HOLOGRAPHIC-TYPE COMMUNICATION: A NEW CHALLENGE FOR THE NEXT DECADE

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

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

NOTE: Corresponding author: Hongzhi Guo, hguo@nsu.edu

Abstract – *Holographic-Type Communication (HTC) has been identified as an important service that will be supported by 6G wireless systems. It provides truly immersive experiences for a large number of novel applications, such as telepresence conferences and holographic healthcare, by displaying multi-view high-resolution 3D holograms of human beings or objects and creating multi-sensory media (Mulsemedia), including audio, haptic, smell, and taste. HTC will play an important role in realizing Metaverse by connecting the physical and virtual world seamlessly. HTC faces great challenges in transmitting high volume data with high throughput and guaranteed end-to-end latency which cannot be addressed by existing communication technologies. The contribution of this paper is two-fold. First, this paper introduces the basics and generic architecture of the HTC systems. The encoding and decoding of hologram and mulsemedia are discussed, and envisioned use cases and technical requirements are introduced. Second, this paper identifies the limitation of existing wireless and wired networks in realizing HTC and points out the promising 6G and beyond networking technologies. Particularly, on the HTC source side, the point cloud encoding and mulsemedia synchronization solutions are introduced and explained. On the HTC networking side, many new directions and according challenges such as semantic communications, deterministic networks, time-sensitive networks, federated networks, distributed encoding and decoding, and predictive networks are covered as they may help to satisfy the high data rate and guaranteed end-to-end latency requirements of the HTC. On the HTC destination side, the heterogeneity of HTC devices, synchronization, and user motion prediction are explored and according research challenges are highlighted.*

Keywords – 6G and beyond, holographic-type communication, holographic teleportation, metaverse, mulsemedia, point cloud encoding, mulsemedia synchronization, semantic communications, deterministic/time sensitive/federated/predictive networks, distributed encoding and decoding.

1. INTRODUCTION

In 2008, the Cisco On-Stage TelePresence Experience was demonstrated where a user from India met virtually with the other two users from the United States. Different from video conferencing where remote users are displayed on a 2D (two-dimensional) screen, this holographic system displayed the two remote users on a 3D (three-dimensional) display as holograms which had the same physical dimension as users. With sufficient resolution, the holograms looked the same as remote users and, thus, the system could provide an immersive experience. More recently, the holographic technologies have been significantly improved and widely used in various entertainment events, such as The Whitney Houston Hologram Tour since 2020 [1] and the ABBA Voyage Concert in 2022 [2].

A hologram is a recording of a light field that consists of the original properties of depth and parallax of 3D human beings and objects. Holography is a technology that is used to record and generate holograms. Holography was developed by Dennis Gabor in 1948 [3], who received the Nobel Prize for Physics in 1971. Since then, various technologies have been developed to create and display holograms. Different from videos and images, which are displayed on 2D screens, a hologram requires a large amount of data to demonstrate the truly 3D structure with additional depth and parallax features. Typically, holograms require several Gbps or even Tbps data rates depending on the hologram size and resolution. As a result, hyper-real holograms are usually generated and (partially) displayed at the same location without any remote transmission due to the limited network bandwidth.

What is Holographic-Type Communication (HTC)? HTC can send holograms and other Multi-sensory media (Mulsemedia) through wireless and wired networks to virtual or remote physical locations. The HTC system mainly comprises three parts, namely, the source, HTC networks, and the destination. Holograms and mulsemedia are created or stored at the source, sent through HTC networks, and rendered and presented at the destination. To provide truly immersive experiences, HTC will leverage all five senses of human perception, including sight, hearing, touch, smell, and taste. Future HTC systems may even include senses such as balance, wind, moisture, etc. Generally, the media that include three or more human senses are mulsemedia [6], [7], which will play an important role in HTC systems. Mulsemedia is created using various sensors on the source side and presented using different actuators on the destination side. HTC is one of the enabling technologies for Metaverse which is a net-

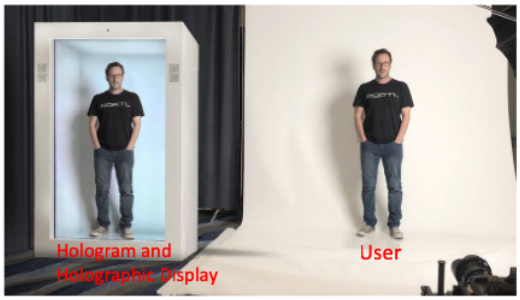


Fig. 1 – David Nussbaum (the founder of US holograms firm Proto) on the right-hand side and his hologram on the left-hand side. The holographic system is Proto. Image from [4], [5]

work of 3D virtual worlds with a focus on social connections [8].

Why is HTC Challenging? HTC requires a large amount of holographic and multi-sensory data which needs to be delivered to the destination with guaranteed and bounded latency. Traditional multimedia communications buffer transmitted or received data to overcome network latency and jitter issues. However, HTC requires high data rates (as high as several Tbps) and unprecedented large buffers, which cannot be supported by existing systems with limited memory. Specifically, the grand challenges for HTC systems include:

- **Develop efficient hologram encoding and decoding techniques**. Directly sending uncompressed holographic data can generate significant network traffic which increases the end-to-end latency and packet losses. Nevertheless, there is a trade-off between encoding and decoding efficiency and the latency due to computation, i.e., a high encoding rate results in long computation latency, and vice versa. It is a challenging problem to jointly design optimal encoding and decoding techniques by considering real-time network status and user Quality-of-Experience (QoE) requirements.

- **Deliver data packets with guaranteed and bounded end-to-end latency that can be defined automatically or by users**. Since buffer HTC data is challenging due to the large data size, it is desirable to deliver HTC data packets at a predefined time in a deterministic way. In other words, multiple flows of HTC data packets can be delivered in a synchronized way. Existing networks use the best-effort delivery, where the probability density function of latency exhibits a long tail, and users cannot configure the latency. HTC Networks have to allow applications/users to define the latency parameters and deliver data packets with guaranteed and bounded end-to-end latency.

- **Precisely synchronize multiple senses**. HTC systems use multiple human perception senses which require synchronization of multiple source sensors, displays, speakers, and actuators. Any asynchronous data can significantly reduce the QoE. Precise synchronization also increases end-to-end latency, memory, and computation complexity. It is challenging to synchronize so many senses with limited resources and guaranteed and bounded end-to-end latency requirements.

- **Design resilient and intelligent rendering algorithms in presence of packet losses and errors**. Packet losses and errors can incur retransmission in traditional networks which in turn increases the end-to-end latency. To avoid this issue, HTC networks do not allow retransmission and the destination should have the intelligence to correct errors and provide high QoE in presence of packet losses.

- **Support high-interactive applications**. High-interactive HTC applications usually require more than 60 frames per second to capture motions. A single frame of holographic data is large, and a high frame rate can dramatically increase the required data rates and network bandwidth. This also requires low end-to-end latency, e.g., smaller than 1 ms. Considering that existing network end-to-end latency is higher than 1 ms and the encoding and decoding latency can be as high as several hundreds of milliseconds, supporting high-interactive HTC application is a great challenge.

These challenges cannot be addressed by existing solutions. For example, the Ultra-Reliable Low Latency Communication (URLLC) can provide high reliability and low latency of 1 ms, but the communication throughput cannot achieve several Gbps to Tbps. The design of HTC will involve many aspects of technical innovations that are not available today.

A Brief History of HTC. The transportation of holograms has been studied and implemented for more than a decade in some ideal scenarios or with reduced QoE. In 2008, the Cisco On-Stage TelePresence Experience was demonstrated, where a user on the Bangalore stage could have a face-to-face meeting with users from San Jose who were displayed on a 3D holographic display. This project was envisioned to enable holographic communication for telepresence conferences. The Telehuman project (the Telehuman in 2012 [9] and the Telehuman2 in 2018 [10]) demonstrated low-cost designs of HTC using cylindrical holographic displays. The Telehuman2 can provide 10 frame-per-second (fps) and the latency from capture to projection is about 200 ms. Also, instead of using advanced light field display, Head-Mounted Displays (HMD) for eXtended Reality (XR) [11], e.g., Microsoft HoloLens, are used to display holograms. The LiveScan3D developed in 2015 can display captured holograms in the same room [12]. In [13], three different communication distances are considered. The source is located at Guildford in the UK and the three destinations are located in London in the UK (Round Trip Time (RTT) is around 4 ms), Virginia in the United States (RTT is around 85 ms), and Seoul in South Korea (RTT is around 285 ms). Gener-

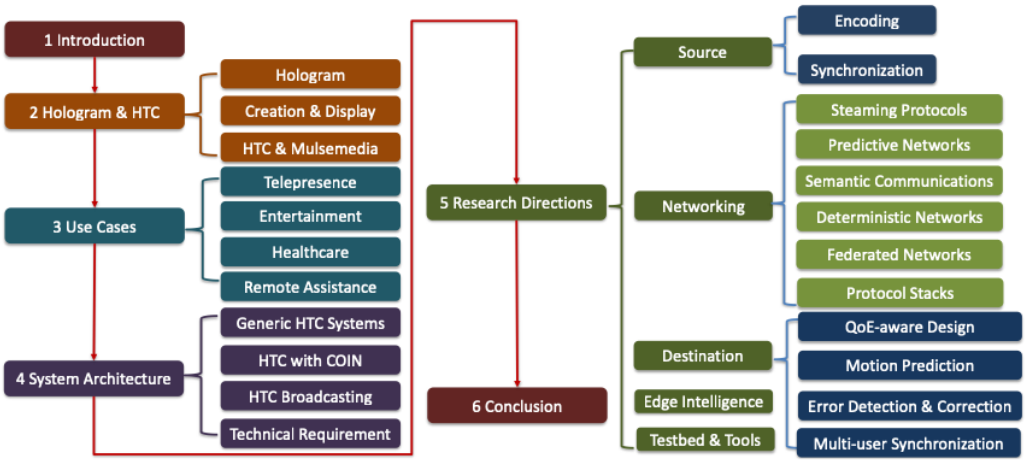


Fig. 2 – The structure of this paper.

ally, the user in London can obtain 30 fps reliably, while the user in Seoul experiences degraded performances. Although the past decade has witnessed a surge in HTC, existing prototypes are constrained by low frame rates and resolution. It is a great challenge to transmit holographic data in existing networks due to the large data size and strict requirement of latency. The development of the next-generation wireless and wireline networks, such as 5G, 6G and beyond, and the employment of Artificial Intelligence (AI)/Machine Learning (ML) in networking and communications, have made it possible to provide high-quality hyperreal HTC. It is anticipated that HTC will be an important use case in the Networking 2030 [14].

