Millimeter Wave and Free-Space-Optics for Future Dual-Connectivity 6DOF Mobile Multi-User VR Streaming

JACOB CHAKARESKI, New Jersey Institute of Technology, USA
MAHMUDUR KHAN, York College of Pennsylvania, USA
TANGUY ROPITAULT, Wireless Networks Division, National Institute of Standards and Technology, USA
STEVE BLANDINO, Wireless Networks Division, National Institute of Standards and Technology, USA

Dual-connectivity streaming is a key enabler of next generation six Degrees Of Freedom (6DOF) Virtual Reality (VR) scene immersion. Indeed, using conventional sub-6 GHz WiFi only allows to reliably stream a low-quality baseline representation of the VR content, while emerging high-frequency communication technologies allow to stream in parallel a high-quality user viewport-specific enhancement representation that synergistically integrates with the baseline representation, to deliver high-quality VR immersion. We investigate holistically as part of an entire future VR streaming system two such candidate emerging technologies, Free Space Optics (FSO) and millimeter-Wave (mmWave) that benefit from a large available spectrum to deliver unprecedented data rates. We analytically characterize the key components of the envisioned dual-connectivity 6DOF VR streaming system that integrates in addition edge computing and scalable 360° video tiling, and we formulate an optimization problem to maximize the immersion fidelity delivered by the system, given the WiFi and mmWave/FSO link rates, and the computing capabilities of the edge server and the users' VR headsets. This optimization problem is mixed integer programming of high complexity and we formulate a geometric programming framework to compute the optimal solution at low complexity. We carry out simulation experiments to assess the performance of the proposed system using actual 6DOF navigation traces from multiple mobile VR users that we collected. Our results demonstrate that our system considerably advances the traditional state-of-the-art and enables streaming of 8K-120 frames-per-second (fps) 6DOF content at high fidelity.

1 INTRODUCTION

Virtual reality holds tremendous potential to advance our society and is expected to impact our quality of life and the economy. Together with another emerging technology, 360° video, VR can suspend our disbelief of being at a remote location, akin to *virtual human teleportation* [1–3]. 360° video streaming to VR headsets is gaining popularity in diverse areas, e.g., gaming and entertainment, education, healthcare, and remote monitoring. The present state of the world (online classes, work from home, telemedicine, etc.) due to the COVID-19 pandemic aptly illustrates the importance of remote 360° video VR immersion and communication.

Traditional wireless communication systems are far from meeting the performance requirements of the envisioned virtual human teleportation. For instance, MPEG recommends a minimum of 12K high-quality spatial resolution and 100 fps temporal frame rate for the 360° video experienced by a VR user [4]. These requirements translate to a data rate of several Gbps, even after applying state-of-the-art High Efficiency Video Coding (HEVC) compression. To enable next-generation societal VR applications, novel non-traditional wireless technologies

Authors' addresses: Jacob Chakareski, New Jersey Institute of Technology, USA; Mahmudur Khan, York College of Pennsylvania, USA; Tanguy Ropitault, Wireless Networks Division, National Institute of Standards and Technology, USA; Steve Blandino, Wireless Networks Division, National Institute of Standards and Technology, USA.

1551-6857/2022/6-ART $15.00
https://doi.org/10.1145/3544494

need to be explored. FSO and mmWave are two such emerging technologies that enable much higher data rates relative to traditional wireless systems. We denote both technologies as *xGen* henceforth.

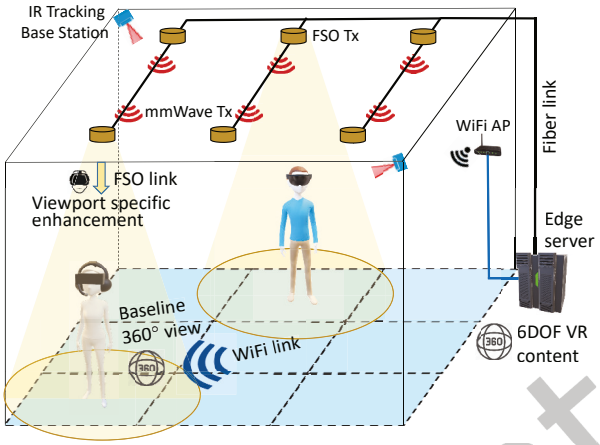


Fig. 1. A 6DOF mobile VR arena WiFi-xGen dual-connectivity scalable streaming system. A WiFi link delivers the present 360° panorama of a user at baseline quality. A directed xGen link delivers a respective viewport-specific enhancement.

Toward the objective above, we investigate an integrated dual-connectivity streaming system for future 6DOF mobile multi-user VR immersion. Our enivisioned system can be useful for both civil and military applications. Students can take virtual field trips to remote destinations. Firefighters can use robot cars or drones to get 6DOF 360° VR videos to assess the condition of a disaster affected area. The system can also be utilized for military training and operations to explore unknown territories. The proposed system is illustrated in Figure 1 and synergistically integrates a traditional sub 6 GHz wireless technology (WiFi) with either FSO or mmWave (xGen), for complementary dual link transmission. Precisely, there is no switching between the WiFi and the xGen (mmwave/FSO) wireless links, they are utilized simultaneously, in parallel. The clients navigate the 6DOF content in the indoor arena equipped with both WiFi and xGen transmitters. The users are equipped with respective WiFi-xGen enabled VR headsets. An edge server identifies the present 360° viewpoint of a user in the 6DOF arena, and partitions the respective 360° video into two embedded representations: a baseline representation of the entire 360° panorama, and a viewport-specific enhancement representation, capturing at high fidelity only the current viewing direction of the user (aka viewport). The server dynamically adapts the two representations to the available transmission rates of the two parallel link (WiFi and xGen). For efficient utilization of the high capacity xGen links and high computation capability of the server, a portion of the viewport-specific enhancement representation may be decoded at the server and streamed as raw data, and the remaining portion is streamed as compressed data. The baseline representation is streamed over WiFi and the enhancement representation is streamed over a directed xGen link. The viewport-specific content from the two representations is then integrated at the user headset to enable high-fidelity 360° remote VR immersion.

The advanced streaming and processing capabilities of the server are enabled by the scalable multi-layer 360° tiling design that we synergistically explore. Our novel dual-connectivity WiFi-xGen architecture aims to use *the best of both worlds*, as follows. Traditional WiFi is used for its robustness, to transmit a lower-quality baseline representation of the VR content and enable application reliability, and xGen is used for its large transmission capacity, to send a complementary high-quality user viewport-specific enhancement representation and enable high-quality immersion. The two representations are then synergistically integrated at the user to considerably augment her quality of immersion and experience. Our system and its key components are described in Section 3. The main contributions of the paper are:

- We enable 6DOF VR-based remote scene immersion using a dual-connectivity multi-user streaming system.
- We formulate an optimization problem that aims to maximize the delivered immersion fidelity across all users in our system. It depends on the WiFi and mmWave/FSO link rates, the computing capabilities of the edge server and user headsets, and system latency requirements.
- We formulate a geometric programming based optimization framework to solve the problem at lower complexity.
- We analyze several methods to guarantee xGen connectivity despite user mobility and head movements.
- We collect 6DOF navigation data to enable realistic evaluation of our framework and demonstrate that both dual-connectivity options, WiFi-mmWave/FSO, enable streaming of high fidelity 8K-120 fps 6DOF content.
- We evaluate the system's performance for two different settings $WiFi - mmWave$ and $WiFi - FSO$, and explore the trade-offs of each xGen technology.
- We compare to the traditional wireless streaming state-of-the-art and demonstrate notable performance advances in delivered immersion fidelity and application reliability.

We note that most existing VR systems enable only three degrees of navigation freedom (3DOF) or capability for a user in the form of head rotation movements (see Section 3.1, Figure 2 right). That is because introducing further spatial mobility for the users (another 3DOF) represents a major challenge on top of an already very demanding application to deliver. Concretely and briefly, enabling the additional spatial mobility introduces further challenges in terms of dynamically identifying the 6DOF coordinates of a user, rendering the content and environmental context appropriately, and delivering them at high fidelity to the user. Simultaneously, the stringent application latency constraints need to be met. All these aspects are especially challenging in the wireless context, where the communication and computational resources and capabilities of the system are traditionally much more limited, when compared to wireline settings. *Our study is the first to investigate a mobile (untethered) multi-user 6DOF VR system, aiming to enable high quality immersion and experience, and understand the fundamental performance trade-offs that arise herein.*

We organize the rest of the paper as follows. We review related work in Section 2. We present our system framework in Section 3. We describe the optimization problem we pursue in Section 4 and present our solution to it in Section 5. We describe the xGen-channel models in Section 6. We carry out our experimental evaluation in Section 8. Finally, we conclude this paper in Section 9.

2 BACKGROUND AND RELATED WORK

FSO exploits the light intensity of a light emitting diode (LED)/laser diode (LD) to modulate a message signal. After propagating through the optical wireless channel, the light message is detected by a photo-diode [5]. Unlike the radio frequency spectrum, plentiful unlicensed spectrum is available for light communications, which has put FSO on the road-map towards sixth generation (6G) networks [6]. While being a novel technology, a few studies of design concepts and experimental testbeds have already appeared [7, 8].

In the radio frequency spectrum, mmWave wireless communication is the enabling technology of new-generation wireless systems, as in the range of 10-100 GHz, more than 20 GHz of spectrum is available for use by cellular or Wireless Local Area Network (WLAN) applications. mmWave has seen its first commercial products, operating in the 60 GHz band, appeared in the early 2010s. More complex transmission schemes to increase even further the achievable data rate are currently being investigated [9].

Emerging VR applications require streaming of high fidelity remote (real or virtual) scene 360° video content, possibly with 6DOF user mobility. Relative to traditional video streaming [10], VR-based 360° video streaming introduces further challenges by requiring an ultra high data rate, hyper intensive computing, and ultra low latency [11]. Though some advances have been made in 360° video streaming using traditional network systems, by intelligent resource allocation and content representation [12–14], the delivered immersion is still limited to low to moderate quality and 4K spatial resolution, encoded at a temporal rate of 30 frames per second. This

outcome is due to fundamental limits in data rate and latency of such systems and their use of traditional server-client architectures. Moreover, wireless VR systems rely on the storage and computing capabilities of the headset itself or the mobile device attached to it. The GPU performance of such devices lags behind desktop GPUs by a factor of 6-10, thus, cannot provide the same quality of remote VR immersion, as experienced in wired VR settings [15].

