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) is an important technology that will be supported by 6G and beyond wireless systems. It provides truly immersive experiences for a large number of novel applications, such as holographic telepresence, healthcare, retail, education, training, entertainment, sports, and gaming, by displaying multi-view high resolution 3D holograms of humans or objects/items and creating multi-sensory media (mulsemedia), including audio, haptic, smell, and taste. HTC faces great challenges in transmitting high volume data with guaranteed end-to-end latency which cannot be addressed by existing communication and networking technologies. The contribution of this paper is two-fold. First, it introduces the basics and generic architectures of HTC systems. The encoding and decoding of hologram and mulsemedia are discussed, and the envisioned use cases and technical requirements are introduced. Second, this paper identifies limitations of existing wireless and wired networks in realizing HTC and points out the promising 6G and beyond networking technologies. Particularly, for HTC sources, the point cloud encoding and mulsemedia creation and synchronization are introduced. For HTC networking, new directions and associated research challenges, such as semantic communications, deterministic networks, time sensitive networks, distributed encoding and decoding, and predictive networks, are discussed as they may enable high data rate communications with guaranteed end-to-end latency. For HTC destinations, the heterogeneity of HTC devices, synchronization, and user motion prediction are explored and associated research challenges are pointed out. Video communication with 2D content has profoundly changed our daily life and working style. HTC is an advanced technology that provides 3D immersive experiences, which will become the next research frontier.*

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

1. INTRODUCTION

A hologram is a recording of a light field that consists of the original properties of depth and parallax of 3D humans and objects/items. Holography is a technology that is used to record and generate holograms. Holography was developed by Dennis Gabor in 1948 [1], 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, hyperreal 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 remote locations. An HTC system mainly comprises three parts, namely, the source, 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 skin senses such as temperature, wind, moisture, etc. Generally, the media that includes three or more human senses is mulsemedia [2]–[4], 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 the metaverse which is a network of 3D virtual worlds with a focus on social connections [5].

A brief history of HTC. The transmission of holograms has been studied and implemented for more than a decade in some ideal scenarios with limited QoE. In 2008, Cisco demonstrated the On-Stage TelePresence Experience, where a user from India met virtually with two users from the United States. Different from videoconferencing where remote users are displayed on a 2D (two-dimensional) screen, the holographic system displayed the two remote users on a 3D (three-dimensional) display as life-size holograms. With sufficient resolution, the holograms looked the same as remote users and, thus, the system could provide immersive experiences. The Telehuman project (the Telehuman in 2012 [6] and the Telehuman2 in 2018 [7]) demonstrated low-cost de-

Fig. 1 – David Nussbaum (the founder of US holograms firm Proto) on the right-hand side and the hologram on the left-hand side. The holographic system is Proto [14], [15].

signs 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 displays, Head-Mounted Displays (HMD) for extended Reality (XR) [8], e.g., Microsoft HoloLens, are used to display holograms. The LiveScan3D developed in 2015 can display holograms in the same room as the user [9]. In [10], three different communication distances are considered. The source was located at Guildford in the UK and three destinations were in London, United Kingdom (Round Trip Time (RTT) is around 4 ms), Virginia, United States (RTT is around 85 ms), and Seoul, South Korea (RTT is around 285 ms). Generally, the user in London could obtain 30 fps reliably, while the user in Seoul experienced degraded performances. More recently, holographic technologies have been widely used in various entertainment events, such as The Whitney Houston Hologram Tour since 2020 [11] and the ABBA Voyage Concert in 2022 [12]. 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 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 Networking 2030 [13].

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

- First, we introduce the basics and the generic HTC system architecture with emphasis on the technologies used for sources, networking, and destinations. Also, we provide typical use cases and discuss the related technical challenges. This part aims to introduce HTC systems in a general way.
- Second, we identify the research gaps and challenges for the HTC systems and point out potential solutions that need to be developed in 6G and beyond wireless systems. Specifically, we discuss the HTC enabling technologies and research directions in the transport layer, network layer, data link layer and physical layer. New protocols that have the potential to improve network throughput and provide services with guaranteed and bounded end to end latency are introduced. Also, we summarize existing research progress 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 [16]. 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, we introduce the basics of hologram and HTC. After that, we discuss the HTC use cases in Section 3. In Section 2 and Section 3, we explain the differences between HTC and existing multimedia and XR technologies. The HTC system architecture and basic technical requirements are given in Section 4. Then, we discuss the research directions and potential solutions 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 [17]. The image-based solutions, such as light field videos, use 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 solutions, such as point cloud, represent real 3D objects using 3D volume pixels. Holograms are different from typical 3D images. 3D images rely on special glasses to generate 3D effects. On the contrary, holograms can be observed with naked eyes.

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 high efficiency in representing 3D objects. Com-

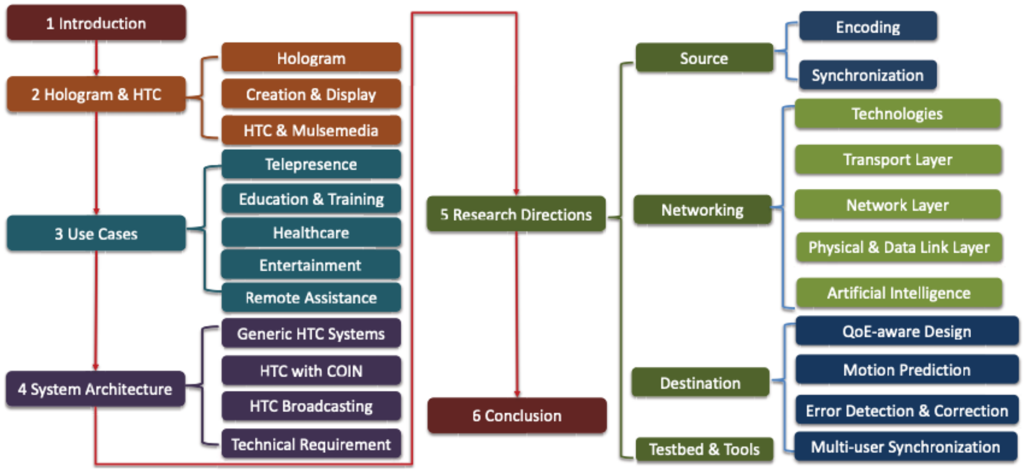


Fig. 2 – The structure of this paper.