Contributions. The contribution of this paper is two-fold:

- First, we introduce the basics and the generic system architecture of HTC with emphasis on the technologies used for sources, networking, and destinations. Also, typical use cases are provided and the related technical challenges are discussed. This part aims to introduce HTC systems in a general way.
- Second, we identify the research gaps and challenges for HTC systems and point out potential solutions that need to be developed in 6G and beyond wireless systems. Specifically, we discuss the use of source encoding and decoding, in-network computation, mulsemedia, and adaptive streaming protocols, wireless sensing, semantic communication, deterministic networks, and federated networks for HTC systems. These technologies have the potential to improve network throughput and provide services with guaranteed and bounded end-to-end latency. Also, we summarize existing research progresses and available open-source research tools, which can facilitate the development of HTC systems.

It should be noted that HTC is different from holographic MIMO surfaces [15]. HTC is a kind of multimedia/mulsemedia communication technology and the hologram in HTC is a 3D representation of objects or users. Holographic MIMO surfaces use a large number of antennas or intelligent surfaces to generate arbitrary beams to improve wireless communication performance. The hologram in holographic MIMO is generated using wireless signals at different frequency bands.

The rest of this paper is organized as follows. In Section 2, the basics of hologram and HTC are introduced. After that, the HTC use cases are discussed in Section 3. Section 2 and Section 3 explain what HTC is and the differences between HTC and existing multimedia and XR technologies. The HTC system architecture and basic technical requirements are given in Section 4. Then, the research directions and potential solutions for HTC are discussed in Section 5. Finally, this paper is concluded in Section 6. The structure of this paper is shown in Fig. 2.

2. HOLOGRAM AND HOLOGRAPHIC-TYPE COMMUNICATION

In this section, we introduce how holograms are created and displayed. Then, we discuss the five senses of human perception and mulsemedia. Last, we introduce HTC and its advantages in creating truly immersive experiences.

2.1 Hologram

A hologram is a recording of the light field which preserves the original depth and parallax of real 3D objects. Computer-generated holograms can be divided into two categories, namely, image-based solutions and volume-based solutions [16]. The image-based solution such as light field video relies on a large number of images from different angles captured by camera arrays. The resolution is determined by the spatial angle interval between cameras. The volume-based solution such as point cloud represents real 3D objects using 3D volume pixels. The hologram is different from typical 3D images. 3D images rely on special glasses to generate 3D effects. On the contrary, the hologram can be observed with naked eyes.

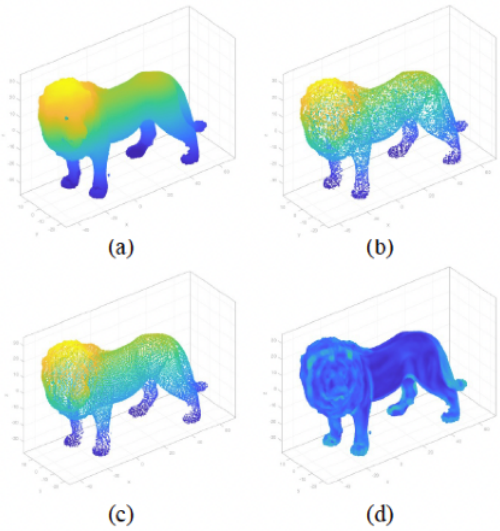


Fig. 3 – Point cloud of a lion: (a) 100% points; (b) randomly sampled 10% points; (c) interval sampling of every 10 points of the point cloud data; and (d) 100% points considering roughness of [0,1.5]. Code was adopted from [17].

2.2 Hologram Creation and Display

Camera arrays are used to capture images of the 3D object from different angles. These images are processed to generate 3D objects based on information of depth and parallax. Next, we mainly focus on the point cloud due to its efficiency in representing 3D objects. Compared with light field holograms using a large number of images, point cloud has a relatively small size that can be efficiently processed and transmitted, and it has become the major trend of holographic representation [16].

In point cloud, 3D objects are represented by points with different locations and attributes. Color and other attributes can be added to point cloud, e.g., $\mathbf{P} \in \mathbb{R}^{N\times 7}$, where N is the number of points, and each point is represented by $[x, y, z, r, g, b, s]$ which includes its location $[x, y, z]$ in a Cartesian Coordinate System, RGB color $[r, g, b]$, and roughness s. In Fig. 3, the point cloud model of a lion is shown. In Fig. 3(a), 100% of the points are shown. In Fig. 3(b), we randomly sampled and plotted 10% of all the points. Here, we can observe that the resolution decreases. In Fig. 3(c), we use interval sampling by selecting 1 point in every 10 points. Last, in Fig. 3(d), we consider the roughness in [0,1.5] and use 100% of the points. The figures were created using the code of [17]. From Fig. 3, we see that the quality of the 3D lion is directly related to the number of points and the associated attributes. However, more points require larger storage space and communication bandwidth.

To display the point cloud objects, it is necessary to render the content for different observation angles. For example, if a user is moving, the user should observe the lion from different angles. Otherwise, if the user observes the same image, then it is a traditional 2D display. There are mainly three types of displays for HTC [18] and a comparison is given in Fig. 4.

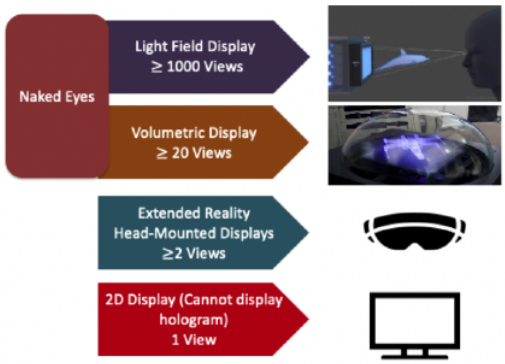


Fig. 4 – Comparison of displays. The volumetric display is Voxon VX1 and image is from [22]. The light field display image is from [23].

- **XR Head-Mounted Display**. HMDs are widely used in XR, including Augmented Reality (AR), Mixed Reality (MR), and Virtual Reality (VR), which provide two different views for the left eye and right eye to create the same effect as real human observation. Users have to use XR HMDs to observe holograms [11]. XR HMDs can only support a limited number of view angles, and users may experience severe fatigue after prolonged use. However, since the display is close to the eyes, there is a limited amount of information that needs to be transmitted to the HMD, and the required bandwidth is small compared to the light field displays.

- **Multi-view Volumetric Display**. It can support multiple view angles without using any glasses or HMDs. However, its capability in terms of the number of view angles is still limited compared to the light field display. It is a relatively small holographic display that can be used for mobile devices and holographic monitors. Usually, it can only support one user, and eye-tracking technologies can be used to adaptively render the holographic content.

- **Light Field Display.** It can provide thousands of view angles and support multiple users simultaneously [19], [20]. External glasses and HMDs are not required for light field display. A 2D array of hogels are used for light field display, and each of the hogel can generate different light intensity in different directions to recreate the light field [21]. In reality, when we observe an object, we see reflect light field from the object. The light field display generates such a light field to create 3D effects.

Light field display is anticipated to be widely used in future HTC systems since it can provide a hyperreal user experience. The required data rates of light field display depend on its size, hogel intensity, and directional resolution (D_r). The directional resolution is determined by the hogel size, e.g., for a square hogel, D_r^2 represents the number of rays that can be generated. For example, for a 1.5 m $\times$ 0.75 m display with 0.5 mm hogels and $D_r = 128$, which can be used to display human beings, the data size

	Sight	Hearing	Touch	Smell	Taste
Holographic-type communication	✓	✓	✓	✓	✓
AR, MR, VR	✓	✓	✓		
Haptic communication	✓	✓	✓		
Video	✓	✓			
Image, text	✓				
Audio		✓			

Fig. 5 – Holographic-type communication with five senses compared with other multimedia technologies.

of a single frame is

$$B = \frac{1.5 \times 0.75}{0.0005^2} \times 128^2 \times 3 \approx 221 \text{ GB}, \qquad (1)$$

where we consider the the RGB bytes per pixel is 3. More details of the calculation can be found in [21]. If the display is used for real-time HTC, consider a 30 frame-per-second refresh rate, the required data rate is 6.63 TBps. Such a high data rate is beyond the transmission capability of today's Internet, which motivates us to investigate new designs of fundamental communication and networking systems to support HTC with light field displays at the destination.

2.3 Holographic-type Communication and Mulsemedia

HTC aims to provide truly immersive experiences for users. Human perceptions use five basic senses: sight, hearing, touch, smell, and taste. Simply sending the hologram from a source to a virtual or remote destination cannot provide immersive experiences. The truly immersive experience should leverage all available human senses. For example, the user at the destination can experience the same surrounding environment as the user at the source, i.e., the user at the destination is virtually in the same space as the user at the source. Moreover, if the hologram is an animal, e.g., a lion, the user at the remote destination can see, hear, touch, and smell the lion as if the user is standing next to it. To provide such an immersive user experience, various sensors have to be employed. Besides the five basic senses, some other senses can further improve the user experiences, such as balance, wind, temperature, ambient light, etc.

As shown in Fig. 5, most multimedia technologies only use one or two senses. Recently, XR and haptic communication have leveraged touch to develop more interactive and immersive applications. Although some XR devices can also offer smell and taste senses, they are not widely adopted. Also, XR uses HMDs which is an integrated platform for various sensors and actuators. For XR, mulsemedia can be created and presented using HMDs. HTC systems without HMDs have to use external sensors and actuators.