Limited recent work has considered the integration of xGen and VR. A fixed mmWave-based VR system was proposed in [16] using a WiGig modules for wireless connectivity. A mmWave reflector was developed in [17] to aid in connectivity maintanence with a VR headset in the event of object blockage. A design concept for using narrow beam FSO transceivers was proposed in [7]. An FSO-enabled VR headset design comprising multiple highly directional photodectors (PDs) was studied to address 3DOF VR user head movements [18]. The application we consider here requires untethered streaming of high fidelity 360° real remote scene content during very challenging 6DOF user mobility, which cannot be addressed by these methods. Our proposed system addresses these challenging requirements via the synergistic integration of WiFi-xGen dual-connectivity, scalabe multi-layer 360° tiling design, and different approaches for xGen connectivity maintenance.

We note that point clouds represent a recent media representation format that captures the shape (geometry) and external coloring of a 3D object. Streaming of point cloud has been considered in the context of augmented reality (AR) applications [19, 20] and is often denoted there as volumetric video streaming. These application and technology are different from the more challenging context we consider, where multiple users are immersed in a VR-reconstructed remote environment with the capability for 6DOF navigation. In the case of volumetric video streaming, one object is digitally rendered on an AR headset and placed in the actual physical surrounding of the user, to be observed from different distances and perspectives. In our case, (a volume of) an entire remote environment is rendered (volumetrically = 6DOF capability) on a VR headset, to enable the user to virtually move through it and observe it from various perspectives.

Finally, broader related work includes prior studies of centralized and distributed multi-user streaming [10, 21, 22] and rate-distortion modeling of packetized content, system performance analysis of transmission resource allocation, and edge-based delivery [23–26], for traditional video applications.

3 SYSTEM FRAMEWORK

3.1 6DOF Volumetric VR

We characterize the 6DOF VR content as a collection of collocated 360° viewpoints (see Figure 2, left) that a user can dynamically navigate over the area of the arena. A single 360° video viewpoint enables three degrees of navigation freedom (3DOF) to a VR user, in the form of rotational head movements around three orthogonal axes (see Figure 2, right), to experience the remote scene immersively from a single location. We extend this concept to the case where the immersion also includes spatial movement of the user in the remote scene, to have a 6DOF immersion experience. The content for each such 360° video viewpoint available to be navigated can be represented using the scalable 360° tiling approach (Section 3.3.1) and the edge server can apply dynamic viewpoint and viewport-driven content adaptation jointly in this case, in response to the 6DOF navigation actions of the user.

3.2 Dual-Connectivity Framework

We consider that there are N_u mobile VR users $U = \{1, 2, ..., N_u\}$ in our system (see Figure 1) navigating the 6DOF VR content in an indoor arena. We divide its spatial area into N_x equal size cells. An xGen transmitter $x \in X$, where $X = \{1, 2, ..., N_x\}$ is installed on the ceiling above the center of each cell. The edge server is linked to the xGen transmitters and a WiFi Access Point (AP). The maximum data transmission rate of each xGen transmitter is C^x and the maximum capacity of the WiFi link is C^w. The VR headsets are dual-connectivity enabled and equipped with a WiFi and an xGen transceiver. Uplink communication between the headset and

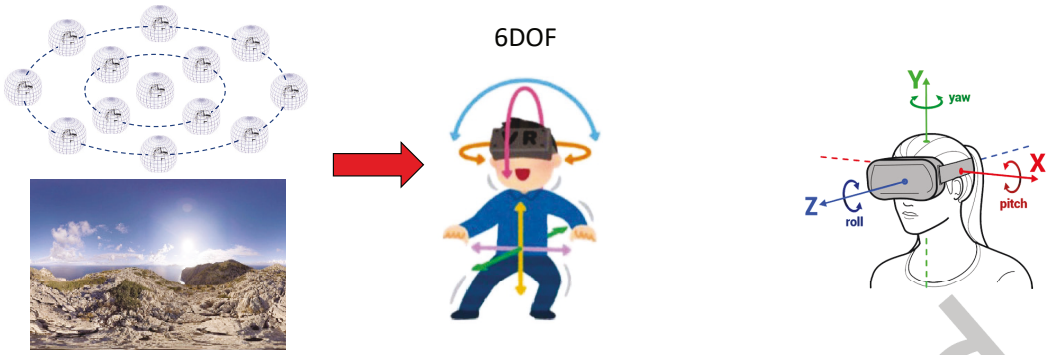


Fig. 2. (left) Multi-viewpoint 360° video 6DOF VR representation of a remote scene, and (right) 3DOF VR head navigation.

the server is carried out via WiFi, to share control information. The server synergistically coordinates the WiFi uplink/downlink and xGen downlink transmissions.

The xGen transmitters in our system can be either FSO or mmWave transmitters. We consider LDs as FSO transmitters. As the beamwidth of an LD is very narrow (typically $< 1°$), a combination of RGB LDs with a diffuser is used to obtain larger beamwidth and increase the FSO communication coverage [27]. Similarly, a mmWave transmitter comprises multiple antenna elements to form a phased antenna array (PAA). The current driving the PAA is split/combined using a splitter/combiner module into each antenna path. The current at each antenna path is phase shifted to steer the mmWave beam towards a target position . Accurate tracking of the 6DOF body and head movements of the users is enabled via two infrared (IR) base stations mounted on the arena walls, and built-in internal-measurement-units (IMUs) and IR sensors on the users' VR headsets. Thanks to the 6DOF information, the edge server identifies the 360° content experienced by the user (viewport), which is defined by the orientation of the VR headset. The edge server partitions the 360° video into two embedded representations: a baseline representation of the entire 360° panorama, and a viewport-specific enhancement representation (see Fig. 3). The server dynamically adapts the two representations to the available transmission rates of the two parallel links. For efficient utilization of the high capacity of the xGen links and high computation capability of the server, a portion of the viewport-specific enhancement representation may be decoded at the server and streamed as raw data, and the remaining portion is streamed as compressed data.

The baseline representation is streamed over WiFi and the enhancement representation is streamed over an xGen link. The viewport-specific content from the two representations is then integrated at the user headset to enable high-fidelity 360° remote VR immersion. We provide a detailed description of the modeling of the different components of our system below.

3.3 Edge server modeling

The edge server is equipped with a graphics processing unit (GPU) for processing high fidelity 360° videos before streaming them to the VR users [28–30]. We describe the server's operation below in detail.

3.3.1 Scalable multi-layer 360° tiling. The server leverages scalable multi-layer 360° video viewpoint tiling design that integrates with the WiFi-xGen dual-connectivity streaming. It partitions each panoramic 360° video frame into a set of tiles $M = \{1, 2, ..., N_M\}$. We denote a block of consecutive 360° video frames compressed together with no reference to other frames, as a group of pictures (GOP). The set of tiles at the same spatial location (i, j) in a GOP is denoted as a GOP-tile m_{ij}. Using the scalable extension of the latest video compression standard (SHVC) [31], the server constructs L embedded layers of increased immersion fidelity l_{ij} for each GOP-tile. The first layer of a compressed GOP-tile is known as the base layer, and the remaining layers are denoted as enhancement

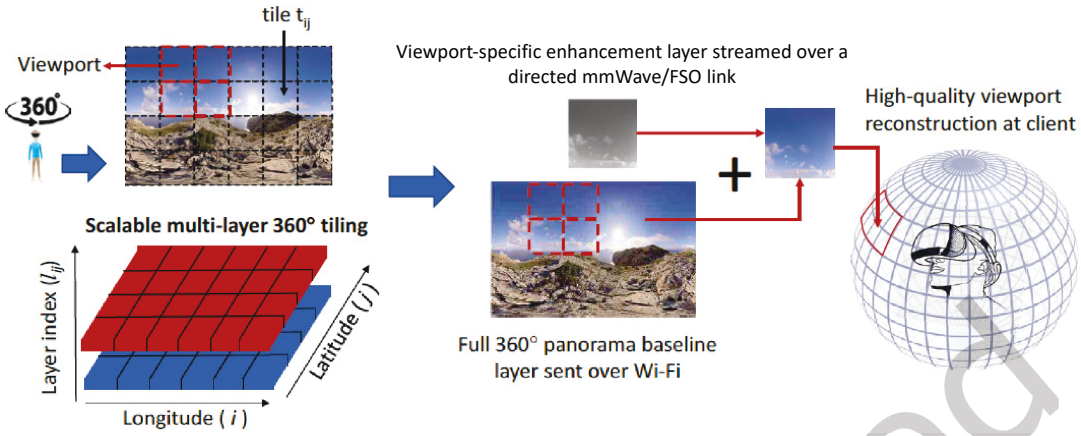


Fig. 3. A user's 360° viewpoint is represented as two embedded layers using scalable 360° tiling. The base layer of the entire 360° panorama is streamed over WiFi. Viewport-specific enhancement layer tiles are sent over a directional mmWave/FSO link. The viewport tiles from the two layers are then integrated at the user to enable high-fidelity immersion.

layers. The reconstruction fidelity of a GOP-tile improves incrementally as more layers are decoded progressively starting from the base layer.

The server constructs a baseline representation of the entire 360° panorama by combining the first n_b embedded layers for each GOP-tile. The induced data rate associated with the baseline representation of a tile $m_{ij} \in M$ is denoted as $R_{ij,w}$. Similarly, the server constructs an enhancement representation by combining the subsequent n_e embedded layers for each GOP-tile comprising the user viewport. The induced data rate associated with the enhancement representation of a tile $m_{ij} \in M_u$ is $R_{ij,x}$. Here, $M_u \subset M$ denotes the subset of GOP-tiles encompassing the user viewport. We formally define this subset as $M_u = \{m_{ij} \in M_u | p^u_{ij} > 0\}$, where p^u_{ij} denotes the probability that user u accesses tile m_{ij} during navigation of the GOP. The minimum and maximum encoding rates for tile m_{ij} available at the server are $R_{ij,\min}$ and $R_{ij,\max}$.

3.3.2 Tile navigation likelihoods. Based on uplinked navigation information, the edge server can develop a set of probabilities $\{p^u_{ij}\}$ that capture how likely user u is to access each GOP tile m_{ij} comprising the 360° panorama associated with her present 360° video viewpoint in the 6DOF content. We leverage our recent advances [32] to enable the server to build this information and benefit our analysis and optimization of the resource allocation carried out by the server.

3.3.3 GOP-tile decoding at the server. As noted in Section 1, we leverage the the multi-Gbps data rate of the xGen link to send some parts of the content as raw data, to alleviate the client from decoding the content prior to rendering on its end. This will reduce the overall end-to-end latency. To enable this, the tiles selected to be sent as raw data would need to be decompressed first at the edge server.