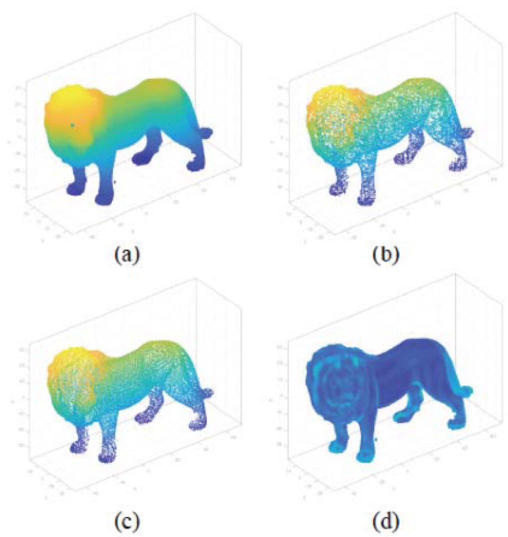


Fig. 3 – Point cloud of a lion: (a) 100% points; (b) randomly sampled 10% points; (c) interval sampling of every 10 points of the point cloud data; and (d) 100% points considering roughness of [0,1.5]. Code was adopted from [18].

pared 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 [17].

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 in [18]. 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 image. There are mainly three types of displays for HTC [19] and a comparison is given in Fig. 4.

- **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 observations. Users have to use XR HMDs to observe holograms [8]. 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 with 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 [20], [21]. External glasses and HMDs are not re-

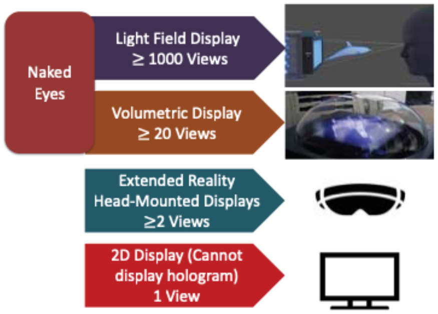


Fig. 4 – Comparison of displays. The volumetric display is Voxon VX1 and the image is from [24]. The image of light field display is from [25].

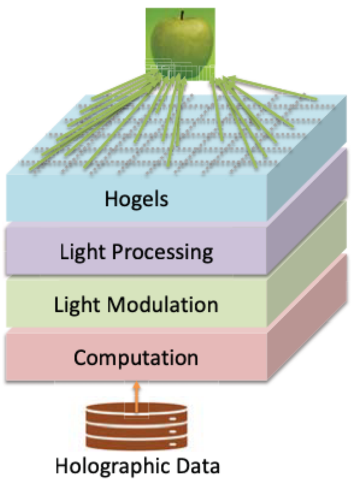


Fig. 5 – Architecture of light field displays.

quired. A 2D array of hogels are used in light field display, and the hogel can generate different light intensity in different directions to recreate the light field [22], [23], as shown in Fig. 5. The holographic data is stored locally or received from remote locations, which are processed to modulate the light in each hogel. A hogel is controlled by using a vector, including the hogel's location in the 2D hogel array and a directional vector to describe the emitted light direction. Similar to antenna arrays, the hogels form an array to generate light fields and create holograms. In reality, when we observe an object, we see reflected light field from the object. The light field display generates such a light field to show 3D effects.

A 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 a 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 of a

	Sight	Hearing	Touch	Smell	Taste
Holographic-Type Communication	✓	✓	✓	✓	✓
XR (AR, MR, & VR)	✓	✓	✓		
Haptic Communication	✓	✓	✓		
Video	✓	✓			
Image & Text	✓				
Audio		✓			

Fig. 6 – Holographic-type communication with five senses compared with other multimedia technologies.

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 that the RGB bytes per pixel is 3. More details of the calculation can be found in [22]. If the display is used for real-time HTC, considering 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 was standing next to it. To provide such an immersive user experience, various sensors and actuators have to be employed. Besides the five basic senses, some other senses can further improve the user experiences, such as the sense of balance and location, the awareness of cold and wind based on our skin, etc.
As shown in Fig. 6, 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 HMD 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, which is a challenging issue.

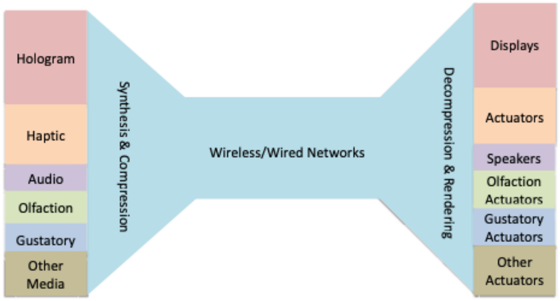


Fig. 7 – Illustration of mulsemedia systems and data transmission.

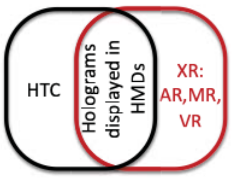


Fig. 8 – The relation between Holographic-Type Communication (HTC) and extended Reality (XR). The intersection between XR and HTC is the hologram displayed using XR HMDs.

Mulsemedia is the media that include three or more senses [2], [3]. Multimedia, including videos, audio, images, etc., uses cameras and microphones at the source to collect data. At the destination, videos and images can be displayed on monitors and the audio files can be played by speakers. Similarly, mulsemedia use various sensors at the source to collect data and various actuators, displays, and speakers at the destination to regenerate the environment. As shown in Fig. 7, the source performs data acquisition and compression. Source 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. 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.

HTC is a unique technology that differs from XR. Their relations are shown in Fig. 8. XR users can leverage HMDs to access HTC content. But XR is not the only way or the major way in the future to get access to HTC content. Users without HMDs can observe and experience HTC content with naked eyes and light field displays. Therefore, HTC and XR have intersections, but each of them has its unique aspects.

3. USE CASES

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

3.1 Holographic telepresence and conferencing

Holographic telepresence will be one of the major applications of HTC. As shown in Fig. 9, the users from different locations can meet as holograms. 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 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. 9. Consider the Whitney Houston Hologram Tour, where the singer's hologram is created by computers rather than using cameras to record the light field.

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 with a large number of products. Users with smart gloves can pick up products and put them into a holographic cart. Such a kind of virtual shopping experience is hyperreal, where the user may not be able to distinguish it from in-person shopping.