Mulsemedia is the media that include three or more senses [6], [7]. Multimedia, including videos, audio, images, etc., uses cameras and microphones at the source to collect data. At the destination, the videos and images can be displayed on monitors and the audio files can be played by speakers. Similarly, mulsemedia uses various sensors at the source to collect data and various actuators, displays, and speakers at the destination to regenerate the environment at the source. As shown in Fig. 6, the source performs data acquisition. Sensors are placed at different locations, and it is important to ensure that the collect data is synchronized. Otherwise, the user at the destination may experience inconsistent senses. Once the destination receives mulsemedia data, it decompresses the data and processes the rendering. It should be noted that HTC systems may not be able to use all of the senses due to the lack of sensors or actuators. Users or applications can select the senses that will be used depending on the hardware and concerns of privacy and security issues.

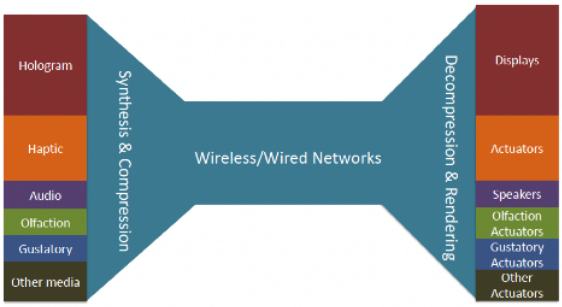


Fig. 6 – Illustration of mulsemedia systems and data transmission.

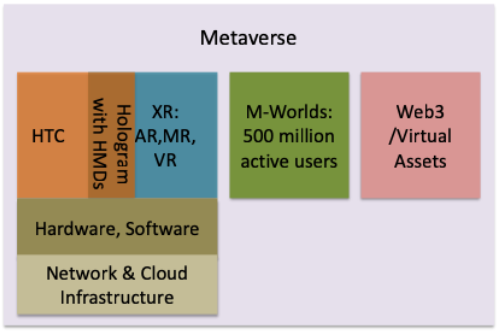


Fig. 7 – The relation among metaverse, holographic-type communication (HTC), and extended reality (XR). The intersection between XR and HTC is the hologram communication and display using XR HMDs.

HTC is a unique technology that differs from XR and Metaverse. Their relations are shown in Fig. 7.

- First, XR users can leverage HMDs to access HTC contents. But XR is not the only way nor the major way in the future to get access to HTC contents. Users without HMDs can observe and experience HTC contents with naked eyes if the light field displays are used. Therefore, HTC and XR have intersections, but each of them has its unique aspects.

- Second, the metaverse is a complex virtual world with a focus on social connections, which uses XR and HTC technologies as gateways. Users can join metaverse with HTC systems or XR systems. For exam-

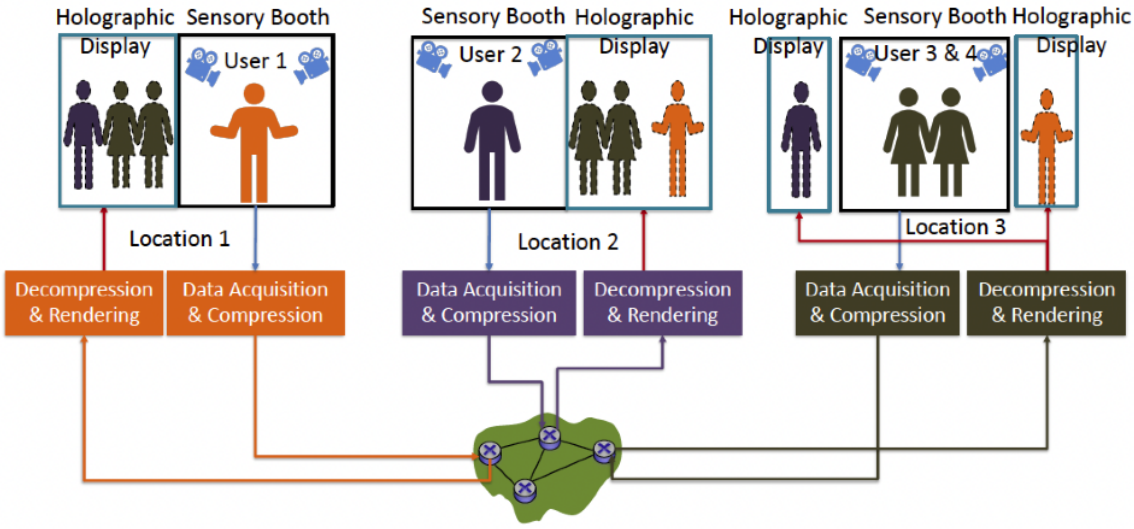


Fig. 8 – Illustration of holographic telepresence.

ple, HTC users can meet in the metaverse with occluded light field displays. Moreover, HTC users can also meet with XR users in the metaverse, which will be discussed in Section 5. Besides the HTC and XR, metaverse has another two pillars, namely, the M-Worlds with around 500 million active users and the Web3 and virtual assets which enable virtual transactions. HTC and XR are important technologies to allow users to get into metaverse.

3. USE CASES

In this section, we introduce the use cases that are enabled by HTC, which are classified into four categories, namely, holographic telepresence, holographic entertainment, holographic healthcare, and holographic remote assistance. It is impossible to enumerate all the HTC use cases. We select these use uses because they will generate profound impacts on the way we live. As the networking 2030 and 6G and beyond technologies become available, these use cases will stimulate technology innovations and create new business models.

3.1 Holographic Telepresence

Holographic telepresence will be one of the major applications of HTC. As shown in Fig. 8, the user from different locations can meet the hologram of other users. The HTC system at each location includes two major components, namely, the sensory booth and the holographic display. The sensory booth has cameras to generate point cloud of the user. Other sensors and actuators may also be installed to create and play received mulsemedia. The holographic display shows the remote users' hologram.

There is another format of holographic telepresence, where holograms are created by computers without the sensory booth in Fig. 8. Consider the Whitney Houston Hologram Tour, where the singer's hologram is created by computers rather than using cameras to record the light field. This kind of holographic telepresence can be used to display objects or characters that do not exist in the physical world.

Next, we introduce the holographic telepresence used for holographic conferences, metaverse, retail, and remote education.

3.1.1 Holographic Conference

Video conferences have played an important role during the COVID-19 pandemic. It can bring users from any location to meet with each other virtually. The format of video conferences is drastically different from real in-person conferences. Video conference users have to look at 2D screens. Compared to video conferences, HTC provides more immersive user experiences. The holographic display is large enough to display remote users with their real physical heights. Users may not be able to distinguish the difference between holographic conferences and real in-person conferences if the QoE is high enough. With smart gloves, holographic conferences can even support handshakes. An example of holographic conference is shown in Fig. 9, where holographic speakers are projected on to the screen to deliver lectures to the audience at Imperial College London.

3.1.2 Metaverse

Metaverse is an immersive virtual world generated by digital technologies [8]. HTC is an important technology to enable multi-user metaverse. Consider that if a user is surrounded by holographic displays, any virtual objects and environments can be generated and remote users can also be displayed. This kind of system can be much more expensive than XR users with HMDs due to the large size of the holographic displays. However, HTC does not require HMDs and it provides better QoE than XR.

Fig. 9 – Hologram lectures at Imperial College London. The life-size holographic speakers in the middle are projected on to a display. The image is from [24].

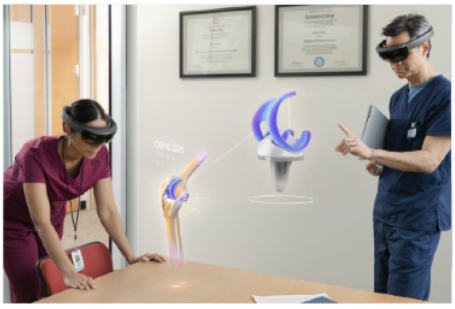


Fig. 10 – Holograms are used to train nurses [25].

3.1.3 Retail

Online shopping has significantly changed our lives, especially during a pandemic. However, it is always challenging to have an accurate idea about the size, color, smell, and many other attributes of online products by simply reading descriptions and looking at images or videos. HTC allows users to view the product with the same size, color, and 3D geometry. Some other senses such as the smell of flowers can also be delivered to HTC users. Advanced HTC systems can also allow users to get into a virtual shopping mall or market with a large number of products. Users with smart gloves can pick up the products they like and put them into a holographic cart. Such a kind of virtual shopping experience is hyperreal that the user may not be able to distinguish it from in-person shopping.

3.1.4 Education and Training

Similar to holographic conferences, students and teachers can also meet using HTC. Besides telepresence, teaching materials can also be developed into holograms to engage students. Currently, students need the imagination to understand complex concepts, especially 3D concepts such as electromagnetic field propagation and gradient of functions. With HTC, students can observe the 3D phenomena directly. As shown in Fig. 10, holograms are used to train nurses using Microsoft HoloLens [25]. High-cost equipment and devices can also be converted into holograms, and students can virtually use them to reduce the cost.

Fig. 11 – 3D reconstruction of a soccer game using XR HMDs [26].

Fig. 12 – Boxing broadcast developed by Condense Reality [27].

3.2 Holographic Entertainment

Entertainment can provide more immersive experiences using HTC than using 2D screens. Television, sports broadcasting, and gaming can be drastically changed.

3.2.1 Holographic Television

The holographic contents, such as news, advertisement, and movies, can be broadcast using HTC technologies by providers. Advanced holographic displays, such as the light field display, can be used by end users as a television. Compared to existing 2D Televisions, holographic televisions provide hyperreal 3D content.

3.2.2 Holographic Sports

Sports broadcasting using 2D screens is widely used. However, users have to follow the view of the camera, and there is no way to see other players outside of the view of the camera. Holographic sports broadcasting can provide a 3D overview from any angle. Consider that a soccer pitch can be projected on a coffee table with a flat holographic display on top. Users can see 3D players with reduced size. Instead of looking at a single or a few players, users can see the whole pitch as if they were in the stadium. Fig. 11 shows an example of using XR HMDs to display the 3D players on a table without using holographic displays. Similarly, any sports such as badminton, tennis, and boxing (as shown in Fig. 12) can be broadcast using HTC.