As noted above, the server can identify the present viewport of user $u \in U$ comprising a subset of GOP-tiles $M_u \subset M$. Among these $|M_u|$ GOP-tiles, a subset of GOP-tiles $M^r_u \subset M_u$ is decoded at the server. Each of these $|M_u|$ tiles is decoded from its highest available data rate $R_{ij,max}$ at the server. The decoding speed of the server is Z and a user $u \in U$ is allocated a speed of $Z_u \leq Z$. Thus, the time delay in decoding the user viewport is $\tau^Z_u = \frac{\sum_{ij \in M^r_u} R_{ij,\max} \Delta T}{Z_u}$. Here, ΔT is the playback duration of a GOP. The size of each decoded GOP-tile is E_r. The ability to transmit raw GOP tiles will provide further performance trade-offs that can be leveraged in our analysis and optimization.

3.3.4 WiFi-xGen dual-connectivity streaming. The server streams the baseline representation of all GOP-tiles to a user over a WiFi link. Each user $u \in U$ is allocated a maximum WiFi data rate of C_u^w and $\sum_{u \in U} C_u^w \leq C^w$. We formulate the delay of streaming the baseline representation of the entire 360° panorama to user u as $\tau_u^w = \frac{\sum_{ij} R_{ij,w} \Delta T}{C_u^w}$.

The server streams to u the $|M_u^r|$ raw GOP-tiles and the enhancement representation of the rest of the GOP-tiles $M_u^e = M_u \setminus M_u^r$ over a directed xGen link. Each user $u \in U_x$ associated with an xGen transmitter is allocated a maximum data rate of C_u^x and $\sum_{u \in U_x} C_u^x \leq C^x$. Here, U_x denotes the set of users associated with x. Thus, we formulate the time delay of streaming the M_u^r raw GOP-tiles and the enhancement representation of the M_u^e tiles over a directed xGen link to user $u \in U_x$ as $\tau_u^x = \frac{|M_u^r| E_r + \sum_{ij \in M_u^e} R_{ij,x} \Delta T}{C_u^x}$.

3.4 User headset modeling

3.4.1 Transceivers for the headset. Each user headset is equipped with a WiFi and an xGen transceiver. For a VR-arena with FSO transmitters, we consider a FSO transceiver on the headset as the xGen transceiver. For a VR-arena with mmWave transmitters, the xGen transceiver on the headset is a mmWave transceiver.

Our FSO-enabled headset design integrates multiple photo-detectors (PDs) placed on hemispherical layers [18]. The number of PDs that can be placed on the headset depends on the size of both the PDs and the headset. Our mmWave receiver comprises 16 antenna elements enabling it to perform analog beamforming both in the horizontal and vertical directions. The dimensions of the receiver is 20mm×5mm and can be easily integrated on top of the VR headset [9].

3.4.2 Decoding and rendering. The headset is also equipped with a mobile GPU for decompressing and rendering the received 360° video to be displayed to the user [33]. Before rendering the viewport on the user's headset, the received compressed content would need to be decompressed or decoded first. The maximum decoding speed of the headset is $z_u \geq z_u^w + z_u^x$, where z_u^w is the speed allocated for decoding the GOP-tiles (baseline representation) received over the WiFi link and z_u^x is the speed allocated for decoding the GOP-tiles (enhancement representation) received over an xGen link. Hence, the time delay in decoding the baseline representation of all M GOP-tiles is $\tau_u^{z,w} = \frac{\sum_{ij} R_{ij,w} \Delta T}{z_u^w}$ and the delay in decoding the enhancement representation of M_u^e GOP-tiles is $\tau_u^{z,x} = \frac{\sum_{ij \in M_u^e} R_{ij,x} \Delta T}{z_u^x}$.

The processing capability of the headset for rendering the viewport is $r_u \geq r_u^w + r_u^x$, where r_u^w is the processing power allocated for rendering the baseline representation of the viewport and r_u^x is the processing power allocated for rendering the combined baseline and enhancement representation of the viewport. Thus, the time delay in rendering the viewport at baseline quality is $\tau_u^{r,w} = \frac{E_v}{r_u^w b_h}$ and at enhanced quality is $\tau_u^{r,x} = \frac{E_v}{r_u^x b_h}$. Here, E_v is the size of the viewport after decoding and b_h is the computed data volume per CPU cycle on the headset.

3.5 User viewport reconstruction distortion

We leverage our recent modeling advances [12] to accurately characterize the reconstruction distortion of a VR user's 360° viewport on her headset as:

$$D_u = \sum_{ij \in M_u^r} p_{ij}^u a_{ij} R_{ij,\max}^{b_{ij}} + \sum_{ij \in M_u^e} p_{ij}^u a_{ij} \left(R_{ij,x} + R_{ij,w} \right)^{b_{ij}},$$

where a_{ij} and b_{ij} are parameters of the model. The modeling above will benefit our problem analysis and optimization framework that are described next.

4 PROBLEM FORMULATION

Our objective is to minimize the aggregate reconstruction distortion of the delivered content experienced by all the users, given the WiFi and xGen link capacities, computing capability of the server and the VR headsets, and

system latency constraints. We formulate our optimization problem of interest as:

$$\min_{\substack{\{M_u^r\},\ \{R_{ij,x}\},\ \{r_u^x\},\ \{z_u^x\},\\ \{Z_u\},\ \{R_{ij,w}\},\ \{r_u^w\},\ \{z_u^w\}}} \sum_x \sum_{u\in U_x} D_u, \tag{1}$$

$$\text{s.t.} \quad \tau_u^w + \tau_u^{z,w} + \tau_u^{r,w} \le \Delta T, \quad u \in U, \tag{2}$$

$$\tau_u^Z + \tau_u^x + \tau_u^{z,x} + \tau_u^{r,x} \le \Delta T, \quad u \in U, \tag{3}$$

$$R_{ij,w} \in [R_{ij,\min}, R_{ij,\max}],\ R_{ij,x} \le R_{ij,\max} - R_{ij,w}, \tag{4}$$

$$\sum_{u\in U} Z_u \le Z, \quad \sum_{u\in U} C_u^w \le C^w, \quad \sum_{u\in U_x} C_u^x \le C^x, \tag{5}$$

$$r_u^w + r_u^x \le r_u, \quad z_u^w + z_u^x \le z_u, \quad \forall u \in U. \tag{6}$$

The constraint in (2) imposes that the total time required to stream the baseline representation of all the tiles from the server to the user over the WiFi link, decode them on the headset, and render the viewport must not exceed ΔT. The constraint in (3) imposes that the total time required to decode $|M_u^r| \ge 0$ tiles on the server, stream these raw tiles and rest of the compressed viewport tiles to the user, decode the compressed tiles on the headset, and render the viewport must not exceed ΔT. The constraint in (4) imposes that the encoding rate for the baseline representation of a GOP-tile must not be less than $R_{ij,min}$ and must not exceed $R_{ij,max}$. It also imposes that the encoding rate of the enhancement representation of a GOP-tile must not exceed $R_{ij,max} - R_{ij,w}$. The constraint in (5) indicates that the total decoding speed of the server allocated to the users is bounded by Z, and the WiFi and xGen resource allocations must not exceed C^w and C^x respectively. The constraint in (6) indicates that the decoding speed of the headset is bounded by z_u and the rendering capability is bounded by r_u.

We set the decoding resources of the server and the WiFi channel data rate to be equally allocated to all users, for fairness. Hence, each user is assigned a decoding speed of $Z_u = Z/N_u$ and a maximum data rate of $C_u^w = C^w/N_u$. Similarly, we set the maximum data rate of each user assigned to xGen transmitter x as $C_u^x = C^x/N_x$. These developments then allow us to decouple (1) into individual subproblems for every user-transmitter pair. We formulate each such sub-problem for user u assigned to xGen transmitter x as

$$\min_{M_u^r,\ \{R_{ij,x}\},\ r_u^w,\ z_u^w,\ \{R_{ij,w}\},\ r_u^x,\ z_u^x} D_u, \tag{7}$$

$$\text{s.t.} \quad (2), (3), (4), \text{ and } (6).$$

The problem in (7) is mixed-integer programming, which is hard to solve optimally in practice. The optimal solution can be achieved via an exhaustive search, which requires searching over all sets $M_u^r \subset M_u$, and then for each such candidate set, finding the optimal streaming data rates for the baseline and enhancement representations, and the user's headset decoding speed and rendering capability allocations. Hence, we propose a lower complexity approach to solve (7), where we first sort the GOP-tiles in the viewport in descending order of their distortion derivative weighted navigation likelihoods. We represent this sorted set of tiles as M_u^s. We then search over $|M_u^s| + 1$ possibilities for M_u^r constructed effectively from M_u^s, instead of carrying out an exhaustive search. We have verified empirically that our strategy captures the optimal solution with high probability.

We present an outline of the proposed approach here. We first construct the set M_u^s as explained above. For each $k \in \{0, 1, ..., |M_u^s|\}$, we construct a candidate set $M_{u,k}^r$ of viewport tiles to be transmitted as raw data over the associated xGen link such that such that $M_{u,k}^r$ comprises the first k tiles from M_u^s. We note here that the set $M_{u,k}^r$ will be empty ($\varnothing$) for the case $k = 0$. Then, all enhancement representation tiles $m_{ij} \in M_u$ will be transmitted as compressed data over the xGen link, and each tile will comprise $n_e(i, j)$ embedded enhancement layers from the scalable 360° tiling, as introduced in Section 3.3. For each $M_{u,k}^r$, we find the streaming data rates

$\{R^{\star}_{ij,x,k}\}$ and $\{R^{\star}_{ij,w,k}\}$ associated with the baseline and enhancement representations, and the user's headset decoding speed allocations $\{z^{x\star}_{u,k}\}$ and $\{z^{w\star}_{u,k}\}$, and rendering speed allocations $\{r^{x\star}_{u,k}\}$ and $\{z^{w\star}_{u,k}\}$, for which the reconstruction distortion $D^{\star}_{u,k}$ is minimum. Finally, we select the value $k^{\star}$ for which for which $D^{\star}_{u,k}$ is the lowest and this completes the solution to (7). We describe our proposed approach in more detail in the following section.