Videoconferencing has 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 videoconferences is drastically different from real in-person conferences. Videoconference users have to look at 2D screens which display remote users with reduced physical dimensions. Compared to videoconferences, 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. 10a, where holographic speakers are projected on to the screen to deliver lectures to the audience at Imperial College London.

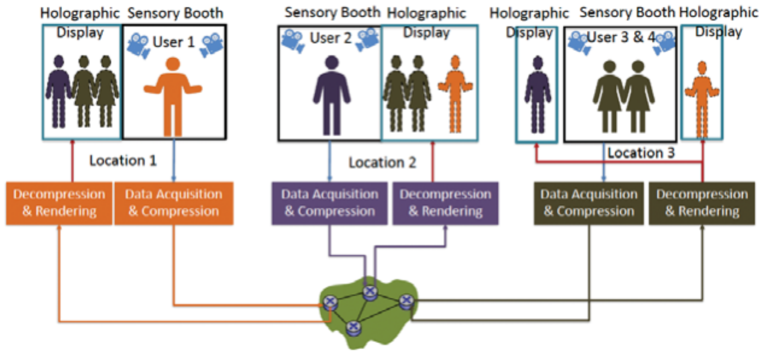
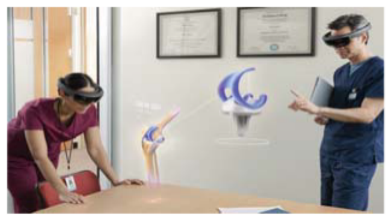
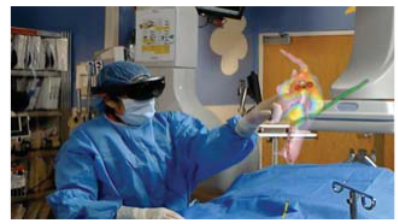


Fig. 9 – Illustration of holographic telepresence

(a) Hologram lectures at Imperial College London [26]. The life-size holographic speakers in the middle are projected onto a display.

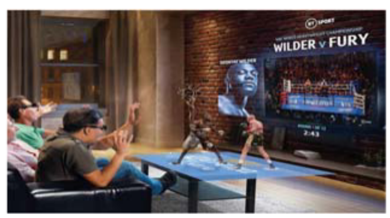


(b) Hologram curriculum developed by Pearson and Microsoft is used to train nurses [27].



(c) A doctor uses the holographic display during a cardiac ablation procedure [28].

(d) 3D reconstruction of a soccer game using XR HMDs [29].



(e) Boxing broadcast developed by Condense Reality [30].

(f) Axiom Holographics (formerly Euclideon Holographics) hologram arcade table [31], [32].

Fig. 10 – Examples of holographic-type communication use cases

3.2 Education and training

Similar to holographic conferencing, 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 gradients of high-order functions. With HTC, students can observe the 3D phenomena directly. As shown in Fig. 10b, holograms are used to train nurses using Microsoft HoloLens [27]. High-cost equipment and devices can also be converted into holograms, and students can virtually use them to reduce the cost.

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. 10c. Remote transmission of holograms will make healthcare more accessible.

3.4 Holographic entertainment

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

Holographic television. The holographic content, 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.

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. 10d shows an example of using XR HMDs to display the 3D players on a table. Similarly, any sports such as badminton, tennis, and boxing (as shown in Fig. 10e) can be broadcast using HTC.

Holographic gaming. Holographic gaming can be non-immersive or immersive. The non-immersive holographic gaming uses holographic displays, e.g., a flat holographic display on a table. Users can control holograms of 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 HMDs, specifically, Augmented Reality (AR) and Mixed Reality (MR) glasses [8], 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.

3.5 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 via a 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.

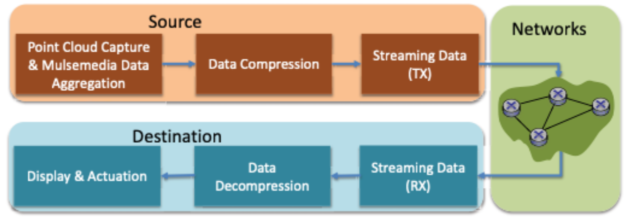


Fig. 11 – HTC system architecture with computing at the source and destination.

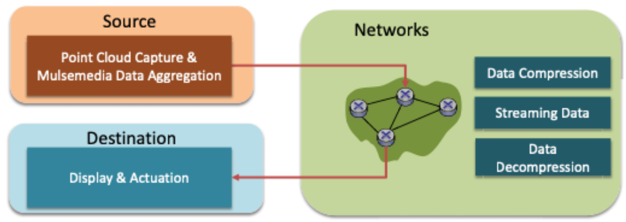


Fig. 12 – HTC system architecture with computing in the network.

4. SYSTEM ARCHITECTURE AND TECHNICAL REQUIREMENTS

Holographic content can be prerecorded and saved in typical devices, such as an external hard drive. Users play the prerecorded holographic content anywhere with light field displays. In this case, there are no end-to-end communication issues. This paper focuses on different scenarios where a source sends holographic content to one or multiple destinations. Specifically, there are the following three scenarios.

- Low-interactive HTC. This can be used for holographic remote education, where teaching materials are streamed from remote servers. Similar to a presentation where we change slides every several minutes, holographic materials need to be updated with a low frequency. As a result, the low-interactive HTC can tolerate relatively long latency requirements.
- High-interactive HTC. This can be used for live conferencing and remote control where ultra-low latency, broad bandwidth, and high reliability are required.
- HTC broadcasting. This can be used for live broadcasting of sports or public events. Users passively receive holographic content without interactions. This application can tolerate long latency requirements. Also, dynamic adaptive data packet transmission can be adopted for efficient utilization of networking resources.

4.1 System architecture

A generic HTC system architecture is shown in Fig. 11, which consists of the source, the destination, and networks. In Fig. 11, the computing tasks, including modulation, demodulation, encoding, decoding, and synchronization of aggregated data are mainly located at the source and destination. The network connects the source and destination. Next, we explain their main functionalities in this context.

- **Source.** The source mainly has three functions.