3.2.3 Holographic Gaming

Holographic gaming can be divided into non-immersive and immersive experiences. The non-immersive holographic gaming uses holographic displays, e.g., a flat holographic display on a table. Users can control holograms of

Fig. 13 – Axiom Holographics (formerly Euclideon Holographics) hologram arcade table [28], [29].

characters, balls, cars, and airplanes. However, for non-immersive holographic gaming, users have to leverage controllers, keyboards, and other tools as inputs. Users can play non-immersive holographic gaming with naked eyes or they can use XR glasses, specifically, Augmented Reality (AR) and Mixed Reality (MR) glasses, to observe holograms.

The immersive holographic gaming is played in a virtual world. Users can use occluded holographic displays to get into a virtual world without seeing the real surrounding environment. Users can play games as if they were the characters in the game. Note that, immersive holographic gaming can also be realized by using Virtual Reality (VR) devices. Users with VR HMDs can observe virtual holographic characters and environments through the near-eye display. The VR HMD is different from the holographic display. Only users with the HMD can observe the holographic content, while the light field display allows any user in front of the display to see the holographic content.

3.3 Holographic Healthcare

HTC can support remote healthcare. For example, for contagious diseases, the doctor and the patient can communicate using HTC without any direct contact. HTC can provide richer information than remote doctor visits using videos. Besides looking at the patient's face, the doctor can also observe the patient's behavior through a holographic display. The doctor can check the patient's body with smart gloves (deliver haptic signals). On the patient's side, sensor and actuator arrays are required to collect haptic signals. Moreover, remote surgery can be conducted with robots. The doctor can see a hologram of the patient and perform surgery on the hologram. The actions of the doctor can be replicated by a robot on the patient's side. Currently, doctors have adopted holograms to assist surgery, as shown in Fig. 14. Remote transmission of holograms will make healthcare more accessible.

3.4 Holographic Remote Assistance

HTC remote assistance can find a large number of applications. Consider that when a user needs assistance to fix appliances, cars, or any machines, remote technical support can only describe the solutions through the phone. It is hard to relate the description to the real location of the problem. Although videos provide better illustration, the size and angle of view make the interpretation challenging. HTC can display the problem location in 3D with real size. Technical support can show a demo to fix the problem. It is easy for users to follow the procedure, which significantly improves the efficiency of after-sales services.

Fig. 14 – A doctor uses the holographic display during a cardiac ablation procedure. Image from [30].

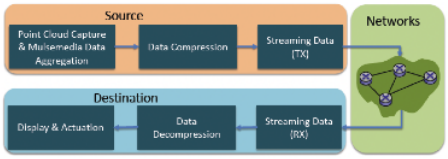


Fig. 15 – HTC system architecture with computing at the source and destination.

4. SYSTEM ARCHITECTURE AND TECHNICAL REQUIREMENTS

4.1 Generic End-to-End HTC Systems

A generic HTC system architecture is shown in Fig. 15, which consists of the source, the destination, and HTC networks. In Fig. 15, the computing tasks, including encoding, decoding, and synchronization of aggregated data are mainly located at the source and destination. The network connects the source and destination. Next, we introduce their main functionalities.

- **Source.** The source mainly has three functions. First, it uses various sensors to capture images, sound, haptic signals, and even smell and taste signals. Second, it processes the aggregated data. For example, it can generate point cloud hologram using images from different angles. Also, the source can synchronize multiple sensory data to ensure that the timestamps are correctly generated. More importantly, the source has to encode and compress the holographic data and mulsemedia to reduce network traffic. Last, the source has to follow HTC networking protocols to send data packets.

- **HTC networks.** The HTC networks deliver a large amount of source data with guaranteed and bounded end-to-end latency. Existing technologies, such as software-defined networking (SDN), automatic network slicing, content caching, etc., need to be improved to meet the requirements. New technologies, such as semantic communications, federated networks, and deterministic networking can make

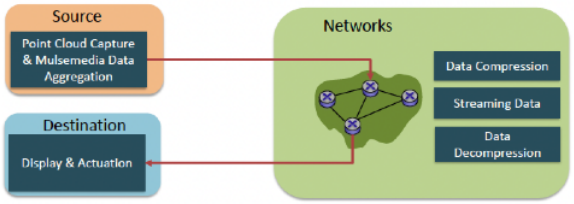


Fig. 16 – HTC system architecture with computing in the network.

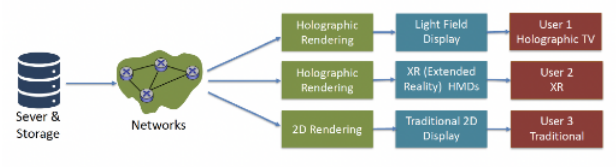


Fig. 17 – HTC broadcasting system with heterogeneous destinations.

HTC networks more robust and powerful.

- **Destination.** The destination receives and renders data for display. It also has various actuators, e.g., olfaction and gustatory, as well as speakers to recreate the environment at the source. The holographic display, actuators, and speakers are placed at different locations, but they need to be synchronized. The synchronization can be conducted at the destination or in the network, i.e., the data packets are delivered at the scheduled time.

4.2 HTC System with Computing in The Network

The system architecture in Fig. 15 require high computation capability at the source and destination. To meet the requirement, high-performance computers and edge servers can be used. To allow low-cost devices at the source and destination, the computation must be offloaded to the network using Computing in the Network (COIN) technologies [31], [32], as shown in Fig. 16. For example, sensors at the source can directly communicate with actuators, displays, and speakers at the destination. The source data can be transmitted without compression. Depending on the network status, such as congestion and available bandwidth, the compression is performed in the network adaptively. Also, the network can decompress and send rendered data to the destination. Edge servers, cloud servers, and routers with computation capabilities can be used. In this way, the computation burden at the source and destination can be dramatically reduced.

4.3 HTC Broadcasting System

Besides end-to-end communication, HTC can also support the broadcasting of prerecorded holographic content, such as holographic TV, holographic video streaming, and asynchronous holographic education. As shown in Fig. 17, holographic contents are stored in a server or data center. HTC is used to connect the server and end users. Note that, end users may use different devices including high-end light field displays and XR HMDs. Some users may use traditional 2D displays. To serve a large number of users, the holographic content has to be rendered for different end users based on QoE requirements.

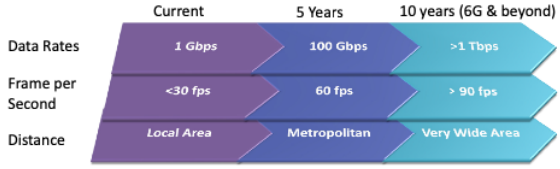


Fig. 18 – Projected future HTC development.

4.4 Key Technical Requirements of HTC

HTC is a complex system using knowledge from many technical areas, such as display, sensors, actuators, data compression, wireless communication, computer networking, etc. In this paper, we focus on the requirements and challenges that are related to HTC networking, which mainly consists of the following four aspects.

- **Data rates.** As discussed in Section 2, the holographic display requires data rates as high as several Terabytes per second. Although efficient encoding and distributed COIN can reduce the required data rates to Gigabytes per second or even several hundreds of Megabytes per second, this increases latency due to computation. Thus, HTC requires unprecedented high data rates, and today's networking technologies need to be reexamined and upgraded to support it. The projected HTC data rates, frame-per-second, and communication distance are shown in Fig. 18. It is anticipated that the 6G and beyond will support high-quality HTC across very wide areas.

- **End-to-end latency.** The latency in HTC networks consists of data acquisition delay, encoding delay, communication, and networking delay, decoding delay, and display and actuation delay. For high-interactive applications, such as holographic gaming, the overall end-to-end latency has to be lower than 20 ms to avoid sickness. Data acquisition, encoding, decoding, and display may use up the 20 ms. As a result, the latency budget for communication and networking can be lower than 1 ms.

- **Lightweight computation.** HTC requires a significant amount of computations to optimize the transmission. Lightweight computation algorithms and frameworks need to be designed to support portable devices that users can access ubiquitously. For example, a user can project holographic contents onto a light field display using smartphones or tablets. Moreover, the lightweight computation can reduce the end-to-end latency and provide high QoE.

- **Mulsemedia synchronization.** HTC systems have a large number of sensors, actuators, displays, and

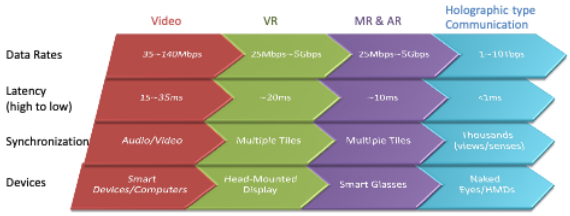


Fig. 19 – Comparison of technical requirements among videos, VR, AR and MR, and HTC.

speakers. Various types of data have to be synchronized and presented at the destination to avoid any mismatch.

A comparison of technical requirements among videos, VR, AR and MR, and HTC is given in Fig. 19. The approximated data was adopted from [33] and [11].

5. RESEARCH DIRECTIONS AND POTENTIAL SOLUTIONS

In this section, we discuss the research directions and potential solutions in 6G and beyond to realize high-QoE HTC systems. We focus on the standards and techniques at the sources, networks, and destinations. Although we discuss them separately, the sources, networks, and destinations are cohesively connected and mutually affect each other. For example, a strong data compression rate at the source generates less traffic for the network. Also, edge intelligence [34] at the source, network, and destination is an important technology that can optimize the HTC system to meet the technical requirements. At the end of this section, we discuss the existing HTC testbeds, as well as available software and hardware for prototype development.

5.1 Source

As discussed in Section 2, the source mainly uses point cloud videos. Next, we discuss the point cloud encoding standards and mulsemedia synchronization.