5 OPTIMIZATION FRAMEWORK

When the selection of GOP-tiles to be streamed in raw format is fixed, i.e., for a given value of k and $M^{\mathrm{r}}_{u,k}$, we can reformulate the problem in (7) as

$$\min_{\substack{\{R_{ij,x,k}\},\ r^{w}_{u,k},\ z^{w}_{u,k},\\ \{R_{ij,w,k}\},\ r^{x}_{u,k},\ z^{x}_{u,k}}} D_{u,k}, \tag{8}$$
$$\text{s.t.} \quad (2), (3), (4), \text{ and } (6).$$

The problem in (8) can be solved optimally by converting it to geometric programming (GP) first. To do so, we first introduce an auxiliary variable $R_{ij,xw} = R_{ij,x} + R_{ij,w}$, where $ij \in M^{e}_{u,k}$, $u \in U_x$. Moreover, we note that once M^{r}_{u} is fixed, its contribution to D_u, as captured by the first sum in the respective expression (see Section 3.5), will be fixed as well. Thus, in the following, we focus on the second sum in the expression for D_u that captures the impact of $R_{ij,xw}$, the remaining variables in the objective function in (8).

Concretely, we rewrite the optimization problem in (8) as:

$$\min_{\substack{\{R_{ij,xw,k}\},\ r^{w}_{u,k},\ z^{w}_{u,k},\\ \{R_{ij,w,k}\}\ r^{x}_{u,k},\ z^{x}_{u,k}}} D^{xw}_{u,k}, \tag{9}$$
$$\text{s.t.} \quad (2) \text{ and } (6),$$
$$\tau^{Z}_{u} + \tau^{xw}_{u} + \tau^{z,xw}_{u} + \tau^{r,x}_{u} \leq \Delta T, \tag{10}$$
$$R_{ij,\min} \leq R_{ij,w} \leq R_{ij,\max},$$
$$R_{ij,w} \leq R_{ij,xw} \leq R_{ij,\max}, \tag{11}$$

where $D^{xw}_{u,k} = \sum_{ij \in M^{e}_{u}} p^{u}_{ij} a_{ij} (R_{ij,xw,k})^{b_{ij}}$, $\tau^{xw}_{u} = |M^{r}_{u,k}| E_r + \sum_{ij \in M^{e}_{u,k}} (R_{ij,xw} - R_{ij,w}) \Delta T / C^{x}_{u}$ and $\tau^{z,xw}_{u} = \sum_{ij \in M^{e}_{u,k}} (R_{ij,xw} - R_{ij,w}) \Delta T / z^{x}_{u}$.

We can convert the problem in (9) to GP using the single condensation method [34]. In particular, according to this method, for a constraint which is a ratio of posynomials, the denominator posynomial can be approximated into a monomial. We formulate an iterative method towards this objective. At each iteration t, we convert the constraints (2) and (10) into respective posynomial functions. The posynomial function for (10) can be expressed as follows.

$$\left(\frac{\Delta T Z_u C_u^x z_u^x r_u^x b_h}{\delta_1(t)}\right)^{-\delta_1(t)} \cdot \left(\frac{\Delta T Z_u z_u^x r_u^x b_h \sum_{ij\in M_{u,k}^e} R_{ij,w}}{\delta_2(t)}\right)^{-\delta_2(t)} \cdot \left(\frac{\Delta T Z_u C_u^x r_u^x b_h \sum_{ij\in M_{u,k}^e} R_{ij,w}}{\delta_2(t)}\right)^{-\delta_3(t)} \cdot (C_u^x z_u^x r_u^x b_h \Delta T \sum_{ij\in M_{u,k}^r} R_{ij,max} + Z_u z_u^x r_u^x b_h (|M_{u,k}^r| E_r + \Delta T \sum_{ij\in M_{u,k}^e} R_{ij,xw}) + Z_u C_u^x r_u^x b_h \Delta T \sum_{ij\in M_{u,k}^e} R_{ij,xw} + Z_u C_u^x z_u^x E_v) \le 1, \quad (12)$$

where $\delta_1(t) = c_1/(c_1+c_2+c_3)$, $\delta_2(t) = c_2/(c_1+c_2+c_3)$, and $\delta_3(t) = c_3/(c_1+c_2+c_3)$. Here, $c_1 = \Delta T Z_u C_u^x z_u^x(t-1) r_u^x(t-1) b_h$, $c_2 = Z_u z_u^x(t-1) r_u^x(t-1) b_h \sum_{ij\in M_{u,k}^e} R_{ij,w}$, and $c_3 = Z_u C_u^x r_u^x(t-1) b_h \sum_{ij\in M_{u,k}^e} R_{ij,w}$. Let $D_k(t) = \sum_{ij\in M_{u,k}^e} p_{ij}^u a_{ij} (R_{ij,xw}(t))^{b_{ij}}$. Then, the optimization problem to be solved at iteration t is:

$$\min_{\substack{\{R_{ij,xw,k}\}, \\ \{R_{ij,w,k}\}}} \; \min_{\substack{r_{u,k}^w, \\ r_{u,k}^x}} \; \min_{\substack{z_{u,k}^w, \\ z_{u,k}^x}} D_{u,k}^{xw}(t), \quad (13)$$
$$\text{s.t.} \quad (2), (6), (10), (11).$$

Here, (13) is a GP problem and we can solve it optimally. We carry out the optimization iteratively until $|D_{u,k}^{xw}(t) - D_{u,k}^{xw}(t-1)| \le \epsilon$, for some small $\epsilon \ge 0$. When this condition is met, we obtain the optimal value of the objective function in (13) as $D_{u,k}^\star = \sum_{ij\in M_{u,k}^r} p_{ij}^u a_{ij} R_{ij,max}^{b_{ij}} + D_{u,k}^{xw}(t)$, the optimal streaming data rate $\{R_{ij,xw,k}^\star\} = \{R_{ij,xw,k}(t)\}$, the optimal headset decoding speed allocations $z_{u,k}^{w\star} = z_{u,k}^w(t)$ and $z_{u,k}^{x\star} = z_{u,k}^x(t)$, and rendering capability allocations $r_{u,k}^{w\star} = r_{u,k}^w(t)$ and $r_{u,k}^{x\star} = r_{u,k}^x(t)$, for a given value of k. This completes the solution to (8).

Finally, we obtain the overall solution that includes the optimal choice of M_u^r by finding the k and $M_{u,k}^r$ that result in the smallest $D_{u,k}^\star$. We formally write this optimization as:

$$D_u^{\text{opt}} = \min D_{u,k}^\star, \quad (14)$$
$$M_u^{r,\text{opt}}, \; z_w^{\text{opt}}, \; r_w^{\text{opt}}, \; \{R_{ij,xw}^{\text{opt}}\}, \; z_x^{\text{opt}}, \; r_x^{\text{opt}} = \arg\min_{\substack{M_{u,k}^r, \\ \{R_{ij,xw,k}^\star\}}} \; \min_{\substack{z_{u,k}^{w\star}, \\ z_{u,k}^{x\star}}} \; \min_{\substack{r_{u,k}^{w\star}, \\ r_{u,k}^{x\star}}} D_{u,k}^\star.$$

This completes the solution to the problem in (7). A formal description of the proposed method to determine the optimum set of raw and compressed GoP-tiles within the viewport, data rates of associated with these GoP-tiles, and the decoding and rendering resource allocation at the headset is provided in Algorithm (1), which is activated every ΔT period of time.

We assess the computational complexity of Algorithm 1 as follows. For a given set of tiles $M_{u,k}^r$, the number of required iterations for Algorithm 1 is $\frac{\log(\mathcal{K})/t_{0,1}\epsilon}{\log \xi}$, where $\mathcal{K}$ is the number of constraints in (13), $t_{0,1}$ is the initial point to approximate the accuracy of the interior point method used to solve (13), $0 < \epsilon < 1$ is the stopping

Algorithm 1 Geometric Programming Solution to (7)

[1] $D = \{\}$ Sort GoP-tiles comprising the viewport in descending order using p^u_{ij}. k = 1 to $|M^s_u|$ Set $t = 1$, $r^x_u(t) = r^w_u(t) = r_u/2$, $z^x_u(t) = z^w_u(t) = z_u/2$, initialize $R_{ij,xw}(t)$. true $t = t + 1$ Determine the optimum $r^w_u(t)$, $r^x_u(t)$, $z^w_u(t)$, $z^x_u(t)$, $R_{ij,xw}(t)$, $D_k(t)$ solving (13) using GGPLAB. $|D_k(t) - D_k(t-1)| \leq \epsilon$ Break

$D^\star_k = \sum_{ij \in M^r_{u,k}} p^u_{ij} a_{ij} R^{b_{ij}}_{ij,max} + D_k(t)$ $D = D \cup D^\star_k$ $D^\star_u = \min(D)$ $M^{opt}_r, \mathbf{R^{opt}_{xw}}, z^{opt}_w, z^{opt}_x, r^{opt}_w, r^{opt}_x = \arg\min D^\star_k$

criterion for the interior point method, and ξ is used for updating the accuracy of the interior point method [35]. This inner optimization is carried out in Algorithm 1 $|M^s_u|$ times, for different sets $M^r_{u,k}$. Thus, the overall computational complexity of Algorithm 1 is on the order of $|M^s_u| \frac{\log(\mathcal{K})/t_{0,1}\epsilon}{\log \xi}$.

6 XGEN CHANNEL AND RECEIVER MODELING

6.1 FSO channel

We formulate a point-to-point line-of-sight (LOS) intensity modulation direct detection (IM/DD) FSO channel model. The optical beam generated by combining light from the RGB LDs and using a diffuser follows the Lambertian radiation pattern [36]. We use the on-off keying (OOK) modulation to modulate the transmit binary signal. For a headset with N_{pd} PDs and an arena with N_x FSO transmitters, we formulate the channel gain between a FSO transmitter $x \in \{1, 2, ..., N_x\}$ and a PD $p \in \{1, 2, ..., N_{pd}\}$ as: $H_{xp} = \frac{(q_l+1)A_{pd}}{2\pi d^2_{xp}} \xi(\phi_{xp}, \psi_{xp})$, if $0 \leq \psi_{xp} \leq \beta$, and $H_{xp} = 0$, if $\psi_{xp} > \beta$, where $\xi(\phi_{xp}, \psi_{xp}) = \cos^{q_l}(\phi_{xp}) T_s(\psi_{xp}) g(\psi_{xp}) \cos(\psi_{xp})$. Moreover, $q_l = \log(2)/\log(\cos(\theta_{div}))$ is the Lambertian emission order, A_{pd} is the area of a PD, ϕ_{xp} is the irradiance angle, ψ_{xp} is the incident angle, and $T_x(\psi_{xp})$ is the optical filter gain. Here, $g(\psi_{xp})$ and β are the optical concentrator gain and field-of-view (FOV) respectively, and $g(\psi_{xp}) = \eta^2/\sin^2(\beta)$, where η is the refractive index.