  1. It uses various sensors to capture images, sound, haptic signals, and even smell and taste signals.
  2. It processes the aggregated data. For example, it can generate point clouds using images from different angles. Also, the source can synchronize multi-sensory data to ensure that timestamps are correctly generated. More importantly, the source has to encode and compress the holographic data and mulsemedia to reduce network traffic.
  3. It utilizes HTC networking protocols to send data packets.

- **Network.** The HTC networks (wireless or wired) 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, terahertz band utilization, etc., need to be tailored to meet the requirements. New technologies, such as semantic communications and deterministic networking, new transport layer solutions, scheduling algorithms for end-to-end latency control, etc., can make 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, to recreate the environment at the source. The holographic display, actuators, and speakers are placed at different locations, but they also 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. Besides receiving and rendering content, the destination can actively correct errors, improve hologram and mulsemedia quality using super-resolution technologies, and perform motion sensing to adaptively request services.

4.2 HTC systems with computing in the network

The system architecture in Fig. 11 requires high computation capabilities at the source and destination. To meet the requirements, 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 [33], [34], as shown in Fig. 12. For

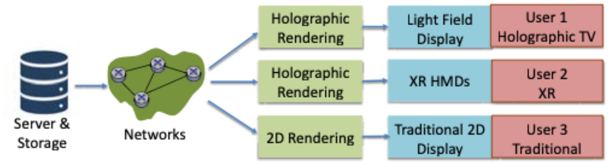


Fig. 13 – HTC broadcasting system with heterogeneous destinations.

Table 1 – Projected future HTC development.

	Current	5 Years	10 Years (6G & beyond)
Data Rates	1 Gbps	100 Gbps	>1 Tbps
Refresh Rates	≤30 fps	60 fps	≥90 fps
Distance	Local Area	Metropolitan	Very Wide Area

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 burdens at the source and destination are dramatically reduced.

4.3 HTC broadcasting systems

Besides end-to-end communication, HTC can also support the broadcasting of prerecorded or live holographic content, such as holographic TV, holographic video streaming, and asynchronous holographic education. As shown in Fig. 13, holographic content is stored in a server or a data center. HTC is used to connect servers and end users. Note that, the 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.

4.4 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 coding, communications and networking, which mainly consist 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 un-

Table 2 – Comparison of technical requirements of videos, VR, AR and MR, and HTC.

	Video	VR	MR & AR	HTC
Data Rates	35-140 Mbps	25 Mbps-5 Gbps	25 Mbps-5Gbps	1-10 Tbps
Latency	15-35 ms	<20 ms	<10 ms	<1 ms
Synchronization	Audio/Video	Multiple Tiles	Multiple Tiles	Thousands Views & Senses
Devices	Smart Devices/ Computers	Head-Mounted Displays	Head-Mounted Displays (Smart Glasses)	Naked Eyes or Head-Mounted Displays

precedented 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 Table 1. It is anticipated that the 6G and beyond systems [35] will support high-quality HTC across very wide areas.

- **End-to-end latency.** The latency in HTC networks consists of data acquisition, encoding, communication, networking, decoding, display and actuation. 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. This will limit the distance between the source and destination because the transmission of data packets cannot be faster than the speed of light.

- **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 content 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 speakers. Various types of data have to be synchronized and presented at the destination to avoid any mismatch. HTC systems need to synchronize thousands of views, as well as five basic senses of human perceptions.

A comparison of technical requirements among videos, VR, AR and MR, and HTC is given in Table 2. The approximated data was adopted from [36] and [8]. We noticed that there are some scenarios where the parameters vary significantly. This figure can provide a general description of the differences.

5. RESEARCH DIRECTIONS AND POTENTIAL SOLUTIONS

In this section, we discuss the research directions and potential solutions 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 [37] at the source, network, and destination is an important technology that can jointly 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.
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 end-to-end latency and packet losses. Nevertheless, there is a trade-off between encoding and decoding efficiency and the computation latency, 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 buffering 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 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 size, and computation complexity.
- **Design resilient and intelligent rendering algorithms in the presence of packet losses and errors**. Packet losses and errors can incur retransmissions 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. These applications also require low end-to-end latency, e.g., smaller than 1 ms. Consider 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. In the following sections, we discuss the source, networks, and the destination separately.

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 [38], [39]. 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 clouds, such as 3D maps.

As shown in [40], 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. Existing encoding technologies need to be improved to satisfy the requirements of HTC. Particularly, the following three research directions have the potential to address this issue.

1. Increase the compression ratio. It has been shown that XR videos can be compressed significantly with relatively low quality to reduce network traffic [41]. The destination uses deep learning-based super-resolution technology to enhance the received video quality. Traditional lossy compression reduces the quality of the decompressed point cloud. However, deep learning-based solutions can overcome this issue by using pretrained models. Future research can focus on understanding the limitation of deep-learning based super-resolution technology and its application in point cloud encoding and decoding. Also, the correlation between adjacent frames can be used to increase the compression ratio.

2. Reduce encoding computation complexity. The high computation complexity increases end-to-end latency and demands significant computation resources. It is necessary to reduce the encoding computation complexity. However, there is a tradeoff between the compression ratio and computation complexity. Usually, a large compression ratio requires high computation complexity, and vice versa. In addition to existing V-PCC solutions, machine learning can be used to train encoding models which can efficiently extract features. These features can be sent to the destination and another machine learning model can be used to recover the original source information. Although the training can cost offline computation resources, the real-time online deployment is simple.

3. Leverage the coordination between the source and the destination. The information of eye-tracking of the destination user 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. In other words, if the source can leverage the feedback information from the destination to plan its transmission, redundant information can be reduced in the encoding and decoding process. It is essential to study the required feedback information for such coordination, the impact of feedback latency, and the effect of inaccurate feedback information.

5.1.2 Mulsemedia synchronization

Mulsemedia includes different types of sensors at different locations, and the synchronization is important [42]. 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 for the following two reasons.

1. Different types of sensors have different response times. For example, cameras can have a capture rate of 50 frames/images 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 [43]. 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. Machine learning classifiers and predictors can be used to verify whether collected senses are consistent and automatically generate senses that are missing or delayed.

2. Human perception has different levels of latency tolerance for different senses. As the study in [42] shows, the latency of air-flow media can be as high as 1 s for acceptable user experiences. 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 lower than 1 ms.

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

1. 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.

2. Study mulsemedia end-to-end latency requirements. Currently, there is no clear requirement of 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.