5.1.1 Representation and Encoding

Multiple RGB-D (depth) cameras are organized as an array to collect images of objects from different angles. Based on the collected information, 3D point cloud objects can be created using computers. MPEG V-PCC (Video-based Point Cloud Compression) and MPEG G-PCC (Geometry-based Point Cloud Compression) are the two major point cloud compression frameworks [35], [36]. V-PCC encodes the 3D point cloud data as a series of 2D videos so that the success of existing 2D video encoding technologies can be fully leveraged. G-PCC encodes the 3D point cloud data directly in the 3D space using geometric-driven approaches. It has been used for LiDAR generated point cloud, such as 3D maps.

As shown in [37], the point cloud encoding can take as long as several seconds (or even longer). Although the compressed data size is small, the latency is not tolerable. This can be improved by optimizing the encoding framework by reducing the redundancy between frames. The information of eye-tracking of the user at the destination is useful to efficiently encode the point cloud at the source, e.g., more data can be allocated to the main field-of-view of the user at the destination.

5.1.2 Mulsemedia Synchronization

Mulsemedia includes different types of sensors at different locations, and its synchronization is important [38]. For example, when a user is talking about flowers with certain gestures, the hologram of the user, olfaction signals of the flower, and audio signals need to be synchronized, otherwise, the destination user can be confused. The synchronization is challenging because of the following two reasons.

- First, different types of sensors have different response times. For example, cameras can have a capture rate of 50 frames/image per second which results in a capture latency of 20 ms, while microphones can have a latency of several milliseconds which is much lower than cameras'. For highly dynamic haptic sensors, the latency should be smaller than 1 ms [39]. Therefore, various sensors need to be calibrated and synchronized by considering complex hardware heterogeneity. Machine learning algorithms can be used for synchronization. Usually, certain behavior or activity includes multiple correlated senses. Classifiers and machine learning predictors can be used to verify whether collected senses are consistent and automatically generate senses that are missing or delayed.
- Second, human perception has different levels of latency tolerance for different senses. As the study in [38] shows, for acceptable user experience the latency of haptic and air-flow media can be as high as 1 s. However, this depends on the applications and the distance between the user and actuators. For time-sensitive applications, such as remote surgery, the latency of haptic signals must be much lower than 1 s.

Currently, the study of HTC-related mulsemedia is sparse. Future research includes the following three directions.

- First, design HTC mulsemedia sensors and actuators. New sensors and actuators are required to provide immersive hyperreal user experiences. For example, the olfaction can be generated by actuators on a desk or bio-stimulators that can directly control human beings' senses by stimulating brain signals. Also, novel haptic and gustatory sensors and actuators are desirable. Currently, these are active research areas.
- Second, study mulsemedia end-to-end latency requirements. Currently, there is no clear require-

ment for end-to-end latency for different applications. Generally, it is always desirable to have low end-to-end latency. However, it is costly to obtain low end-to-end latency for every user. Thus, it is necessary to understand the end-to-end latency requirements of various applications.

- Third, develop intelligent mulsemedia synchronization frameworks. With calibrated HTC sensors and the knowledge of end-to-end latency requirements, intelligent mulsemedia synchronization frameworks that can synchronize mulsemedia at the source, in the network, or at the destination can be designed.

Note that, MPEG-V provides a framework to create mulsemedia for interactions between the physical world and a virtual world [40]. The discussed mulsemedia research directions can be built on top of existing MPEG-V standards.

5.2 Networking

HTC networking technologies need to support high data rates while maintaining a guaranteed and bounded end-to-end latency. However, the data rates and end-to-end latency are not independent of each other due to the constraints of networking protocols. Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) are widely used transport layer technologies. The TCP throughput can be obtained by [41]

$$T \leq \frac{MSS}{RTT}\frac{C}{\sqrt{\rho}}, \tag{2}$$

where MSS is the maximum segment size, RTT is round trip time, and ρ is the packet loss rate. The constant C is implicitly assumed to be 1 for simplification. It can be different values around 1 depending on different assumptions [41]. When $MSS = 1460$ Bytes, the throughput in log scale is shown in Fig. 20. As we can see, to achieve 369 Gbps throughput, the required round trip time is 1 ms and the packet loss rate is $1.0 \times 10^{-7}\%$. Such low round trip time and high reliability cannot be supported by existing networks. Also, note that there is a physical limitation of round trip time due to the speed of signal propagation. For example, for 1 ms, the signal can only propagate around 300 km at the speed of light.

High-quality HTC requires data rates in the range of several Gbps to Tbps. Also, considering a large number of users, the required network throughput is unprecedented. Therefore, novel approaches are desirable to reduce the packet loss and round trip time to increase the network throughput. Next, we first introduce the basic networking streaming protocols, upon which we point out the enablers for HTC networks, including predictive networks, semantic communication networks, deterministic networks, and federated networks.

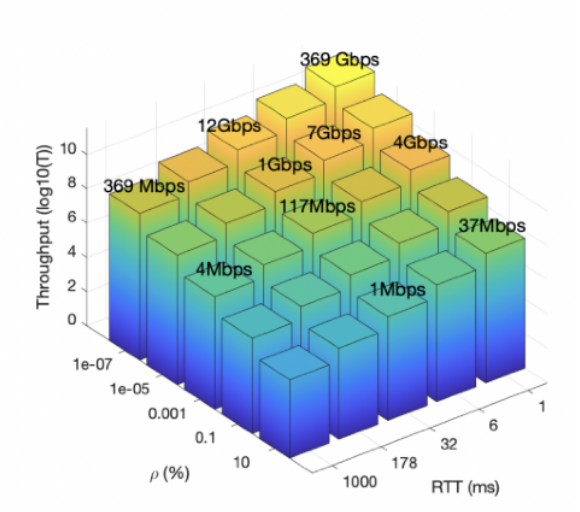


Fig. 20 – TCP throughput under different packet loss and round trip time. The throughput (T) is shown in logscale $log_{10}(T)$.

5.2.1 *Networking Streaming Protocols*

Currently, TCP and UDP are the fundamental streaming protocols. TCP can provide reliable connections between the source and destination, i.e., the packet loss rate is low. However, the end-to-end latency of TCP is higher than that of UDP. UDP does not build a reliable connection using handshaking between the source and destination. As a result, the packet loss rate is higher, and loss recovery is needed at the destination. The MPEG-DASH (Dynamic Adaptive Streaming over HTTP) is a promising standard for HTC networking. It encodes the stream into different segments with different qualities. For point cloud videos, the segmentation can be performed in space and time. The destination can decide the quality based on the estimation of network status. The protocol is based on TCP, and it can leverage existing HTTP web infrastructure to deliver content. UDP-based protocols, such as WebRTC (Web Real Time Communications), need more investigation and redesign to transmit HTC data. Since HTC is an emerging technology, currently, there are no widely used steaming protocols. TCP-based solutions need dedicated networking resources and network traffic information to reduce the end-to-end latency, while UDP-based solutions need loss recovery solutions. Next, we introduce the potential solutions that can improve TCP and UDP protocols for HTC networks.

5.2.2 *AI-empowered Network Prediction and Adaptive Control*

AI can be used for Quality-of-Service (QoS) prediction, network planning, and network control in SDN. Most existing network protocols respond to abnormal events after they have generated negative impacts. It is more efficient if the network can predict these events and respond early. The Cisco Predictive Networks is an example of this technology. Also, based on partial observation of the net-

work status, the QoS parameters can be estimated, such as latency and bandwidth, to determine if the user's requests can be served [42], [43]. AI can also be used to solve complex networking problems, such as network planning. Network planning is an important part to support HTC networks. Based on user service demand, it can continuously update the network topology, schedule maintenance, and upgrade network hardware and software. However, due to the large scale of the network, especially wide area networks, the optimal solution is challenging to obtain. Deep learning solutions, such as deep reinforcement learning and graph neural networks, are efficient solutions that can obtain optimal solutions and enable self-driving networking [44]. Last, AI can be used in real-time network control, such as network monitoring, adaptive routing, and network slicing [45]–[48]. AI-empowered networks can efficiently address the challenges that are faced by TCP and UDP by predicting and controlling network traffics, allocating sufficient bandwidth, and preventing packet losses. This will provide reliable and efficient networks for HTC.

5.2.3 *Semantic Communication Networks*

In [49], communication problems were divided into three levels by Weaver:

- The technical problem studies how accurately we can transmit communication symbols, which can be addressed by using Shannon's communication models.
- The semantic problem studies how precisely the transmitted symbols can convey the desired meaning.
- The effectiveness problem studies how effectively the received meaning can affect the desired conduct.

The technical problem has been extensively studied. The well-designed transmitter and receiver have various technologies to efficiently compress data, mitigate multipath fading, and optimally detect received signals. The semantic problem is only recently received attention due to the advancement of machine learning technologies [50]–[52]. The hologram and mulsemedia require significantly large bandwidth to transmit data. Although AI-empowered technologies can improve networking performances, it is also desirable to adopt efficient source coding and channel coding schemes to reduce the amount of data that is transmitted. Existing research works have studied the joint source-channel coding of transmitting text and images [50], [53], [54]. Instead of focusing on the technical problem of transmitting symbols correctly, semantic communication focuses on correctly delivering the meaning. Deep learning plays an important role in semantic communication by understanding the meaning of transmitted information, which can efficiently compress the source data. This significantly reduces the transmitted data and network traffic.

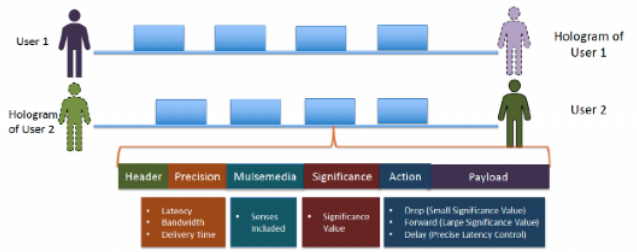


Fig. 21 – New IP-type packet for HTC networks using semantic communication and packet wash.