We formulate the noise power at a PD as $\kappa^2 N_0 B_L$, where κ is the ratio of DC optical power to the square root of electrical signal power, N_0 is the noise power spectral density, which follows a Gaussian distribution, and B_L denotes the baseband modulation bandwidth [37]. The signal received by a user from a transmitter with which he is not associated, is regarded as interference. We combine the FSO signals received across all PDs on a headset to produce a higher quality aggregate resulting signal. We adopt maximal-ratio-combining (MRC) as a diversity combining technique to integrate the incoming optical signals across all PDS, with the maximum possible benefit [38]. In MRC, the output signal of each PD is multiplied by a weight $\lambda_{xp} = (R_{PD} H_{xp} P_x)^2 / \left(\kappa^2 N_0 B_L + \sum^{N_x}_{x'=1, x' \neq x} (R_{pd} H_{xp} P_x)^2\right)$, where P_x is the transmitted optical power of the transmitter and R_{pd} is the responsivity of the PD. We formulate the signal-to-interference-and-noise ratio (SINR) associated with our MRC technique as: $SINR^{MRC}_x = (\sum^{N_{pd}}_{p=1} \lambda_{xp} R_{pd} H_{xp} P_x)^2 / \Xi$, where $\Xi = \sum^{N_{pd}}_{p=1} \left(\sum^{N_{xr}}_{x'=1, x' \neq x} (\lambda_{xp} H_{xp} P_x)^2 + \lambda^2_{xp} \kappa^2 N_0 B_L\right)$. Finally, we model the maximum throughput of each VR user as: $T_r = B_L \log_2(1 + SINR^{MRC}_x)/N^x_u$, where N^x_u is the number of users assigned to the same FSO transmitter x.

6.2 mmWave channel

We adopt a Quasi-deterministic (Q-D) methodology for modeling the mmWave channel [39]. We characterize the channel using a set of strong rays (D-rays) containing most of the power and a set of relatively weak random rays (R-rays) with predefined distribution. We obtain the D-rays using ray-tracing techniques and model the R-rays in a statistical way. We formulate the power received by a mmWave receiver on a subcarrier k as:

$$P_{mmW} = |H_{umW}|^2 P^{mmW}_x, \tag{15}$$

where P^{mmW}_x is the transmit power of a mmWave transmitter, and H_{umW} is the channel frequency response (CFR) at the subcarrier k. We formulate the CIR as:

$$h_{umW} = \sum_{i=0}^{n-1} \left(A^{(i)} \left((W_{ch}^{i})^{H} W (V_{ch}^{i})^{H} V \right) e^{-j2\pi f t_i} \right), \tag{16}$$

where n is the number of multipath components (MPCs), $A^{(i)}$ is the complex amplitude of the i-th MPC, and W and V denote the receiver and transmitter antenna weight vectors respectively. W_{ch}^{i} and V_{ch}^{i} represent beam steering vectors for angle of arrival and angle of departure respectively, of the i-th MPC. Here, f is the operational frequency, and t_i denotes the delay of the i-th path. The complex amplitude of the i-th MPC is given as: $A^{(i)} = 10^{-PL_i/20} e^{j*ps_i}$, where PL_i and ps_i denote the pathloss and phase shift of the i-th path, respectively. The radiation pattern of the receiver array is given by $Y_{Rx}^{(i)} = (W_{ch}^{i})^{H} W$ and that of the transmitter array is given by $Y_{Tx}^{(i)} = (V_{ch}^{i})^{H} V$. Thus, we re-write (16) as: $h_{umW} = \sum_{i=0}^{n-1} \left(A^{(i)} \left(Y_{Rx}^{(i)} Y_{Tx}^{(i)} \right) e^{-j2\pi f t_i} \right)$.

The bandwidth of the mmWave channel is B_M, which we divide into a number of equally spaced sub-bands. The received power per sub-band is determined using (15). We obtain the total received power at the VR headset over the mmWave link as: $P_{Rx}^{mmW} = \sum_{i=0}^{N_{subband}} (P_{Rx_i} b_{m_i})$, where $N_{subband}$ denotes the number of sub-bands, P_{Rx_i} is the power received for the i-th sub-band, and b_{m_i} is the bandwith of the i-th sub-band. We formulate the SINR between a user and a mmWave transmitter as: $SINR_{mmW} = \frac{P_{Rx}^{mmW}}{N_{mmW} + I_{mmW}}$. Here, N_{mmW} is the noise floor given as $N_{mmW} = 290 \times N_{figure} k_B B_M$, where N_{figure} is the noise figure and k_B is Boltzmann constant. Finally, we formulate the maximum thoughput of each VR user as: $T_R^{mmW} = B_M \log_2(1 + SINR_{mmW})/N_x$.

7 XGEN CONNECTIVITY MAINTENANCE

7.1 Free-Space Optics

We present three different FSO connectivity maintenance methods: electronic steering, mechanical steering, and electro-mechanical steering.

7.1.1 Outline. We first propose two xGen connetivity maintenance methods: electronic steering (ES) and mechanical steering (MS), which have different transmission beamwidth requirements. In ES, transmitters with large beamwidth provide full communication coverage to a mobile VR user via electronic (cell) switching. In MS, transmitters with narrower beamwidth are mechanically steered towards the user to maintain xGen connectivity. We propose two schemes for user-to-transmitter assignment in MS. We also investigate a third connectivity maintenance method (EMS) to integrate the best aspects of ES and MS. We describe these methods below in more detail.

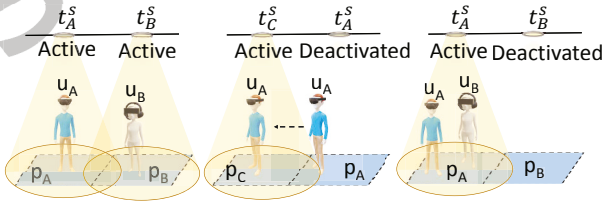


Fig. 4. Electronic steering.

7.1.2 Electronic steering. In this method, a transmitter is assigned to one or more users navigating within its corresponding playing area (cell). As a user moves to an adjacent cell, the server uses the tracking information to assign the transmitter of this cell to him. For example, as shown in Fig. 4, left, when users u_A and u_B are located within cells p_A and p_B respectively, transmitters t_A^s and t_B^s are assigned to them respectively. As u_A moves from p_A to p_C, t_A^s is deactivated and t_B is activated and assigned to him (4, middle). We define this switching of

transmitter assignment for a user as *electronic steering*. Also, as shown in Fig. 4, right, when both u_A and u_B are within p_A, they are both assigned to t_A^s and allocated an equal share of its data rate.

Here, the minimum transmitter beamwidth $\theta_{div^{min}}$ is chosen such that a user with height h_{max} is always covered by at least one of the transmitter beams. For a VR arena with dimensions $l_x \times l_y \times l_z$ and each playing area of size $l'_x \times l'_y$, we select $\theta_{div}^{min} = \tan^{-1} \frac{\sqrt{l'^2_x + l'^2_y}}{2(lz - h_{max})}$.

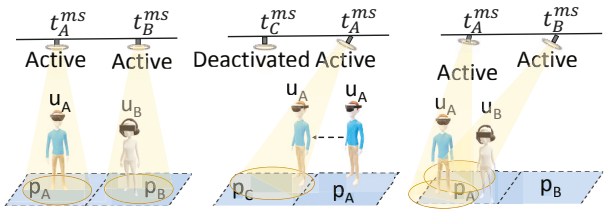


Fig. 5. Mechanical steering.

7.1.3 Mechanical steering. Here, each transmitter is mounted on a mechanically steerable platform. Each transmitter is assigned to only one user during a given time period. The server uses the tracking information to steer a transmitter towards its assigned user to maintain connectivity. Each user is served by a separate transmitter even if multiple users a located within the same cell. For example, as shown in Fig. 5, left, u_A and u_B are initially assigned to t_A^{ms} and t_B^{ms} respectively. As u_A moves from p_A to p_C (Fig. 5, middle), t_A^{ms} is steered to point towards him and maintain connectivity. Even when both u_A and u_B are with p_A, they are served separately by t_A^{ms} and t_B^{ms} respectively (Fig. 5, right). Here, the transmitter beamwidth can be set smaller compared to *ES* as they can mechanically steered to maintain connectivity.

We explore two different user-to-transmitter assignment schemes here, *MS* with fixed assignment (*MSF*) and *MS* with dynamic assignment (*MSD*).

MSF. In this scheme, a transmitter is initially assigned to the user with whom it has the least distance. The transmitter serves the same user for the entire duration of the VR session.

MSD. Here, a transmitter is assigned to a user with whom it has the least distance at the start of the VR session. As the users move within the arena, the server performs a user-to-transmitter re-assignment in a periodic manner based on the signal-to-noise-ratio (SNR) experienced by the users. Let $s_{u,x}$ denote the SNR experienced by user $u \in U$ when he is served by transmitter $x \in X$ and $d_{u,x}$ denote the distance between u and x. A one-to-one mapping exists between $s_{u,x}$ and $d_{u,x}$. At every ΔT time unit, a user-to-transmitter re-assignment is performed such that the $\min(s_{u,x})$ is maximized, or equivalently the $\max(d_{u,x})$ is minimized.

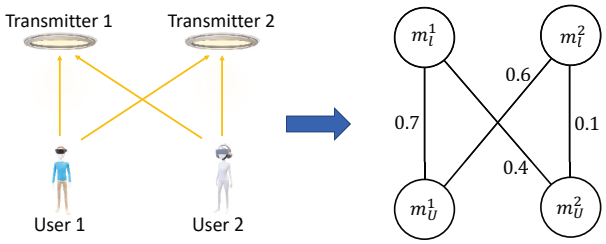


Fig. 6. Bipartite graph example for 2 transmitters and 2 users.

The optimal solution to the user-to-transmitter assignment problem can be obtained via an exhaustive search, which is computationally expensive. Thus, we explore a lower complexity approach to solve the problem

optimally using graph-theoretic concepts. We construct a weighted bipartite graph, where each user $u \in U$ and each transmitter $x \in X$ is represented by vertices $f_u^1 \in \mathcal{F}^1$ and $f_x^2 \in \mathcal{F}^2$ respectively, and the weight of and edge (f_u^1, f_x^2) is $w_{f_u^1, f_x^2} = d_{u,x}$. We show a bipartite graph for two users and two transmitters in Fig. 6 where the two possible maximum matching are $\{(f_x^1, f_u^1), (f_x^2, f_u^2)\}$ and $\{(f_x^1, f_u^2), (f_x^2, f_u^1)\}$.