3. 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.

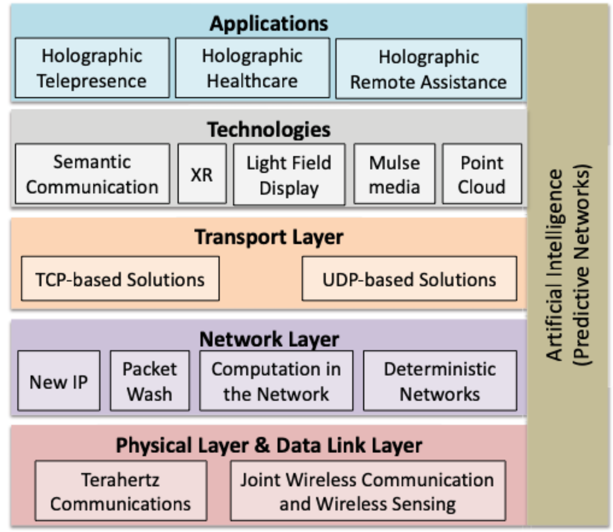


Fig. 14 – Networking protocol stacks and key technologies.

Note that, MPEG-V provides a framework to create mulsemedia for interactions between the physical world and a virtual world [44]. 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 network throughput and the end-to-end latency are not independent of each other due to the constraints of networking protocols. Next, we first introduce the generic HTC networking protocol stacks and then we introduce each layer of the stack.

Despite HTC introducing many novel technologies that can transform existing computer networks, it still needs to be gradually developed based on existing networking protocols. In Fig. 14, we show the key novel technologies that will be used to enable HTC. The use cases that were discussed in Section 3 are supported by HTC-enabling technologies, such as semantic communication, XR, mulsemedia, and point cloud encoding and decoding. Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) are widely used transport layer protocols. TCP-based protocols are reliable, but the end-to-end latency is large. On the contrary, UDP-based protocols experience packet losses, but the end-to-end latency is small. Streaming HTC content has unprecedented challenges and, thus, it is essential to develop and identify the optimal transport layer protocols. The network layer also demands novel technologies such as the New IP [45], [46], packet wash, deterministic networks and computa-

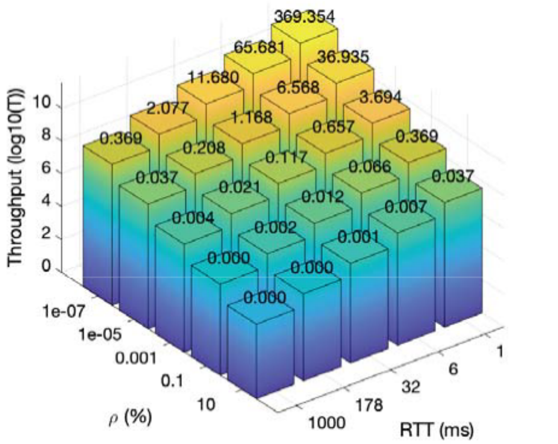


Fig. 15 – TCP throughput under different packet losses and round trip time (the unit of the text notations above each bar is Gbps, and three fractional digits are used). The throughput (T) is shown in logscale $log_{10}(T)$.

tion in the network to effectively reduce the end-to-end latency and support the source encoding and the destination decoding. The physical layer and data link layer use 6G and beyond technologies, including terahertz communications, visible light communications, among others. Also, the joint (or coexistence of) wireless communication and wireless sensing will play an important role in HTC. Wireless sensing can free users from wearing sensors which provide more flexibility and usability. In addition, artificial intelligence can be applied in each layer to provide prediction and classification capabilities.

Existing computer networks cannot support high-quality HTC. For example, if TCP is used to stream HTC content, the throughput can be obtained by [47]

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

where MSS is the maximum segment size, RTT is the 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 [47]. When $MSS = 1460$ Bytes, the throughput in logscale is shown in Fig. 15. 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 a 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.

The development of HTC network protocols may result in a new network paradigm. In other words, if successful, the network protocol stacks in Fig. 14 will gradually evolve into the one in Fig. 16. The holographic applications are supported by hologram and mulsemedia technologies. The source and the destination need to coordinate and select the quality of holograms and the number of senses that can be supported based on their available

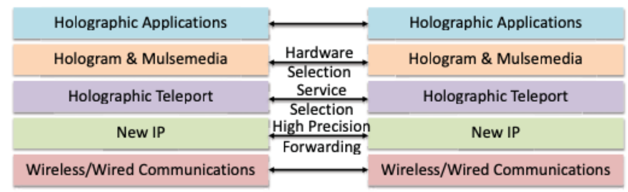


Fig. 16 – HTC networking protocol stacks.

hardware. The holographic teleport [48] can provide various services to meet the low latency and high throughput requirement. The teleport is more towards real-time applications than the transport. The New IP layer can forward packets from the source to the destination with predefined and guaranteed end-to-end latency. These layers are supported by wireless or wired physical communications.

Next, we discuss the challenges and potential research directions for each layer in Fig. 14 from the top to the bottom. Since holographic applications have been discussed in Section 3, our discussion starts from the technologies.

5.2.1 Technologies

Holograms that are represented by point cloud and mulsemedia are the major content of HTC, which have been introduced in Section 5.1. XR is an affordable way to use HTC, and a technical review is presented in [8]. Light field displays present high-end HTC content that can be observed using naked eyes. However, existing light field displays are expensive. Novel materials and displaying technologies need to be developed to make light field display more accessible. Semantic communication based on deep learning architectures has been proposed which can efficiently reduce network traffic by sending meaning instead of bits. Since it has not been employed in HTC, we mainly introduce semantic communication in this subsection.

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

- 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 only recently received attention due to the advancement of machine learning technologies [50]–[52]. Semantic communication can reduce network traffic to efficiently send holograms and mulsemedia. Instead of focusing on the technical problem of transmitting symbols correctly, semantic communication focuses on correctly delivering the meaning by extracting semantic information from the raw data. Deep learning plays an important role in semantic communication by understanding the meaning of transmitted information, which can efficiently compress the source data.