Semantic communication for 3D videos, especially point cloud videos, is an open research problem. If we can deliver the same meaning without transmitting the whole hologram, the required data rates will be much lower. A potential solution is to use deep learning to understand the meaning at the source and detect the major differences between adjacent frames. At the destination, the meaning is rendered using deep learning together with the hologram of the source user, which is only transmitted once, to create the expressions of feeling and body motions. Redundant information can be removed when the network does not have sufficient bandwidth. However, when the network throughput is high, redundant information can be kept to increase the QoE. Therefore, adaptive semantic communication needs to be designed to fully leverage the network resources to provide the optimal QoE. In [55], [56], the qualitative communication is studied. It transmits packets with different significance and selectively drops packets based on their significance when there are congestions. The significance value can be derived from the meaning of semantic communication. For example, if a packet carries most of the meaning at the source, its significance value is high. On the contrary, if a packet does not carry information about the meaning at the source, but includes a certain amount of background information, its significance value is small. At the destination, dropped packets with small significance values can be recovered using deep learning algorithms.

The significance value can be included in New IP packets [56], [57]. An example of a New IP-type packet is shown in Fig. 21. The packet gives clear information for network routers about how and when to deliver the packet to the destination. Packets can be dropped based on their significance value using packet wash [58] when there are network congestions. The New IP offers a way to implement semantic communication networks. The powerful New IP packets provide reconfigurability and flexibility to deliver large data packets with guaranteed latency requirements.

5.2.4 *Deterministic Networks*

Existing computer networks mainly deliver packets using best-effort delivery. The end-to-end latency is a random variable and jitters can be significant. Existing multimedia communication typically can buffer 100 ms data for decoding and rendering to provide smooth user ex-

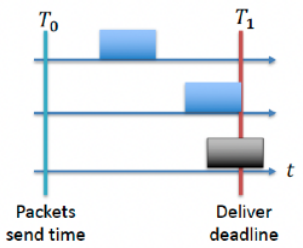


Fig. 22 – In-time guarantee. The packet delivery deadline is T_1. The top and the middle packets are delivered before the deadline, which meet the requirements. The bottom packet is fully delivered after the deadline.

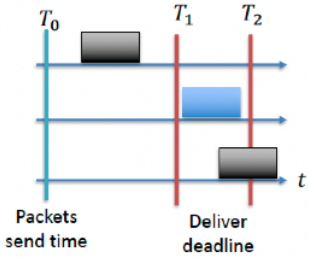


Fig. 23 – On-time guarantee. The packet delivery deadline is between T_1 and T_2. The top packet is delivered early and the bottom packet is delivered late. The packet in the middle is delivered on time, which meets the requirement.

periences. However, 100 ms of HTC streaming generates a large amount of data; it is challenging or impossible to buffer a long sequence of received packets. Thus, the packets need to be delivered to the destination with a guaranteed time, so that they can be processed immediately without occupying the buffer. This requires a deterministic network where packets can be delivered as scheduled. In [57], [59], [60], three different time-related services are proposed for Network 2030:

- **In-time guarantees.** Data packets arrive before a specific time, as shown in Fig. 22.
- **On-time guarantees.** Data packets arrive at a specific time with guaranteed variance, as shown in Fig. 23.
- **Coordinated guarantees.** Multiple flows of data packets arrive with in-time or on-time guarantees, as shown in Fig. 24.

The on-time guarantee is more challenging than the in-time guarantee since the packets cannot be delivered early. Existing networks cannot provide such kinds of services. The SDN, network function virtualization, and automatic network slicing are key enablers to realizing these services. For HTC networks, the coordinated guarantee is of paramount importance. The mulsemedia may include multiple flows of different senses. They need to be delivered on time with a coordinated guarantee. If any flow is delivered early, it needs to be buffered. Holographic data and haptic data sizes are large, which consume significant storage resources. On the contrary, if all the flows are delivered on time, the destination can synthesize and render the received mulsemedia immediately.

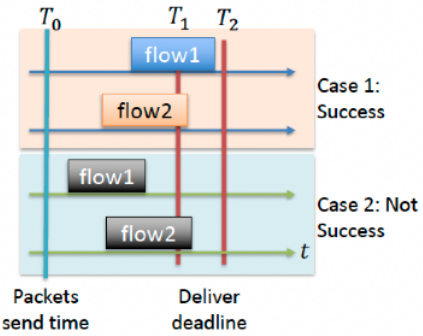


Fig. 24 – Coordinated guarantees. The upper two flows of packets are delivered with on-time guarantee, while the lower two flows are not.

The Deterministic Networking (DetNet) Architecture is proposed by the IETF DetNet working group to provide bounded network latency and reduce packet loss. It can reduce the unexpected randomness in networks. The Time-Sensitive Network (TSN) following IEEE 802.1 is [61] a similar technology that aims to provide zero congestion loss and bounded latency by reserving network resources. The development and implementation of DetNet and TSN can facilitate the design of deterministic networks for HTC.

5.2.5 *Federated Networks*

As discussed in previous sections, HTC needs to reserve network resources, such as bandwidth and computing resources, in order to obtain high data rates, guaranteed and bounded latency, and high reliability. Reserving resources in public networks is costly due to the large number of users. It is anticipated that high-quality HTC can be first implemented in private networks with fewer users and more networking resources, e.g., campus networks, industrial networks, etc. It is relatively simple to connect HTC users within a private network, but inter-network HTC requires more effort to define a universal policy across different private networks. Federated networks enforce consistent configurations and policies and manage shared networking resources among separated private networks. If an HTC user in a private network communicates with another user in a different network, federated networking configurations and policies can be used to allow the two users to be connected. Also, networking resources, such as network services and gateways, can be shared and dynamically allocated among private networks to support HTC applications.

5.2.6 *HTC Networking Protocol Stacks*

The Open Systems Interconnection model (OSI model) has been extensively used for data communication. The networking protocol stack is divided into seven layers with different functions. The OSI model is developed for communication using data packets, which has enjoyed

great success during the past several decades. HTC calls for a new design of the network protocols that focuses on content/meaning rather than data because HTC has a significant amount of data that is not efficient to transmit. First, the updated application layer needs to evaluate the available hardware, such as sensors, actuators, and displays, then determine the QoE. Second, the transport layer can be augmented or replaced by a layer that can support data transportation, hologram transportation, and mulsemedia transportation. Both data bits and meanings can be supported. Third, instead of forwarding packets, the network layer needs to include AI/ML and the New IP [56] with deterministic network performance. Last, the physical layer at the end users' side should include mmWave and Terahertz wireless sensing as built-in functions to monitor users' motion and mapping surrounding environments.

5.3 Destinations

The destination renders HTC content and provides feedback to the source and HTC networks to improve and maintain high-quality QoE. Specifically, the functionalities of the destination include:

- First, the destination monitors the network status and defines the desired QoE. Such information can be sent back to the source to optimize the encoding. For example, when the network is free of congestion and sufficient network bandwidth is available, the destination can request high-quality HTC content. On the contrary, when the network bandwidth is not sufficient, the destination can request the semantic meaning without other detailed data. This task includes the QoE-aware design and AI-empowered motion prediction at the destination.

- Second, the destination detects and corrects errors, improves the resolution and quality of holograms and mulsemedia, and synchronizes received packets from multiple sources.

Next, we discuss the detailed challenges at the destination.

5.3.1 QoE-aware Design

Although HTC systems need to transmit a significant amount of data, the user at the destination may only pay attention to a part of the received data [16]. For instance, in a telepresence conference, a user may only focus on the hologram's face. All other parts of the body may use a reduced resolution. Also, the destination needs to communicate with the source to request the desired QoE. The source can use adaptive encoding to stream data. If the received data cannot provide the requested QoE, the destination can use AI-empowered fine-tuning and holographic enhancement technology to improve the QoE. Although similar technologies have been developed for image and video transmissions [62], it is still an open research problem in HTC networks.

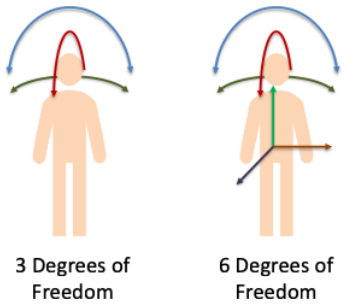


Fig. 25 – Illustration of the 3 degrees of freedom and 6 degrees of freedom.

5.3.2 AI-empowered Wireless Sensing and Motion Prediction

Motion prediction of the user at the destination is necessary to allow the source to adaptively transmit data and reduce the end-to-end latency. For example, when the user is looking in different directions, the source with predicted knowledge can transmit high-resolution information to the destination user's FoV. Motion prediction uses various sensors, such as cameras and inertial measurement units, to track the movement of the head, hands, and body.

For XR users, sensors are integrated with HMDs. While they are using HMDs to get access to holographic content, their motion can be tracked by these sensors. However, for light field display users, they can see holographic content with naked eyes. As a result, it is challenging to find optimal locations for sensors. On one hand, if sensors are placed far away from the user, the sensing accuracy cannot be guaranteed. On the other hand, if sensors are wearable for the user, the QoE will be reduced since the user has to use external devices.

Generally, the motion prediction for HTC encounters three major challenges. First, how to optimally perform sensing for light field display users. Some XR devices only provide 3 Degrees of Freedom (DoF) to track the motion of the head. HTC requires 6 DoF, including the head and the body, as shown in Fig. 25. This provides more immersive experiences to fully utilize the high-quality hologram. As a result, the motion sensing is more challenging for HTC. Wireless sensing is a promising solution. In 5G and 6G, mmWave and Terahertz radios are widely used. Due to the short wavelength and thus high resolution, the sensing accuracy is high [63]. Wireless access points can be leveraged to track users' motion due to its ubiquitous availability [64]. Second, how to predict users' motion based on collected sensing data. Deep learning-based architecture has been extensively used for eye tracking [65] and other motion sensing and prediction, which has demonstrated high accuracy. Third, how to mitigate the impact of prediction errors. If the prediction is not accurate, users may experience sickness [66]. To address

this issue, it is necessary to develop a prediction error detection framework. Once the error is detected and evaluated, the destination may use AI-empowered solutions to create computer-rendered content based on previous content.