We can express the user-to-transmitter assignment problem as a bottleneck matching (BM) problem of the graph defined by the maximum matching whose largest edge weight is a small as possible, i.e.,

$$\min_{\pi \in \Pi} \max_{(f_u^1, f_x^2) \in \pi} \omega_{(f_u^1, f_x^2)}, \tag{17}$$

where Π comprises all the possible maximum matching. For the graph in Fig. 6, the bottleneck matching is $\{(f_u^1, f_x^2), (f_u^2, f_x^1)\}$ and the corresponding assignment is: Transmitter 1 is assigned to User 2 and Transmitter 2 is assigned to User 1. We formulate an algorithm to solve the problem in (17) inspired by the approach studied in [40].

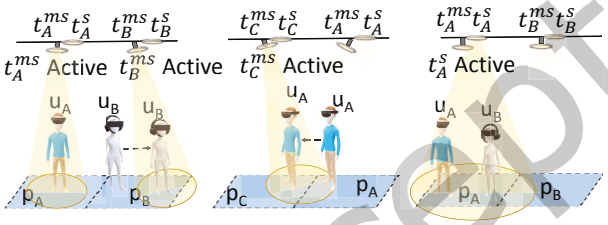


Fig. 7. Electro-mechanical steering.

7.1.4 Electro-mechanical steering. In this scheme, two transmitters are installed on the ceiling at the center of each cell, one stationary and another mechanically steerable. We aim to integrate best aspects of *ES* and *MS* here. In this method, a user is served by a mechanically steerable transmitter as long as he navigates within the corresponding cell and is the sole user in that cell. When more than one user are located within a cell the corresponding stationary transmitter serves them instead of the mechanically steerable one. For example, in Fig. 7, cell p_A has a stationary transmitter t_A^s and a mechanically steerable transmitter t_A^{ms}. Users u_A and u_B are initially located at cells p_A and p_B respectively (7, left). As u_A moves within p_A, t_A^{ms} is mechanically steered to maintain connectivity with him, similar to *MS*. When u_A moves to the adjacent cell p_C, similarly to *ES*, he is switch to a new transmitter t_C^{ms} and t_A^{ms} is deactivated (Fig. 7, middle). Again, similarly to *ES*, when both u_A and u_B move to the same cell p_A, they are served by the stationary transmitter t_A^s (Fig. 7, right).

7.2 mmWave

In the mmWave band, the signal experiences higher path loss compared to the traditional sub-6-GHz band due to freespace, oxygen-absorption and/or penetration losses. This loss can be compensated by employing beamforming, i.e, by steering the antenna elements of the phased array antenna to focus the power towards a specific direction, yielding directional beams. One of the most challenging problems in the mmWave band is to properly select the BF to apply to enable communication between a pair of devices. This process is referred as Beamforming Training (BFT). The BT phase operates using three main concepts: sectors, quasi-omni antenna pattern, and Sector Sweep (SS). First, the sectors correspond to a discretization of the antenna space by using a set of precomputed antenna weight vectors, reducing the number of possible directional beams to try. In our case, we employ 54 sectors. Then, the quasi-omni antenna pattern is the antenna operating mode with the widest beamwidth attainable and thus the one used when the direction of a communicating node is unknown. In our framework, the quasi-omni pattern radiates in an hemispherical way as we neglect radiation below the substrate of the PAA. Finally, SS is the

process in which one node, referred as the responder receives in quasi-omni mode while the other, refered as the initiator sends a sequence of frames transmitting on different directional sectors. The SS phase allows to perform the BFT as the responder measures the quality of each transmitted frames in each sector by the initiator to enable the selection of the best sector. We define two different mmWave connectivity maintenance: mmWave Same Channel (MMWSC) and mmWave Different Channel (MMWDC). It is worth mentioning that for this study, the BFT signaling overhead is not taken into account as we are interested in exploring the upper-limit performance of the proposed dual-connectivity system. In future work, we plan to investigate the impact of BFT overhead on the mmWave communication in the VR arena environment, and to evaluate the impact of novel beam-tracking techniques developed to significantly reduce BFT overhead [41].

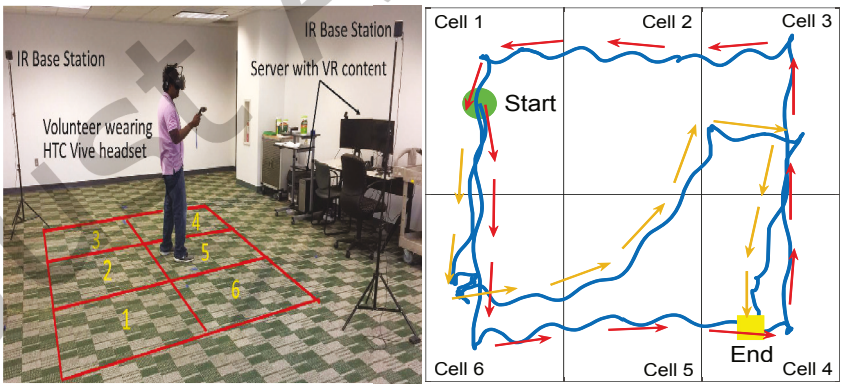
7.2.1 MMWSC. In this scheme, all the mmWave transmitters and receivers, denoted as Access Points (APs) and Stations (STAs) in the IEEE 802.11 terminology, are configured to operate in the same (frequency) channel. Each AP performs a BFT with all the STAs every 4 ms and each STA, based on the BFT results, is associated to the best AP, i.e., the one yielding the highest received power for a given 6DOF position. As all the mmWave transmitters and receivers operate in the same channel, a given AP to STA transmission is considering every other ongoing AP to STA transmission as interference for the computation of its SINR.

7.2.2 MMWDC. In this scheme, each of the APs is configured to operate on a different channel. At the beginning of a simulation run, each STA is associated to the AP yielding the highest receive power and remains associated with this AP, *the selected AP*, for duration of the entire VR session. Thus, the STA performs its BFT only with the selected AP for the entire simulation. It is worth mentioning that for this case, we do not consider any co-channel interference so every AP to STA transmission is interference free.

8 WIFI-XGEN PERFORMANCE EVALUATION

8.1 6DOF VR navigation data capture and setup

Here, we carry out performance evaluation of the proposed system. Our simulation experiments leverage real 6DOF navigation measurements, to incorporate realistic body and head movements comprising VR navigation in the evaluation.



(a) (b)

Fig. 8. (a) 6DOF VR navigation data capture. (b) A sample navigation path (x, y) for 6DOF content *Virtual Museum*.

The 6DOF body and head movement VR navigation measurements were collected with the help of users who were provided with an HTC Vive wireless headset. The measurements were collected in the indoor environmented shown in Figure 8, where the users navigated the 6-DOF VR content *Virtual Museum* [42] across a spatial area

of 6m × 4m, divided into six playing areas (cells) of size 2m × 2m each (height is 3m). We used the software packages SteamVR SDK [43] and Opentrack [44] to record the navigation information for the users in our arena system, as they were being tracked during a session (see Section 3.2). We captured data for three volunteer users individually, across six sessions per user, one for each cell used as the starting navigation point for the user. A total of 30,000 tracking samples were captured per session, at a sampling rate of 250 samples per second. The collected navigation data is publicly shared as part of this publication, to foster further investigations and broader community engagement [45].

8.2 Simulation Setup

8.2.1 WiFi-FSO. We equip the VR arena with six FSO transmitters, each of which is installed on the ceiling above the center of each cell. We set the divergence angle of each stationary FSO transmitter as 51° and that of each mechanically steerable transmitter as 25°. Each user is equipped with a multi-photodetector (PD) VR headset. The headsets comprise 47 PDs with an angular distance of $\Theta_d = 25°$ between two PDs. We set the half-angle field-of-view (FOV) of each PD as $\beta = 0.75\Theta_d$ [18]. The tracking data accuracy is ±1 mm. The system-level results are obtained via a Matlab implementation.

8.2.2 WiFi-mmWave. We equip the VR arena with six mmWave transmitters, one for each cell. Each transmitter and VR headset are equipped with a 16 PAA organized in a rectangular 2×8 configuration to enable beamforming in both azimuth and elevation angles. The mmWave propagation is generated using the open source NIST Quasi-Deterministic channel model implementation [46] that can accurately predict the channel characteristics for mmWave frequencies. The system level results are obtained via an ns-3 IEEE 802.11ad implementation [47, 48].

For each scenario, we assess the viewport immersion fidelity/quality for user u via the Peak Signal to Noise Ratio (Y-PSNR) of its expected viewport distortion experienced by the user over a GOP, computed as $10 \log_{10}(255^2 / \sum_{ij \in M_u} p_{ij}^u D_{ij})$. Benefiting from our earlier advances [12], we accurately model the distortion terms D_{ij} associated with the GOP tiles m_{ij} comprising the present spatial 360° video viewpoint of the user in the 6DOF content. Each such viewpoint comprising the content (see Section 3.1) is scalable encoded at different data rates and 8K-120fps spatial resolution and temporal frame rate. We compute the average Y-PSNR per GOP and across the entire session.

8.3 Results and analysis

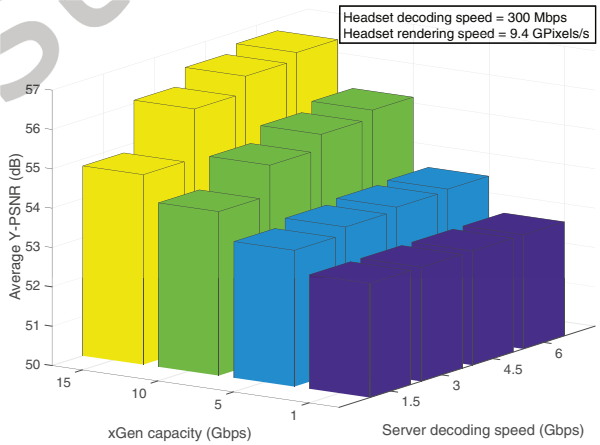


Fig. 9. Impact of xGen channel capacity and server decoding speed on delivered immersion fidelity.