Different from text, image and audio [50], [53], [54], holograms and mulsemedia are more complex, and it is challenging to extract meaning from them. Potential research directions include:

1. Sense-based semantic communication. The joint source-channel coding can efficiently extract and reliably transmit semantic information. Existing research work has studied the joint source-channel coding of transmitting text and images [50], [53], [54]. Semantic communication for holograms and mulsemedia, especially point cloud videos, is an open research problem. The raw data collected by different sensors and cameras can be considered as inputs of different deep learning models which can perform joint source-channel coding. Since different senses have drastically different data patterns, the sense-based semantic communication treats different senses using different deep learning models.

2. Environment-based semantic communication. Instead of extract meaning from individual senses, the environment-based semantic communication can jointly understand the environment using all available senses. Since the human perceptions and reactions are correlated, it is more efficient to leverage such correlations to reduce the deep learning model complexity and rendering coherent HTC content on the destination side.

5.2.2 Transport layer

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 Dynamic Adaptive Streaming over HTTP (MPEG-DASH) is a promising standard for low-interactive HTC applications [55]. 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. However, the MPEG-DASH saves content with different qualities in servers which may occupy large memories for HTC applications. Also, it cannot support high-interactive HTC applications due to long latency.

UDP-based protocols, such as Web Real Time Communications (WebRTC), need more investigation and redesign to transmit HTC data [56]. They have the potential to support high-interactive HTC applications. QUIC has been used for 360-degree video transmission in XR systems. It can multiplex streams with prioritized schedulers and achieve low end-to-end latency. Prioritized transmission is essential to meet the latency requirements, i.e., data packets with high priorities can obtain sufficient network resources. QUIC will be an important transport protocol in HTC. Holograms can be divided into tiles and only the tiles that are in the destination user's field-of-view will be given high priorities.

Since HTC is an emerging technology, there are no widely used steaming protocols currently. 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. In the literature, TCP is barely used to stream high-volume data with strict latency constraints. There are two potential research directions to develop transport protocols for HTC.

1. UDP protocols and TCP protocols can be jointly designed to meet HTC latency and reliability requirements. In [57], DASH is used on top of the QUIC protocol. First, it leverages the simplicity of DASH to encode source data into tiles. Then, the QUIC protocol can transmit tiles with different priorities. This is an important direction to combine the strength of existing protocols in order to meet HTC requirements.

2. UDP protocols can be improved and enhanced for HTC applications. Existing UDP protocols have some limitations, e.g., precisely controlling the stream delivery time or coordinating multiple streams to meet the deadline are not possible. Since HTC has multisensory data which can be encoded into different streams, coordination and precisely controlling the delivery deadline are important. Existing UDP protocols can be enhanced by adding more functionalities. For example, a deadline-aware QUIC protocol is proposed in [58] to stream 360-degree videos.

5.2.3 Network layer

The HTC network layer needs to be highly programmable, reconfigurable, and adaptable, which can be achieved by the New IP design [45], [46].

The New IP allows packets to be delivered based on service requirements which are defined in the packet contract. In [45], [59], qualitative communication is studied. It sends packets with different significance values and selectively drops packets based on their significance values when there is congestion. The significance value can be included in New IP packets [45], [46]. An example of a New IP-type packet is shown in Fig. 17. The packet gives clear information for network routers about how

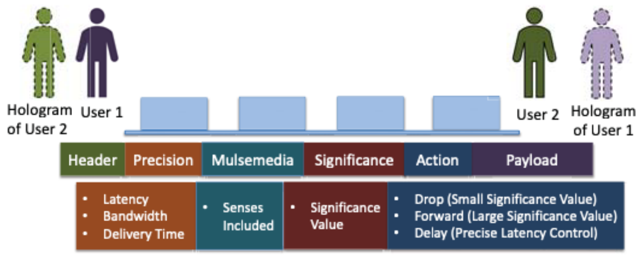


Fig. 17 – New IP-type packet for HTC networks using semantic communication and packet wash.

and when to deliver the packet to the destination. Packets can be dropped based on their significance value using packet wash [60] when there is network congestion. 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.

The New IP can be used to deliver DASH or semantic packets based on priorities. For example, HTC packets can be categorized based on the level of semantic information. The significance value can be derived from the meaning of semantic communication. For example, if a packet carries most of the meaning of the source, its significance value is high. On the contrary, if a packet did not carry information about the meaning of the source, but it includes a certain amount of background information, its significance value is small. When the network bandwidth is sufficient, raw data packets can be sent together with semantic packets, while when there is network congestion, the raw data can be gradually dropped as it moves towards the destination. Adaptive semantic communication needs to be designed to fully leverage the network resources to provide the optimal QoE. At the destination, dropped packets with small significance values can be recovered using deep learning.

Besides new protocols in the network layer, new services are also needed to support HTC applications. 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 experiences. 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 deadline, so that they can be processed immediately without occupying the buffer. This requires a deterministic network where packets can be delivered as scheduled. In [46], [61], [62], three different time-related services are proposed for Network 2030:

- **In-time guarantees.** Data packets arrive before a specific time, as shown in Fig. 18.
- **On-time guarantees.** Data packets arrive at a specific time with guaranteed variances, as shown in Fig. 19.
- **Coordinated guarantees.** Multiple flows of data packets arrive with in-time or on-time guarantees, as shown in Fig. 20.

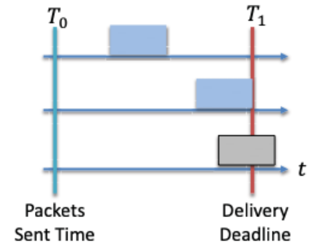


Fig. 18 – In-time guarantee. The packet delivery deadline is T_1. The top and the middle packets are delivered before the deadline, which meet the requirements. The bottom packet is fully delivered after the deadline.

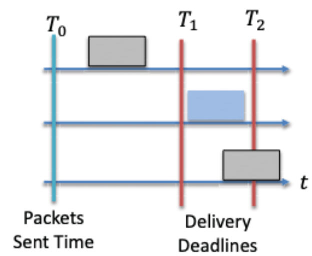


Fig. 19 – On-time guarantee. The packet delivery deadline is between T_1 and T_2. The top packet is delivered early and the bottom packet is delivered late. The packet in the middle is delivered on time, which meets the requirement.

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 realize 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 and haptic data size is large and consumes significant storage resources. On the contrary, if all the flows are delivered on time, the destination can synthesize and render the received mulsemedia immediately.

The Deterministic Networking (DetNet) architecture is proposed by the IETF DetNet working group to provide bounded network latency and reduce packet loss [63]. It can reduce the unexpected randomness in networks. The Time-Sensitive Network (TSN) following IEEE 802.1 is [64] 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.

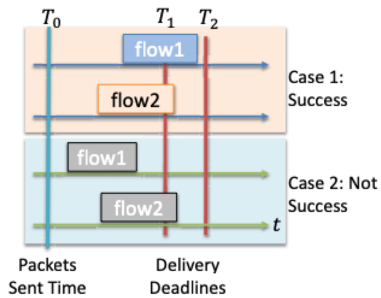


Fig. 20 – Coordinated guarantees. The upper two flows of packets are delivered with on-time guarantee, while the lower two flows are not.

5.2.4 *Physical layer & data link layer*

The physical layer and data link layer face challenges in providing high data rates with low latency and high reliability. Terahertz communication is the major technology that will be used due to its broad bandwidth. It can achieve several terabits per second data rates that can support high-end light field displays [65]. Besides research challenges in the physical and data link layers of 6G and beyond wireless systems, HTC has the following unique challenges.

1. High energy and spectrum efficiency. The high-volume data of holograms and mulsemedia requires extremely high energy efficiency of communication systems, especially for battery-powered devices. High energy efficiencies of communication systems avoid frequently changing/recharging batteries. Terahertz communication can achieve high energy efficiency with ultra-high data rates. Existing work has studied the antenna design, channel modeling, modulation waveform design, MIMO systems [65]–[67], etc. More recently, terahertz communication systems have been used for XR applications, where servers and HMDs communicate in the terahertz band [68], [69]. Future research can further improve the energy and spectrum efficiency to achieve beyond 1 Tbps data rates that can support high-quality light field displays.

2. High reliability. Terahertz communication signals are highly directional and prone to blockages. User mobility and dynamic environments can affect the communication reliability. The terahertz reconfigurable intelligent surfaces can reflect wireless signals towards optimized directions, which avoid blockages, align interferences, and enhance received signals [70]. Although terahertz reconfigurable intelligent surfaces have been designed, it is not clear how to optimally deploy these surfaces for HTC applications in various scenarios [71].

3. Coexistence of wireless communication and wireless sensing. Wireless sensing is an essential part of HTC systems since it can replace wearable sensors for hand tracking, head tracking, and body tracking. The data link layer needs to enable coexistence of wireless communication and wireless sensing. Since the terahertz band has a broad bandwidth, a small sub-band can be allocated for wireless sensing. Also, the terahertz reconfigurable intelligence surface can be used as a sensing hub, which passively monitor the environment based on incident and reflected waves [71].

5.2.5 *Artificial intelligence: predictive networks and edge intelligence*

AI can be used in each layer in Fig. 14. Here, we primarily introduce the intelligence of the network, i.e., predictive networks, and edge intelligence.

Network control and management using model-based solutions have been widely used. However, existing networks still experience failures occasionally. AI provides another solution that can strengthen existing network control and management. 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 network is an example of this technology [72]. Also, based on partial observation of the network status, the QoS parameters can be estimated, such as latency and bandwidth, to determine if the user's requests can be served [73], [74].

AI can be used to solve complex networking problems, such as network planning. Network planning is an important part for supporting 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 [75].

Last, AI can be used in real-time network control, such as network monitoring, adaptive routing, and network slicing [76]–[79]. AI-empowered networks can efficiently address the challenges that are faced by TCP and UDP by predicting and controlling network traffic, allocating sufficient bandwidth, and preventing packet losses. This will provide reliable and efficient services for HTC networks.

Edge intelligence [37] will play an important role in HTC networks. HTC and edge services can be sold to users together in order to ensure that HTC services can be smoothly provided. The cloud-based solutions aggregate data from various sensors and perform point cloud compression, which creates a long delay due to the long distance between the user and the cloud. Also, the cloud may serve many users simultaneously, which 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. Source sensors may have limited computation capabilities, and they offload computation tasks to the edge servers. 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. The sensors, actuators, displays, and speakers with edge intelligence at the destination become more powerful in processing received data. Different senses may require different intelligent algorithms for decoding and synchronization, and the specialized edge intelligence for different devices can improve their performance in error detection and correction, sensing, and prediction.

5.3 Destination

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

1. 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 requests the semantic meaning without other detailed data. This task includes the QoE-aware design and AI-empowered motion prediction at the destination.

2. 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.

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 [17]. For instance, in a telepresence conference, a user may only focus on the hologram's face. All other parts of the body use a reduced resolution. Also, the destination needs to communicate with the source to request the desired QoE. The source employs 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 [80], it is still an open research problem in HTC networks.

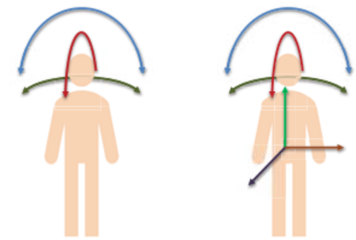


Fig. 21 – Illustration of the 3 degrees of freedom and 6 degrees of freedom

AI-empowered wireless sensing and motion prediction. Motion prediction of the destination user is necessary to allow the source to adaptively transmit data and reduce the end-to-end latency. For example, when the user is looking towards a direction, the source with predicted knowledge can transmit high-resolution information to the destination user's Field of View (FoV). Motion prediction uses various sensors, such as cameras and inertial measurement units, to track the movement of the head, hands, and the 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 extra devices.

Generally, the motion prediction for HTC encounters three major challenges.

1. 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. 21. 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. Terahertz signals will be used in 6G and beyond wireless systems. Due to the short wavelength and thus high resolution, the sensing accuracy is high [65]. Wireless access points can be leveraged to track users' motion due to its ubiquitous availability [81].

2. How to predict users' motion based on collected sensing data. Deep learning-based architectures have been extensively used for eye tracking [68] and other motion sensing and prediction, which have demonstrated high accuracy.

3. How to mitigate the impact of prediction errors. If the prediction is not accurate, users may experience sickness [82]. 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.

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 [83]. As shown in [83], 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 [84], [85], the point cloud is considered as a graph. Graph signal processing and graph neural networks are used to develop a robust communication system in the presence of channel fading and noises. Based on this work, more advanced AI-empowered error detection and correction schemes can be developed. Deep learning architectures, such as Generative Adversarial Networks (GAN), can be used to correct errors and improve the QoE.

Multi-user synchronization. The scalability of HTC networks is a challenging problem. Consider the telepresence conferencing, when there are two users, the number of sensors, actuators, displays and speakers is 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 is 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 Testbed design

Existing HTC testbeds mainly transmit holographic data without mulsemedia [86]. Currently, mulsemedia testbed design is an independent research direction [44]. 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 testbeds use XR HMDs or even 2D computer monitors as the display to evaluate the quality of the received hologram [9], [10], [86]. Point cloud data can be generated from multiple RGB-D cameras, e.g., Microsoft RGB-D Kinect 2.0. The LiveScan3D toolkit is an efficient tool to generate point cloud data based on synchronized 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 [10], 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 [7], [20]. Although it can successfully display holograms streamed from remote locations, the quality still needs to be improved. Also, compared to XR HMDs, the light field display is more expensive and development toolkits are sparse. 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 enables novel applications, such as telepresence conferencing and remote surgery. This paper provides the fundamentals 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, networks, and the destination. Promising solutions that will be developed in 6G and beyond wireless systems and Network 2030 to realize high-quality HTC are identified and introduced. Although basic HTC testbeds have been developed, there is still a significant technical gap to develop advanced HTC systems with light field displays that can connect long-distance users for high-interactive applications. It is anticipated that this technical gap will be addressed when the light field displays are more accessible and artificial intelligence technologies are more widely used in HTC systems.

ACKNOWLEDGMENT

The authors would like to thank Martin Reisslein and the three anonymous reviewers for their insightful comments and valuable suggestions that have significantly improved the quality of this paper.

REFERENCES

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

[2] 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.

[3] 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.

[4] A. Covaci, L. Zou, I. Tal, G.-M. Muntean, and G. Ghinea, "Is multimedia multisensorial?-a review of mulsemedia systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–35, 2018.

[5] 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.

[6] 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.

[7] 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.

[8] 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, pp. 1–15, 2022.

[9] 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.

[10] 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.

[11] "An Evening With Whitney," 2022. [Online]. Available: `https://basehologram.com/productions/whitney-houston`.

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

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

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

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] M. Lucente, "The first 20 years of holographic video—and the next 20," in *SMPTE 2nd annual international conference on stereoscopic 3D for media and entertainment*, 2011, pp. 21–23.

[24] "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/`.

[25] 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`.

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

[27] "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/.

[28] 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.

[29] 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.

[30] "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/.

[31] 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/.

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

[33] 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*.

[34] 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.

[35] 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.

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

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] M. Dasari, A. Bhattacharya, S. Vargas, P. Sahu, A. Balasubramanian, and S. R. Das, "Streaming 360-degree videos using super-resolution," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, IEEE, 2020, pp. 1977–1986.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] R. Li *et al.*, "Towards a new internet for the year 2030 and beyond," in *Proc. 3rd Annu. ITU IMT-2020/5G Workshop Demo Day*, 2018, pp. 1–21.

[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] I. Sodagar, "The mpeg-dash standard for multimedia streaming over the internet," *IEEE multimedia*, vol. 18, no. 4, pp. 62–67, 2011.

[56] S. Zhao, Z. Li, and D. Medhi, "Low delay mpeg dash streaming over the webrtc data channel," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, IEEE, 2016, pp. 1–6.

[57] S.-C. Yen, C.-L. Fan, and C.-H. Hsu, "Streaming 360 videos to head-mounted virtual reality using DASH over QUIC transport protocol," in *Proceedings of the 24th ACM Workshop on Packet Video*, 2019, pp. 7–12.

[58] H. Shi, Y. Cui, F. Qian, and Y. Hu, "Dtp: Deadline-aware transport protocol," in *Proceedings of the 3rd Asia-Pacific Workshop on Networking 2019*, 2019, pp. 1–7.

[59] 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.

[60] 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.

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

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

[63] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein, and H. ElBakoury, "Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 88–145, 2018.

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

[65] 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*, vol. 70, no. 6, pp. 4250–4285, 2022.

[66] Jornet, Josep Miquel and Akyildiz, Ian F, "Graphene-based plasmonic nano-antenna for terahertz band communication in nanonetworks," *IEEE Journal on selected areas in communications*, vol. 31, no. 12, pp. 685–694, 2013.

[67] Akyildiz, Ian F and Jornet, Josep Miquel, "Realizing ultra-massive mimo ($1024\times$ 1024) communication in the (0.06–10) terahertz band," *Nano Communication Networks*, vol. 8, pp. 46–54, 2016.

[68] 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.

[69] J. Du, F. R. Yu, G. Lu, J. Wang, J. Jiang, and X. Chu, "MEC-assisted immersive VR video streaming over terahertz wireless networks: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9517–9529, 2020.

[70] S. Dash, C. Psomas, I. Krikidis, I. F. Akyildiz, and A. Pitsillides, "Active control of thz waves in wireless environments using graphene-based ris," *IEEE Transactions on Antennas and Propagation*, 2022.

[71] 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.

[72] J. Vasseur, "Predictive networks: Networks that learn, predict and plan," *Cisco White Paper*, pp. 1–9, Jul. 2022.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] 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.

[84] 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.

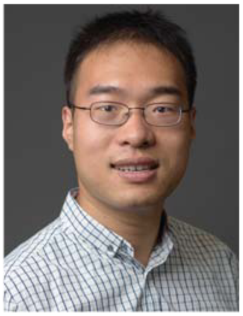
[85] T. Fujihashi, T. Koike-Akino, S. Chen, and T. Watanabe, "Wireless 3D point cloud delivery using deep graph neural networks," in *ICC 2021-IEEE International Conference on Communications*, IEEE, 2021, pp. 1–6.

[86] 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 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 has also been 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 Editor-in-Chief of the International Telecommunication Union Journal on Future and Evolving Technologies (ITU J-FET), which was established in 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 has had many international affiliations during his career and has 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 has received numerous awards from IEEE, ACM, and other professional organizations, including the Humboldt Award from Germany. In August 2022, according to Google Scholar his H-index is 133 and the total number of citations to his articles is more than 136K+. 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 CAREER award in 2022, 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.