5.3.3 AI-empowered Error Detection and Correction

By using the New IP and semantic communication, some of the packets may be dropped or distorted during transmission. MPEG V-PCC converts the point cloud data into different parts, such as the occupancy video data, geometry video data, and attribute video data. Different parts play different roles in reconstructing the original point cloud [67]. As shown in [67], packet losses can be addressed well using simple error correction techniques, such as copying from nearby points. Besides packet loss, wireless communication channels, e.g., from wireless access points to holographic displays or HMDs, experience fading and noises. In [68], [69], the point cloud is considered as a graph. Graph signal processing and graph neural networks are used to develop a robust communication system in presence of channel fading and noises. Based on these works, more advanced AI-empowered error detection and correction schemes can be developed. Deep learning architectures, such as Generative Adversarial Networks, can be used to correct errors and improve the QoE.

5.3.4 Multi-user Synchronization

The scalability of HTC networks is a challenging problem. Considering the telepresence conferencing, when there are two users, the included sensors, actuators, displays, and speakers may be manageable. However, as the user number increases, the included HTC components increase dramatically. Users may even come from widely spread locations and the communication latency are drastically different. Therefore, multi-user HTC applications require more handshakes. For example, when a user requests to join a multi-user HTC application, the network needs to evaluate the user's networking and HTC hardware and software. If the bandwidth, latency, and other application-specified requirements can be met, then the user is allowed to join. Otherwise, the user is rejected. Once multiple users join the HTC application, the synchronization can be achieved by using deterministic networking functions, such as bounded latency. Each user or the application can define the packet delivery time to ensure that data packets will be delivered in a synchronized way to avoid congestion.

5.4 Edge Intelligence

Edge intelligence [34] will play an important role in HTC networks due to the limited network throughput and end-to-end latency constraints. The cloud-based solution aggregates data from various sensors and performs point cloud compression. This creates a long delay due to the long distance between the user and the cloud. Also, the cloud may serve many users simultaneously, which also incurs latency due to queuing. Edge intelligence provides source sensors with learning and computing capabilities. Source sensors can intelligently compress raw data, e.g., extract and send features instead of raw data. This can reduce network traffic. The network edge intelligence can serve as a gateway. First, it can accomplish the offloaded tasks from source sensors. Second, it can intelligently offload high-complexity tasks to the cloud in order to obtain advanced computation resources to accomplish the tasks on time. In addition, network edge intelligence can perform distributed computation in the network for encoding and decoding HTC data. Edge intelligence can also improve the performance of the sensors, actuators, displays, and speakers at the destination. Different senses may require different intelligent algorithms for decoding and synchronization, the specialized edge intelligence for different devices can improve their performance in error detection and correction, sensing, and prediction.

5.5 Testbed Design

Existing HTC testbeds mainly transmit holographic data without mulsemedia [70]. Currently, Mulsemedia testbed design is an independent research direction [40]. It will be an important step towards truly immersive HTC by integrating the holographic data transmission, e.g., point cloud video streaming, with mulsemedia to build a comprehensive HTC testbed.

Most existing HTC testbed uses XR HMDs or even 2D computer monitors as the display to evaluate the quality of the received hologram [12], [13], [70]. Point cloud data can be generated from multiple RGB-D (Depth) cameras, e.g., Microsoft RGB-D Kinect 2.0. The LiveScan3D toolkit is an efficient tool to generate point cloud data based on synchronized multiple images captured by RGB-D cameras. To save network bandwidth, only the human body is captured and transmitted, and the background information is usually neglected. Also, some existing testbeds have a limited communication range, and the experiments are performed in labs. In [13], long-range HTC was tested. However, the performance degrades significantly when the source and destination are in different countries. HTC using light field display has been reported in [10], [19]. Although they can successfully display holograms streamed from remote locations, the quality still needs to be improved. Also, compared to the XR HMDs, the light field display is more expensive and development toolkits are spare. For future research, XR HMDs are easier to use. For example, Microsoft HoloLens is supported by Microsoft Azure services and other XR development software, such as Unity.

6. CONCLUSION

Holographic-type communication (HTC) can provide truly immersive user experiences by fully using the five basic senses of human perception. It is an emerging technology that can enable novel applications, such as telepresence conferences and remote surgery. This paper provides fundamental knowledge of HTC and outlines the research roadmaps. First, this paper provides an introduction to HTC, including the difference between HTC and existing multimedia communication, HTC system architectures, and promising HTC use cases. Second, this paper points out HTC research challenges from the perspectives of the source, HTC networks, and the destination. Promising solutions that will be developed in 6G and beyond and Network 2030 to realize high-quality HTC are identified and introduced.

REFERENCES

[1] "An Evening With Whitney," 2022. [Online]. Available: https://baseholo gram.com/productions/whitney-houston.

[2] "ABBA Voyage Concert," 2022. [Online]. Available: https://abbavoyage.com.

[3] D. Gabor, "Holography, 1948-1971," *Science*, vol. 177, no. 4046, pp. 299–313, 1972.

[4] A. Murad and W. Smale, "How hologram tech may soon replace video calls," 2021. [Online]. Available: https://www.bbc.com/news/business-59577341.

[5] "Proto," 2022. [Online]. Available: https://www.protohologram.com.

[6] T. Bi, A. Pichon, L. Zou, S. Chen, G. Ghinea, and G.-M. Muntean, "A dash-based mulsemedia adaptive delivery solution," in *Proceedings of the 10th International Workshop on Immersive Mixed and Virtual Environment Systems*, 2018, pp. 1–6.

[7] G. Ghinea and O. Ademoye, "A user perspective of olfaction-enhanced mulsemedia," in *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, 2010, pp. 277–280.

[8] L.-H. Lee, T. Braud, P. Zhou, L. Wang, D. Xu, Z. Lin, A. Kumar, C. Bermejo, and P. Hui, "All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda," *arXiv preprint arXiv:2110.05352*, 2021.

[9] K. Kim, J. Bolton, A. Girouard, J. Cooperstock, and R. Vertegaal, "Telehuman: Effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 2531–2540.

[10] D. Gotsch, X. Zhang, T. Merritt, and R. Vertegaal, "Telehuman2: A cylindrical light field teleconferencing system for life-size 3d human telepresence.," in *CHI*, vol. 18, 2018, p. 552.

[11] I. F. Akyildiz and H. Guo, "Wireless communication research challenges for Extended Reality (XR)," *ITU Journal on Future and Evolving Technologies*, vol. 3, no. 1, 2022.

[12] M. Kowalski, J. Naruniec, and M. Daniluk, "Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors," in *2015 international conference on 3D vision*, IEEE, 2015, pp. 318–325.

[13] I. Selinis, N. Wang, B. Da, D. Yu, and R. Tafazolli, "On the internet-scale streaming of holographic-type content with assured user quality of experiences," in *2020 IFIP networking conference (networking)*, IEEE, 2020, pp. 136–144.

[14] I. FG-NET2030, "Representative use cases and key network requirements for network 2030," *FG-NET2030 document NET2030-O-027*, 2020.

[15] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di Renzo, and M. Debbah, "Holographic mimo surfaces for 6g wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 118–125, 2020.

[16] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters, and F. De Turck, "Toward truly immersive holographic-type communication: Challenges and solutions," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 93–99, 2020.

[17] P. Glira, N. Pfeifer, C. Briese, and C. Ressl, "A correspondence framework for als strip adjustments based on variants of the icp algorithm," *Photogrammetrie-Fernerkundung-Geoinformation*, vol. 2015, no. 4, pp. 275–289, 2015.

[18] D. Blinder, A. Ahar, S. Bettens, T. Birnbaum, A. Symeonidou, H. Ottevaere, C. Schretter, and P. Schelkens, "Signal processing challenges for digital holographic video display systems," *Signal Processing: Image Communication*, vol. 70, pp. 114–130, 2019.

[19] P. A. Kara, A. Cserkaszky, M. G. Martini, A. Barsi, L. Bokor, and T. Balogh, "Evaluation of the concept of dynamic adaptive streaming of light field video," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 407–421, 2018.

[20] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, 2017.

[21] T. L. Burnett, "Invited paper: Light-field display architecture and the challenge of synthetic light-field radiance image rendering," in *SID Symposium Digest of Technical Papers*, Wiley Online Library, vol. 48, 2017, pp. 899–902.

[22] "Voxon VX1 Volumetric Display is a Real-Life Hologram Table That Doesn't Require Special Glasses," 2019. [Online]. Available: https://www.techeblog.com/voxon-vx1-hologram-table/.

[23] A. Pennington, "Towards the holodeck: An in-depth look at light field," 2018. [Online]. Available: https://www.ibc.org/trends/towards-the-holodeck-an-in-depth-look-at-light-field/2809.article.

[24] A. Hollender, "ARHT Introduces Real-Time Holographic Lecturers At London Business School," 2018. [Online]. Available: https://vrscout.com/news/arht-holographic-lecturers/.

[25] "Using holograms to train nurses: Pearson and Microsoft launch mixed-reality curriculum," 2018. [Online]. Available: https://news.microsoft.com/en-gb/2018/01/22/using-holograms-train-nurses-pearson-microsoft-launch-mixed-reality-curriculum/.

[26] K. Rematas, I. Kemelmacher-Shlizerman, B. Curless, and S. Seitz, "Soccer on your tabletop," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4738–4747.

[27] "Condense Reality and BT collaborating on 3D hologram technology," 2020. [Online]. Available: https://www.sportspromedia.com/news/condense-reality-bt-hologram-technology-streaming-volumetric-video-boxing/.

[28] L. Blain, "Interview: Euclideon prepares to storm the arcade world with 3D hologram games," 2018. [Online]. Available: https://newatlas.com/euclideon-hologram-arcade-games/57334/.

[29] "Axiom Holographics," [Online]. Available: https://axiomholographics.com.

[30] B. Miller, "WashU-developed holograms help physicians during cardiac procedure," 2020. [Online]. Available: https://engineering.wustl.edu/news/2020/WashU-developed-holograms-help-physicians-during-cardiac-procedure.html.

[31] A. Sapio, I. Abdelaziz, A. Aldilaijan, M. Canini, and P. Kalnis, "In-network computation is a dumb idea whose time has come," in *Proceedings of the 16th ACM Workshop on Hot Topics in Networks*, 2017, pp. 150–156.

[32] D. Zeng, N. Ansari, M.-J. Montpetit, E. M. Schooler, and D. Tarchi, "Guest editorial: In-network computing: Emerging trends for the edge-cloud continuum," *IEEE Network*, vol. 35, no. 5, pp. 12–13, 2021. DOI: 10.1109/MNET.2021.9606835.

[33] R. Li, "Enabling holographic media for future applications: Identifying the missing pieces and limitations in networks," in *Sigcomm NEAT 2019 Panel*, ACM, 2019.

[34] D. Xu, T. Li, Y. Li, X. Su, S. Tarkoma, T. Jiang, J. Crowcroft, and P. Hui, "Edge intelligence: Empowering intelligence to the edge of network," *Proceedings of the IEEE*, vol. 109, no. 11, pp. 1778–1837, 2021.

[35] C. Cao, M. Preda, and T. Zaharia, "3d point cloud compression: A survey," in *The 24th International Conference on 3D Web Technology*, 2019, pp. 1–9.

[36] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (v-pcc) and geometry-based (g-pcc)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, 2020.

[37] H. Liu, H. Yuan, Q. Liu, J. Hou, and J. Liu, "A comprehensive study and comparison of core technologies for mpeg 3-d point cloud compression," *IEEE Transactions on Broadcasting*, vol. 66, no. 3, pp. 701–717, 2019.

[38] Z. Yuan, T. Bi, G.-M. Muntean, and G. Ghinea, "Perceived synchronization of mulsemedia services," *IEEE Transactions on Multimedia*, vol. 17, no. 7, pp. 957–966, 2015.

[39] K. Antonakoglou, X. Xu, E. Steinbach, T. Mahmoodi, and M. Dohler, "Toward haptic communications over the 5g tactile internet," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3034–3059, 2018.

[40] E. B. Saleme and C. A. S. Santos, "Playsem: A platform for rendering mulsemedia compatible with mpeg-v," in *Proceedings of the 21st Brazilian Symposium on Multimedia and the Web*, 2015, pp. 145–148.

[41] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, "The macroscopic behavior of the tcp congestion avoidance algorithm," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 3, pp. 67–82, 1997.

[42] K. Rusek, J. Suárez-Varela, P. Almasan, P. Barlet-Ros, and A. Cabellos-Aparicio, "Routenet: Leveraging graph neural networks for network modeling and optimization in sdn," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 10, pp. 2260–2270, 2020.

[43] J. Suárez-Varela, M. Ferriol-Galmés, A. López, P. Almasan, G. Bernárdez, D. Pujol-Perich, K. Rusek, L. Bonniot, C. Neumann, F. Schnitzler, *et al.*, "The graph neural networking challenge: A worldwide competition for education in ai/ml for networks," *ACM SIGCOMM Computer Communication Review*, vol. 51, no. 3, pp. 9–16, 2021.

[44] H. Zhu, V. Gupta, S. S. Ahuja, Y. Tian, Y. Zhang, and X. Jin, "Network planning with deep reinforcement learning," in *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, 2021, pp. 258–271.

[45] F. Tang, B. Mao, Z. M. Fadlullah, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "On removing routing protocol from future wireless networks: A real-time deep learning approach for intelligent traffic control," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 154–160, 2017.

[46] P. Pinyoanuntapong, M. Lee, and P. Wang, "Distributed multi-hop traffic engineering via stochastic policy gradient reinforcement learning," in *2019 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2019, pp. 1–6.

[47] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, IEEE, 2017, pp. 1–6.

[48] A. Thantharate, R. Paropkari, V. Walunj, and C. Beard, "Deepslice: A deep learning approach towards an efficient and reliable network slicing in 5g networks," in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, IEEE, 2019, pp. 0762–0767.

[49] W. Weaver, "Recent contributions to the mathematical theory of communication," *ETC: a review of general semantics*, pp. 261–281, 1953.

[50] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.

[51] Z. Q. Liew, Y. Cheng, W. Y. B. Lim, D. Niyato, C. Miao, and S. Sun, "Economics of semantic communication system in wireless powered internet of things," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 8637–8641.

[52] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44–50, 2021.

[53] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, 2019.

[54] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 2018, pp. 2326–2330.

[55] R. Li, K. Makhijani, H. Yousefi, C. Westphal, L. Dong, T. Wauters, and F. De Turck, "A framework for qualitative communications using big packet protocol," in *Proceedings of the ACM SIGCOMM 2019 Workshop on Networking for Emerging Applications and Technologies*, 2019, pp. 22–28.

[56] R. Li, L. Dong, C. Westphal, and K. Makhijani, "Qualitative communication for emerging network applications with new ip," in *2021 17th International Conference on Mobility, Sensing and Networking (MSN)*, IEEE, 2021, pp. 628–637.

[57] R. Li, K. Makhijani, and L. Dong, "New ip: A data packet framework to evolve the internet," in *2020 IEEE 21st International Conference on High Performance Switching and Routing (HPSR)*, IEEE, 2020, pp. 1–8.

[58] L. Dong and A. Clemm, "High-precision end-to-end latency guarantees using packet wash," in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, IEEE, 2021, pp. 259–267.

[59] I. FG-NET2030, "Network 2030 - a blueprint of technology, applications and market drivers towards the year 2030 and beyond," *FG-NET2030 document*, 2019.

[60] ——, "New services and capabilities for network 2030: Description, technical gap and performance target analysis," *FG-NET2030 document NET2030-O-027*, 2019.

[61] N. Finn, "Introduction to time-sensitive networking," *IEEE Communications Standards Magazine*, vol. 2, no. 2, pp. 22–28, 2018.

[62] H. Liu, Z. Ruan, P. Zhao, C. Dong, F. Shang, Y. Liu, L. Yang, and R. Timofte, "Video super-resolution based on deep learning: A comprehensive survey," *Artificial Intelligence Review*, pp. 1–55, 2022.

[63] 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," *IEEE Transactions on Communications*, 2022.

[64] M. Kotaru and S. Katti, "Position tracking for virtual reality using commodity wifi," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 68–78.

[65] X. Liu, Y. Deng, C. Han, and M. D. Renzo, "Learning-based prediction, rendering and transmission for interactive virtual reality in ris-assisted terahertz networks," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 2, pp. 710–724, 2022. DOI: 10.1109/JSAC.2021.3118405.

[66] T. Hoeschele, C. Dietzel, D. Kopp, F. H. Fitzek, and M. Reisslein, "Importance of internet exchange point (ixp) infrastructure for 5g: Estimating the impact of 5g use cases," *Telecommunications Policy*, vol. 45, no. 3, p. 102 091, 2021.

[67] C.-H. Wu, X. Li, R. Rajesh, W. T. Ooi, and C.-H. Hsu, "Dynamic 3d point cloud streaming: Distortion and concealment," in *Proceedings of the 31st ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*, 2021, pp. 98–105.

[68] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "Holocast: Graph signal processing for graceful point cloud delivery," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, IEEE, 2019, pp. 1–7.

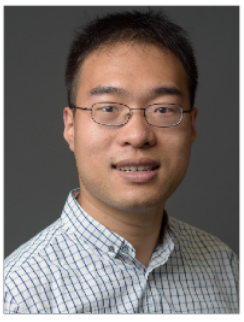
[69] T. Fujihashi, T. Koike-Akino, S. Chen, and T. Watanabe, "Wireless 3d point cloud delivery using deep graph neural networks," in *ICC 2021-IEEE International Conference on Communications*, IEEE, 2021, pp. 1–6.

[70] Z. Liu, Q. Li, X. Chen, C. Wu, S. Ishihara, J. Li, and Y. Ji, "Point cloud video streaming: Challenges and solutions," *IEEE Network*, vol. 35, no. 5, pp. 202–209, 2021.

AUTHORS

Ian F. Akyildiz received the BS, MS, and PhD degrees in Electrical and Computer Engineering from the University of Erlangen–Nurnberg, Germany, in 1978, 1981, and 1984, respectively. Currently he is the Founder and President of the Truva Inc., a consulting company based in Georgia, USA, since 1989. He is also a member of the Advisory Board at the Technology Innovation Institute (TII) Abu Dhabi, United Arab Emirates, since June 2020. He is the Founder and the Editor-in-Chief of the newly established of International Telecommunication Union Journal on Future and Evolving Technologies (ITU J-FET) since August 2020.

He served as the Ken Byers Chair Professor in Telecommunications, the Past Chair of the Telecom Group at the ECE, and the Director of the Broadband Wireless Networking Laboratory, Georgia Institute of Technology, from 1985 to 2020. He had many international affiliations during his career and established research centers in Spain, South Africa, Finland, Saudi Arabia, Germany, Russia, India, and Cyprus. Dr. Akyildiz is an IEEE Life Fellow and an ACM Fellow. He received numerous awards from IEEE, ACM, and other professional organizations, including Humboldt Award from Germany. In June 2022, according to Google Scholar his H-index is 133 and the total number of citations to his articles is more than 135+K. His current research interests include 6G/7G wireless systems, TeraHertz communication, reconfigurable intelligent surfaces, nano-networks, Internet of Space Things/CUBESATs, Internet of Bio-Nano Things, molecular communication, and underwater communication.



Hongzhi Guo is an Assistant Professor of Electrical Engineering at Norfolk State University. He received his Ph.D. degree from the University at Buffalo, the State University of New York in 2017, and his MS degree from Columbia University in 2013, both in Electrical Engineering. His broad research agenda is to develop the foundations for wireless sensor networks and networked robotics to automate dangerous dirty dull tasks in extreme environments, such as underground and underwater. He received the NSF CRII award in 2020, the Jeffress Trust Awards Program in Interdisciplinary Research in 2020, the NSF HBCU-UP RIA award in 2020, and the Best Demo Award in IEEE INFOCOM 2017.