8.3.1 Impact of xGen channel capacitiy and server decoding speed. In Fig. 9, we show how the available xGen link rate and the decoding capability of the edge server affects the delivered immersion fidelity. We can see that when the xGen channel capacity is small (1 Gbps or 5 Gbps), increasing the server decoding speed does not improve the Y-PSNR. At such link capacities, all the GOP-tiles comprising the enhancement representation of the user viewport are streamed in compressed format. No GOP-tile is decoded at the server to be streamed as raw data. Thus, the server decoding speed does not impact the delivered immersion fidelity in such cases. As the xGen link capacity increases, the number of GoP-tiles to be decoded at the server and streamed as raw data increases, and the GOP-tiles to be streamed as compressed data are encoded at higher rates. Thus, the immersion fidelity improves with an increase in the server decoding speed. For example, when the xGen channel capacitiy is 10 Gbps, the Y-PSNR increases from 54.17 dB to 55.51 dB as the server decoding speed increases from 1.5 Gbps to 6 Gbps. Similarly, when the channel capacity is 15 Gbps, a Y-PSNR gain of 1.86 dB is obtained for a server decoding speed of 6 Gbps compared to that of 1.5 Gbps.

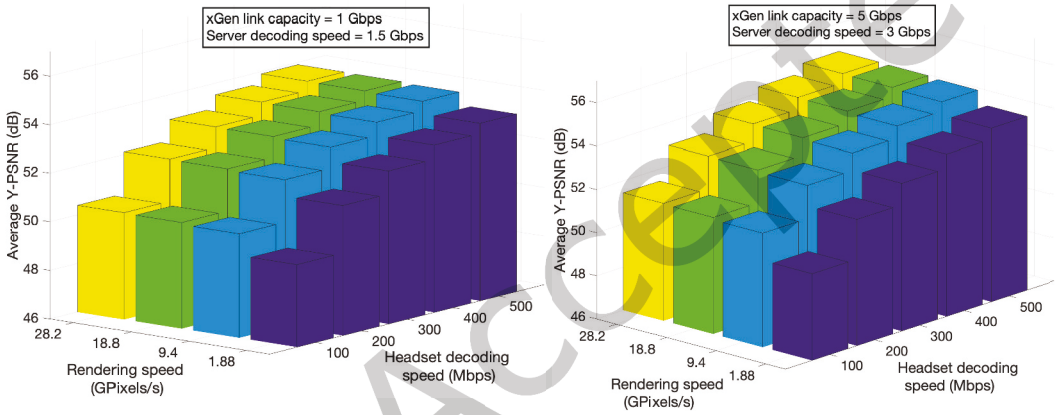
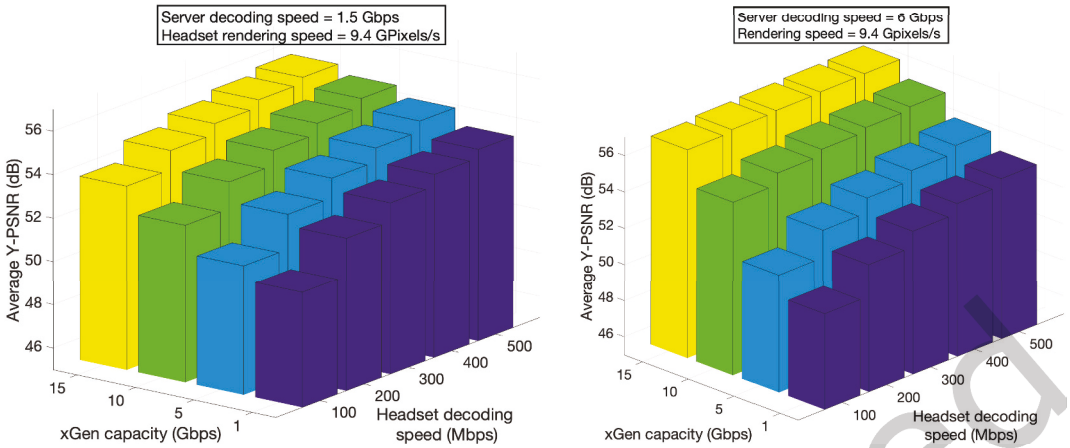


Fig. 10. Immersion fidelity for different user headset computing capability. (a) xGen capacity 1 Gbps, server decoding speed 1.5 Gbps. (b) xGen capacity 5 Gbps, server decoding speed 3 Gbps.

8.3.2 Impact of headset computing capability. We show the the impact of the computing capability of the user VR headset on the delivered immersion fidelity in Fig. 10 for different xGen capacities and server decoding speeds. We can see that the Y-PSNR increases ≈ 1 dB as the rendering speed increases from 1.88 Gpixels/s to 9.4 Gpixels/s. Further increasing the rendering speed does not impact the delivered immersion fidelity. But we can observe from Fig. 10a that when the xGen capacity is 1 Gbps and the server decoding speed is 1.5 Gbps, an increase in the decoding speed of the headset improves the Y-PSNR. For any value of the rendering speed, a gain of ≈ 3.5-4 dB is obtained as the headset decoding speed is increased from 100 Mbps to 500 Mbps. Similarly, for an xGen capacity of 5 Gbps and server decoding speed of 3 Gbps, a 3.3-4 dB Y-PSNR gain is achieved by increasing the headset decoding speed from 100 Mbps to 500 Mbps. As the decoding speed of the headset increases, it can decode GOP-tiles encoded at higher data rates, hence the improvement in the delivered immersion fidelity.

8.3.3 Impact of headset decoding speed and xGen capacity. In Fig. 11, we show the combined effect of the xGen link capacity and the headset decoding capability on the delivered immersion fidelity. We can see that the Y-PSNR imroves significantly as the xGen capacity and the headset decoding speed increases. For example, the Y-PSNR increases more than 3 dB when the xGen capacity increases from 1 Gbps to 15 Gbps for headset decoding speed of 100 Mbps (Fig. 11a). Similarly, when xGen capacity is 1 Gbps, the Y-PSNR increases from 50.28 dB to 53.83 dB as



(a) Server decoding speed = 1.5 Gbps (b) Server decoding speed = 6 Gbps

Fig. 11. Immersion fidelity for different headset decoding speeds and xGen capacities.

the headset decoding speed increases from 100 Mbps to 500 Mbps. Moreover, a Y-PSNR gain of ≈ 5 dB is achieved when the link capacity is increased from 1 Gbps to 15 Gbps and the headset decoding speed is increased from 100 Mbps to 500 Mbps. Increasing the value of either of these parameters allows GOP-tiles encoded at higher rates to be streamed from the server and decoded in time on the user headset, thus improving the Y-PSNR significantly.

We can additionally observe from Fig. 11b that when the server decoding speed is 6 Gbps, and the xGen capacity is 10Gbps/15 Gbps, the improvement in the delivered immersion fidelity with an increase in the headset decoding speed is very small. At such high link capacities and server decoding speeds, most of the GOP-tiles are decoded at the server and streamed as raw data. Thus, the number of GOP-tiles streamed as raw data is small. Hence, the headset has to decode a very small number of GOP-tiles, and increasing its decoding speed does not improve the delivered immersion fidelity significantly.

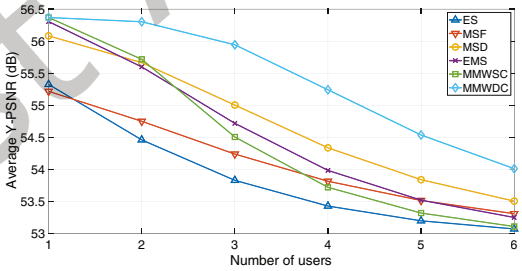


Fig. 12. Immersion fidelity for ES, MSF, MSD, EMS, MMWSC, and MMWDC.

8.3.4 Impact of number of users and connecitivity maintenance methods. We can see in Figure 12 that the immersion fidelity decreases, as expected, across all connectivity maintenance methods and both dual-connectivity systems, as the number of simultaneous VR users in the arena is increased. The first reason is that the WiFi channel data rate and the server's encoding speed are equally allocated to the users in the arena. Moreover, the probability of multiple users being located within the same cell increases, as the number of simultaneous users increases. Thus, the throughput per user decreases when the transmitter's data rate need to be shared among several users. Moreover, the server's comupting resources are also equally allocated to the users when there are more than one

user in the arena. Hence, the average Y-PSNR decreases as the number of users increases. Similary, for *MSF* and *MSD*, as the number of simultaneous users increases, the probability of having multiple users closely located to each other increases. This increases the interference they experience, as their respective transmitters point towards the same area. This causes the throughput to drop for such users. Additionally the server's computing resources are shared by the users. Hence, the average Y-PSNR decreases with more users in the arena.

In the WiFi-FSO system, *EMS* provides higher Y-PSNR than *ES* for any number of VR users. It also enables higher delivered immersion fidelity than *MS*, when there are less than 6 users. *MSD* enables the highest immersion fidelity using its narrow transmitter beamwidth, which helps to achieve higher throughput, and its optimized dynamic user-to-transmitter assignment. In the WiFi-mmWave system, *MMWDC* provides higher immersion fidelity than *MMWSC*, as the users are allocated higher data rates through separate mmWave channels.

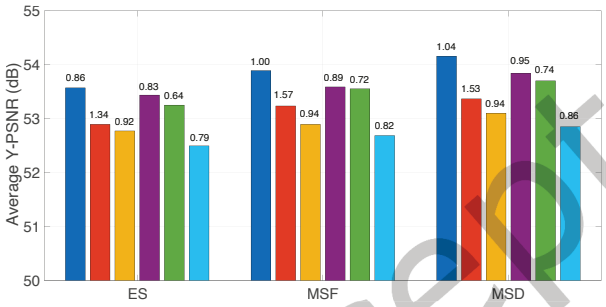


Fig. 13. Immersion fidelity and its variation, per user for ES, MSF, and MSD.

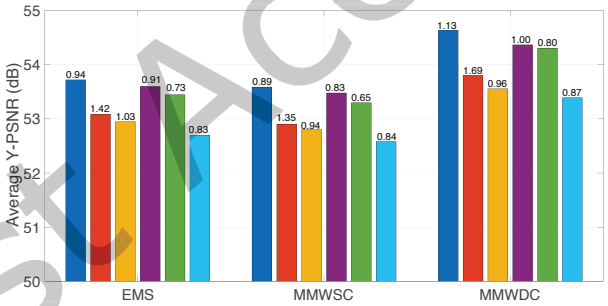


Fig. 14. Immersion fidelity and its variation, per user for EMS, MMWSC, and MMWDC.

8.3.5 Performance trade-offs. Figure 13 and Figure 14 show the expected value and standard deviation of the GOP Y-PSNR per user, with six simultaneous users in the arena system. In the WiFi-FSO setting, We can see that *MSD* provides better Y-PSNR than all the other connectivity maintenance methods. The higher Y-PSNR gain is enabled by the different aspects of *MSD* as discussed in Section 8.3.4. the delivered immersion fidelity provided by *MSF* and *EMS* is very similar but higher than *ES*. Although the Y-PSNR provided by *ES* is lower than the other methods, its variation is also the smallest. With an increase in the number of simultaneous users, the probability of having multiple users in the same cell and equally sharing its transmitter's data rate increases, which causes the Y-PSNR variation to be lower for *EMS*. Thus, it enables a more consistent performance in this regard.

In the WiFi-mmWave setting, we can observe that *MMWDC* provides higher immersion fidelity compared to *MMWSC*. Since each user is served by a separate dedicated channel in *MMWDC*, there is less interference, thus the delivered immersion fidelity is higher.

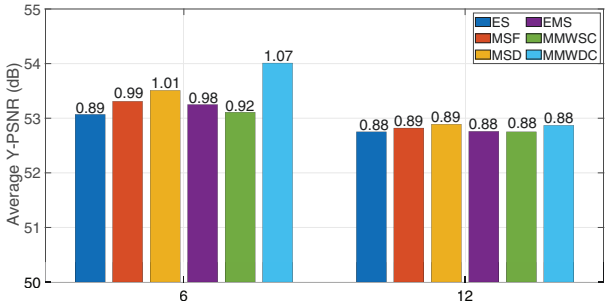


Fig. 15. Performance vs. user load for ES, MS, EMS, MMWSC, and MMWDC.

Finally, we examine the robustness of the connectivity maintenance methods to increased user load, considering 12 simultaneous users in the system. This setting corresponds to having two users in each cell at the start of the VR session. Here, for *MSF* and *MSD*, by design, the number of transmitters are increased to be equal to the number of users in the arena.

We can see from Figure 15 that though the delivered immersion fidelity slightly decreases when the number of served users is increased from six to 12, the enabled viewport Y-PSNR is still well above 52 dB, for all connectivity maintenance methods. This is expected as the probability of interference increases with user load. Also, the computing resources of the server allocated to each user decreases. In the WiFi-FSO setting, for *ES* and *EMS*, the probability of multiple users sharing the same transmitter increases with an increase in the user load, which results in reduced throughput, hence lower Y-PSNR for each user. The average standard deviation of the Y-PSNR reduces as well for *ES* and *EMS*, as the load is increased due to the same reason. Similarly, for *MSF* and *MSD*, we can see that the average Y-PSNR decreases with user load. For these methdos, as the number of users increases, the probability of multiple transmitters pointing towards the same direction increases when multiple users are located close to each other, causing higher interference. Thus, the throughput decreases for each user. Additionally, the server's computing resources allocated to the users also reduces. Hence, the average Y-PSNR decreases for *MSF* and *MSD* too. We can see that *MSD* provides higher delivered immersion fidelity for the increased user load too. Finally, in the WiFi-mmWave setting, the reduction in allocated computing resources of the server per user, as the number of users increases, causes the delivered immersion fidelity to drop significantly for *MMWDC*.

8.3.6 Comparison to the conventional state-of-the-art. To understand the benefits of our dual-connectivity streaming system relative to the state-of-the-art that relies on conventional network systems and single (traditional wireless) connectivity, we implemented a reference method that leverages the latest MPEG-DASH streaming standard, to deliver the content to users in our arena via WiFi [49]. As expected, this reference method could not stream the content at viewport quality higher than 38 dB, which is quite inadequate for this context. This outcome merits the benefits of our system design and the advances it integrates. We note that as the performance demonstrated by the reference system above was quite inferior to the results shown in the figures presented here, we omitted its respective graphs therein, for better presentation.

Finally, in Figure 16, we study the delivered viewport quality over time for a user, where we also included a variant of our system that uses only FSO/mmwave (single path) to stream the content. We can observe that the viewport quality varies over time for all the compared methods, due to transient FSO/mmwave link drops or a mismatch between the delivered enhancement representation viewport and the actual user viewport, caused by rapid head/body navigation movements. Still, the observed viewport quality variation is much lower for our system, which considerably increases its quality of experience and reliability, relative to MSDNoWiFi and MMWSCNoWiFi that experience an application downtime during such instances. Concretely, viewport quality

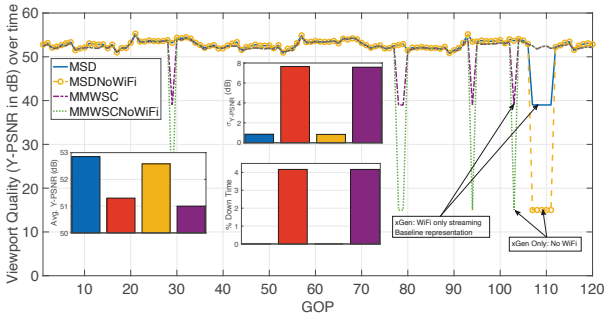


Fig. 16. Sample temporal viewport quality (user load is 6).

gains of ≈ 2 dB and eight times smaller standard deviation of viewport quality are enabled over MSDNoWiFi and MMWSCNoWiFi. We can also observe that the user experienced 4.17% application downtime for PMSDNoWiFi and MMWSCNoWiFi, compared to 0% downtime for our method. These benefits merit the integrated dual connectivity streaming and scalable 360° rate allocation.

9 CONCLUSION

We explored a novel WiFi-mmWave/FSO dual-connecitivity scalable streaming system to enable 6DOF VR-based remote scene immersion. Our system comprises an edge server that paritions the present 360° video viewpoint of a user into a baseline representation of the entire 360° panorama streamed to the user over WiFi, and a viewport-specific enhancement representation streamed to the user over a directed mmWave/FSO link. The server dynamically adapts the two representations to the available transmission rates of the two parallel links (WiFi and mmWave/FSO) to the user. These advanced capabilities of the server are enabled by the scalable multi-layer tiling of the 360° content that we synergistically explore. At the user, the two received representations are integrated to provide high fidelity VR immersion. We formulated an optimization problem to maximize the delivered immersion fidelity, which depends on the WiFi and mmWave/FSO link rates, and the computation capability of the server and the user's VR headset. We designed a geometric programming optimization framework that captures the optimal solution at lower complexity. Another key advance of the proposed system is the enabled dual-connectivity, which increases the reliability and delivered immersion fidelity, and the novel integrated approaches we investigate to maintain it. These are ES, MSF, MSD, EMS, MMWSC, and MMWDC. Moreover, we collected 6DOF navigation data of mobile VR user to evaluate the performance of the proposed system. We showed that MSD provides the best performance in the WiFi-FSO setting and MMWDC enables higher immersion fidelity in the WiFi-mmWave setting. Our results demonstrate that our system considerably advances the traditional state-of-the-art and enables streaming of 8K-120 frames-per-second (fps) 6DOF content at high fidelity.

10 ACKNOWLEDGEMENT

The work of Jacob Chakareski and Mahmudur Khan has been supported in part by the National Science Foundation under awards CCF-2031881, ECCS-2032387, CNS-2040088, CNS-2032033, and CNS-2106150; by the National Institutes of Health under award R01EY030470; and by the Panasonic Chair of Sustainability at the New Jersey Institute for Technology. To Nima, Kian, and Maya, for their love.

REFERENCES

[1] J. Chakareski, "Drone networks for virtual human teleportation," in *Proc. ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications*, Niagra Falls, NY, USA, June 2017, pp. 21–26.

[2] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, "The road to immersive communication," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 974–990, Apr. 2012.

[3] J. Chakareski, "UAV-IoT for next generation virtual reality," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5977–5990, 2019.
[4] T. S. Champel, T. Fautier, E. Thomas, and R. Koenen, "Quality requirements for VR," in *Proc. 116th MPEG Meeting of ISO/IEC JTC1/SC29/WG11*, Chengdu, China, October 2016.
[5] S. Dimitrov and H. Haas, *Principles of LED light communications: Towards networked Li-Fi*. Cambridge University Press, 2015.
[6] E. Calvanese Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez *et al.*, "6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 42–50, 2019.
[7] M. S. Rahman, K. Zheng, and H. Gupta, "FSO-VR: Steerable free space optics link for virtual reality headsets," in *Proc. ACM Workshop on Wearable Systems and Applications*, Munich, Germany, June 2018, pp. 11–15.
[8] J. Beysens, Q. Wang, A. Galisteo, D. Giustiniano, and S. Pollin, "A cell-free networking system with visible light," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 461–476, 2020.
[9] S. Blandino, G. Mangraviti, C. Desset *et al.*, "Multi-user hybrid mimo at 60 GHz using 16-antenna transmitters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 2, pp. 848–858, 2019.
[10] J. Chakareski and P. Frossard, "Distributed collaboration for enhanced sender-driven video streaming," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 858–870, Aug. 2008.
[11] B. Begole, "Why the Internet pipes will burst when virtual reality takes off," *Forbes Magazine*, Feb, 2016.
[12] J. Chakareski, R. Aksu, X. Corbillon, G. Simon, and V. Swaminathan, "Viewport-driven rate-distortion optimized 360° video streaming," in *Proc. IEEE Int'l Conf. Communications*, Kansas City, MO, May 2018.
[13] X. Corbillon, A. Devlic, G. Simon, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *Proc. Int'l Conf. Communications*. Paris, France: IEEE, May 2017.
[14] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer!" in *Proc. IEEE Int'l Symp. Multimedia*, Dec. 2016.
[15] E. Cuervo, K. Chintalapudi, and M. Kotaru, "Creating the perfect illusion: What will it take to create life-like virtual reality headsets?" in *Proc. HotMobile*. ACM, 2018, pp. 7–12.
[16] R. Zhong, M. Wang, Z. Chen, L. Liu, Y. Liu, J. Zhang, L. Zhang, and T. Moscibroda, "On building a programmable wireless high-quality virtual reality system using commodity hardware," in *Proceedings of the 8th Asia-Pacific Workshop on Systems*. ACM, 2017, p. 7.
[17] O. Abari, D. Bharadia, A. Duffield, and D. Katabi, "Enabling high-quality untethered virtual reality." in *Proc. Symposium on Networked Systems Design and Implementation*, Boston, MA, March 2017, pp. 531–544.
[18] M. Khan and J. Chakareski, "Visible light communication for next generation untethered virtual reality systems," in *Proc. IEEE International Conference on Communications Workshops*, Shanghai, China, May 2019, pp. 1–6.
[19] A. Clemm, H. Torres Vega, Maria Kumar Ravuri, T. Wauters, and F. De Turck, "Viewport-adaptive scalable multi-user virtual reality mobile-edge streaming," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 93–99, Jan. 2020.
[20] Z. Liu, Q. Li, X. Chen, C. Wu, S. Ishihara, J. