. Next, we study the grand challenges of realizing the wireless ultimate XR. Potential solutions and future research directions are also provided.

A. Data Rates

Our vision is that WAVAR will use 6G wireless systems, whereas LAVAR will use the next-generation Wi-Fis. In this way, we can achieve seamless ubiquitous connectivity. Note that using only 6G wireless systems may not not practical because the mmWave, Terahertz, and visible light signals experience significant propagation losses due to building blockages. Deploying more 6G base stations cannot effectively solve this problem because the base stations are much more expensive than the Wi-Fi access points. It is more economical to use Wi-Fi for LAVAR instead of 6G wireless systems.

Though 5G promises a peak data rate of 20 Gbps [6], recent network measurements show that the achievable data rates are around 0.1 to 2.0 Gbps [8], [9]. Since the requirement of existing entry-level XR is lower than 1.0 Gbps, 5G can provide sufficient data rates. However, the ultimate XR requires much higher data rates than that provided by 5G. The envisioned 6G wireless system has a peak data rate of 1 Tbps and an experience data rate of 1.0 Gbps [6]. Such high data rates will enable the use of high-quality ultimate WAVAR.

Most existing XR devices support Wi-Fi 5 which cannot provide sufficient data rates for ultimate XR applications. LAVAR will employ the next-generation Wi-Fi systems, such as 802.11be (around 46 Gbps) [7] and 802.11ay (around 100 Gbps) [25]. Such high data rates together with the data compression techniques, such as H.266 (Versatile Video Coding) [26], Wi-Fi systems can support ultimate LAVAR. WAVAR and LAVAR have the following specific challenges to achieve and maintain high data rates.

1) Novel Wireless System Design: The high data rates in 6G wireless systems rely on novel wireless communication systems, such as Terahertz, mmWave, and Visible Light Communication (VLC). mmWave bands have received significant attention in 5G systems; they may still play an important role in 6G. Terahertz wireless communication systems have been developed for more than a decade, but there are still open research problems, such as optimal resource allocation in the Terahertz band, co-design of sensing, communication, and intelligence, and beamforming among others [27], [28].

2) Unreliable/Blocked Wireless Environment: A novel design is required to avoid blockages in indoor and outdoor environments. For example, in VR gaming, the human body may block mmWave or Terahertz signals intermittently. In [29]–[33], Reconfigurable Intelligent Surfaces (RIS) have been used to create extra propagation paths. This increases the system reliability by providing redundant propagation paths in case of blockages, as shown in Fig. 5. It is challenging to provide reliable high data rates considering the stochastic nature of the wireless channel. Adaptive protocols are needed

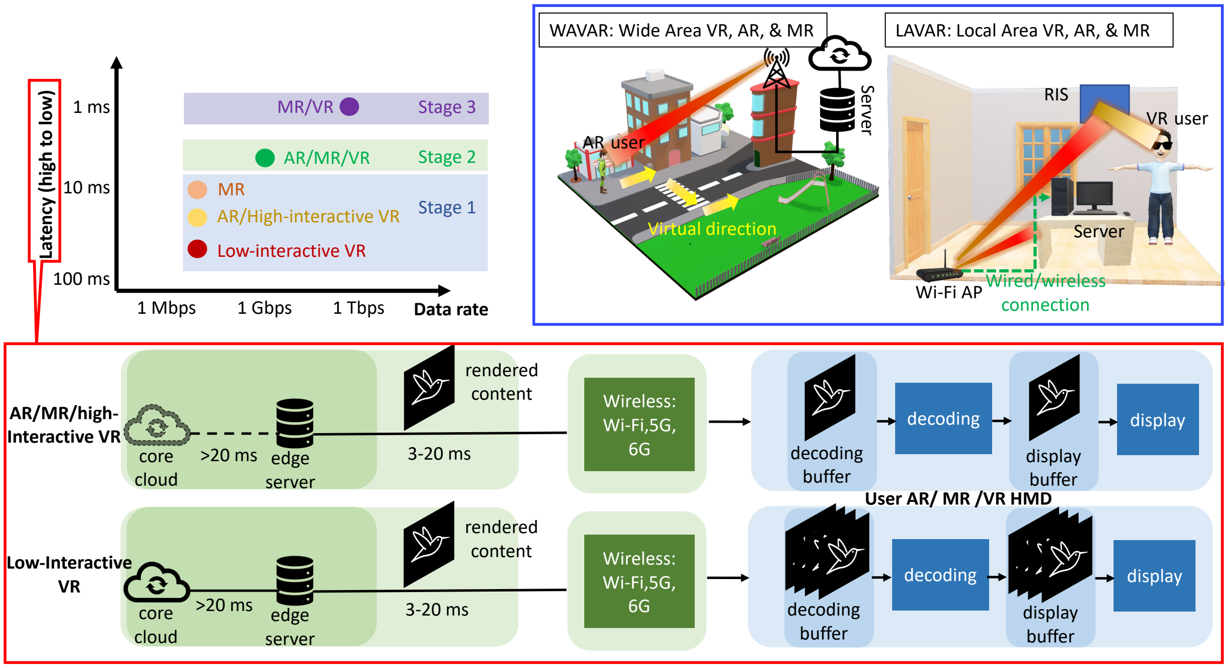

Fig. 5. Stages of XR development in terms of data rate and latency (top left); XR use cases for Wide Area VR, AR, and MR (WAVAR) and Local Area VR, AR, and MR (LAVAR) (top right); Sources of XR latency (bottom). Values of latency are from [24].

to optimally control communication systems to meet required QoE.

3) Multi-Users: WAVAR has to support a large number of devices (tens of thousands) simultaneously. Consider the AR sports where fans can see players' names and performances that are displayed close to players in real-time. A potential solution can divide the large stadium into small LAVAR and use multiple Wi-Fi access points to provide AR service. Interference should be considered when planning router locations.

B. Latency

The major sources that generate latency in wireless XR systems are shown in Fig. 5. For AR, MR, and high-interactive VR, video frames are displayed immediately which can tolerate extremely low latency, while for low-interactive VR, video frames can be buffered and the latency tolerance is high.

1) Wireless and Wireline Latency Minimization: Today's 5G end-to-end delay is much higher than 10 ms which needs significant improvements. For example, recent network measurements show that the end-to-end delay in 5G networks is around 21.8 ms and 27.4 ms [8], [9]. 6G proposes to reduce the delay to around 1 ms. The radio access networks using mmWave and Terahertz can achieve very low latency, but it is usually neglected that the wireline communication also needs to be updated to support 6G networks. Also, some cellular core network functions can be moved to base stations to reduce the access delay. The Device-to-Device (D2D) communication at mmWave and Terahertz bands can further reduce the delay by allowing the user to directly communicate with the local server.

2) Trade-off between Video Encoding and Wireless Communication: High-quality video encoding and decoding may take longer than 10 ms which is even larger than the overall latency requirements. Usually, on one hand, video is encoded/compressed before wireless transmission to reduce the communication latency. On the other hand, video encoding and decoding increase the latency, which is usually much smaller than the communication latency in wireless networks. Since 5G and 6G have high communication data rates, the communication latency can be significantly reduced. It is not clear whether we still need high compression ratios if the communication channel can allow high volume data transmission. Adaptive encoding algorithms considering wireless communication channels can be more efficient.

3) Edge Computing and Caching: Wireless XR will leverage edge computing and caching due to the following reasons.

- High-bandwidth cloud computing services are expensive which may cost several thousand dollars per month. Compared to existing XR devices which cost around \$300 to \$5,000, cloud computing services with GPU servers may not be practical.
- The latency is also affected by the path length. Using cloud services may create significant traffic in the network and increase latency.

As shown in Fig. 5, AR, MR, and high-interactive VR rely on edge servers because the communication with the core cloud may generate significant latency, which cannot meet the latency requirements. Edge servers are close to users which incur short delays. Also, local information, such as indoor environment and street information, can be cached in edge servers which will significantly reduce the latency and improve computing efficiency. For example, the pictures or videos created in the same apartment have significant identical contents which can be cached and reused.

4) Software-Defined Networking (SDN), Network Function Virtualization (NFV), and Automatic Network Slicing: 6G wireless systems are envisioned to support a wide variety of applications and XR is only one of them. Considering the fast-growing network traffic, the automatic network slicing with the support of SDN and NFV is necessary to prioritize XR applications in order to reduce the latency [6], [34]. Optimal network slicing algorithms for XR applications are desirable in the era of 6G to efficiently manage and use networking resources [35].

C. Mobility

Mobile WAVAR can be used for navigation for automobiles and pedestrians. Due to the short range of mmWave and Terahertz wireless systems, the mobility incurs frequent handoffs which cause long latencies. The soft handoff that allows multiple connections is necessary for seamless connections. Also, deep learning-based motion prediction in conjunction with network scheduling can be used to plan resource allocation for the user. UAVs provide a large coverage area which can reduce the number of handoffs. The UAV trajectory and location can be jointly designed with XR users' mobility. Motion prediction also allows LAVAR to pre-render content and reduce the latency [32]. Since mmWave and Terahertz wireless systems are highly directional, beam steering considering the user's mobility is challenging. Motion prediction is necessary for accurate and efficient beamforming.

D. Weight and Power Consumption

Different from laptops and smartphones, XR HMDs are worn on the head. The weight should be small in order to improve the QoE. A heavy HMD may not be accessible to everyone. However, advanced computation, communication, sensing, and display require bulky devices. Moreover, high-power consumption also requires large batteries to prolong the operation time. All these factors can increase the weight of HMD. Also, the high-power consumption may generate heating which makes the HMD not wearable. To make wireless XR practical, low-power communication, computation, and networking protocols have to be employed. Simultaneous wireless information and power transfer at the mmWave and Terahertz bands have the potential to partially address this issue.

E. Collaborative XR

Collaborative XR will enable multiple users to work on the same task simultaneously [36]. For example, LAVAR can support several doctors to work collaboratively on a surgery. This is a challenging problem because it requires higher network throughput and ultra-low latency. Serving multiple users simultaneously includes various wireless communication and networking problems, such as synchronization and end-to-end latency minimization. Using one mobile edge server may not be sufficient, and multiple RISs are necessary to control wireless signal propagation. The intelligent communication environments [6] can be an efficient solution to meet such high requirements.

F. Other Challenges

XR is a complex system and there are many other challenges that are directly or indirectly related to wireless communication and networking.

Security and privacy of XR are of paramount importance, especially AR and MR that combines real and virtual environments. For example, if an attacker modifies the traffic light, speed limit, and road symbol signs, users or autonomous vehicles may be misled to make life-threatening decisions. The data storage and communication have to be protected, and intelligent applications can be installed in XR devices to detect and correct malicious information.

Wireless sensing can be integrated into XR devices to reduce the use of peripheral sensors. In this way, XR devices can be made more compact. The use of Terahertz wireless communication for high-data-rate communication can also provide unprecedented wireless sensing accuracies due to its short wavelength. Although the human body can block Terahertz signals, this also provides information about the motion of the human body. In conjunction with optical cameras, this can provide accurate motion sensing.

Operating systems dedicated for XR are also desirable to efficiently manage applications, hardware, energy, data communication, security and privacy, and display among others. XR devices will support a plethora of applications simultaneously and integrate a wide variety of intelligent things. The future networked XR will connect a large number of XR devices. Such complicated systems require operating systems to manage resources and tasks accordingly.

VII. Conclusion

Extended Reality (XR) consisting of Augmented Reality (AR), Mixed Reality (MR), and Virtual Reality (VR) will soon become the next generation mobile computing platform that can make rapid and profound changes in our lives, just as the changes that laptops and smartphones have brought to us. However, today's XR devices are mainly tethered using cables which limit their mobility and potentials. In this paper, we introduce wireless XR systems and discuss their requirements of wireless data rates and latency, as well as their use cases. Research challenges and potential solutions to realize the environed indoor and outdoor applications are provided. 6G wireless systems and the next-generation Wi-Fis will allow XR users to move without hindrance and they can also support multiple users simultaneously, which are the enablers of high-quality ultimate XR.

References

[1] M. L. Heilig, "Sensorama simulator," Aug. 28 1962, US Patent 3,050,870.

[2] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 1968, pp. 757–764.

[3] S. Shekhar, S. K. Feiner, and W. G. Aref, "Spatial computing," *Communications of the ACM*, vol. 59, no. 1, pp. 72–81, 2015.

[4] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, "A survey on mobile augmented reality with 5g mobile edge computing: Architectures, applications, and technical aspects," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1160–1192, 2021.

[5] T. Braud, F. H. Bijarbooneh, D. Chatzopoulos, and P. Hui, "Future networking challenges: The case of mobile augmented reality," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 1796–1807.

[6] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[7] E. Khorov, I. Levitsky, and I. F. Akyildiz, "Current status and directions of IEEE 802.11 be, the future Wi-Fi 7," *IEEE Access*, vol. 8, pp. 88 664–88 688, 2020.

[8] D. Xu, A. Zhou, X. Zhang, G. Wang, X. Liu, C. An, Y. Shi, L. Liu, and H. Ma, "Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption," in *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication*, 2020, pp. 479–494.

[9] A. Narayanan, E. Ramadan, J. Carpenter, Q. Liu, Y. Liu, F. Qian, and Z.-L. Zhang, "A first look at commercial 5G performance on smartphones," in *Proceedings of The Web Conference 2020*, 2020, pp. 894–905.

[10] H. Zhang, Z. Yang, and P. Mohapatra, "Wireless access to ultimate virtual reality 360-degree video," in *Proceedings of the International Conference on Internet of Things Design and Implementation*, 2019, pp. 271–272.

[11] F. Hu, Y. Deng, W. Saad, M. Bennis, and A. H. Aghvami, "Cellular-connected wireless virtual reality: Requirements, challenges, and solutions," *IEEE Communications Magazine*, vol. 58, no. 5, pp. 105–111, 2020.

[12] E. Bastug, M. Bennis, M. Médard, and M. Debbah, "Toward interconnected virtual reality: Opportunities, challenges, and enablers," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 110–117, 2017.

[13] X. Hou, Y. Lu, and S. Dey, "Wireless vr/ar with edge/cloud computing," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2017, pp. 1–8.

[14] A. Szajna, R. Stryjski, W. Woźniak, N. Chamier-Gliszczyński, and M. Kostrzewski, "Assessment of augmented reality in manual wiring production process with use of mobile ar glasses," *Sensors*, vol. 20, no. 17, p. 4755, 2020.

[15] Y. Kang and K. C. Yang, "Employing digital reality technologies in art exhibitions and museums: A global survey of best practices and implications," in *Virtual and augmented reality in education, art, and museums*. IGI Global, 2020, pp. 139–161.

[16] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[17] C. Cao, Z. Li, P. Zhou, and M. Li, "Amateur: augmented reality based vehicle navigation system," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, pp. 1–24, 2018.

[18] F. Cutolo, U. Fontana, M. Carbone, R. D'Amato, and V. Ferrari, "[poster] hybrid video/optical see-through hmd," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, 2017, pp. 52–57.

[19] C. Merenda, H. Kim, K. Tanous, J. L. Gabbard, B. Feichtl, T. Misu, and C. Suga, "Augmented reality interface design approaches for goal-directed and stimulus-driven driving tasks," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 11, pp. 2875–2885, 2018.

[20] C. Westphal, "Challenges in networking to support augmented reality and virtual reality," *IEEE ICNC*, 2017.

[21] Z. Tan, Y. Li, Q. Li, Z. Zhang, Z. Li, and S. Lu, "Supporting mobile vr in lte networks: How close are we?" *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 2, no. 1, pp. 1–31, 2018.

[22] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran, and M. D. Silva, "VR is on the edge: How to deliver 360 videos in mobile networks," in *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network*, 2017, pp. 30–35.

[23] Q. Liu, S. Huang, J. Opadere, and T. Han, "An edge network orchestrator for mobile augmented reality," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 756–764.

[24] GSMA Future Networks, "Cloud ar/vr whitepaper," 2019. [Online]. Available: https://www.gsma.com/futurenetworks/wiki/cloud-ar-vr-whitepaper/

[25] Y. Ghasempour, C. R. Da Silva, C. Cordeiro, and E. W. Knightly, "Ieee 802.11 ay: Next-generation 60 ghz communication for 100 gb/s wi-fi," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.

[26] B. Bross, J. Chen, J.-R. Ohm, G. J. Sullivan, and Y.-K. Wang, "Developments in international video coding standardization after avc, with an overview of versatile video coding (vvc)," *Proceedings of the IEEE*, 2021.

[27] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical communication*, vol. 12, pp. 16–32, 2014.

[28] I. F. Akyildiz, C. Han, Z. Hu, S. Nie, and J. M. Jornet, "Terahertz band communication: An old problem revisited and research directions for the next decade," *arXiv preprint arXiv:2112.13187*, 2021.

[29] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, 2018.

[30] C. Liaskos, G. Pyrialakos, A. Pitilakis, A. Tsioliaridou, M. Christodoulou, N. Kantartzis, S. Ioannidis, A. Pitsillides, and I. Akyildiz, "The internet of metamaterial things and their software enablers," *Int. Telecommun. Union J*, vol. 1, no. 1, pp. 55–77, 2020.

[31] C. Liaskos, L. Mamatas, A. Pourdamghani, A. Tsioliaridou, S. Ioannidis, A. Pitsillides, S. Schmid, and I. Akyildiz, "Software-defined reconfigurable intelligent surfaces: From theory to end-to-end implementation," *IEEE Proceedings*, 2022.

[32] X. Liu, Y. Deng, C. Han, and M. Di Renzo, "Learning-based prediction, rendering and transmission for interactive virtual reality in ris-assisted terahertz networks," *IEEE Journal on Selected Areas in Communications*, 2021.

[33] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, and P. Popovski, "Risk-based optimization of virtual reality over terahertz reconfigurable intelligent surfaces," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[34] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, "5g network slicing using sdn and nfv: A survey of taxonomy, architectures and future challenges," *Computer Networks*, vol. 167, p. 106984, 2020.

[35] S. Vassilaras, L. Gkatzikis, N. Liakopoulos, I. N. Stiakogiannakis, M. Qi, L. Shi, L. Liu, M. Debbah, and G. S. Paschos, "The algorithmic aspects of network slicing," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 112–119, 2017.

[36] M. Billinghurst and H. Kato, "Collaborative augmented reality," *Communications of the ACM*, vol. 45, no. 7, pp. 64–70, 2002.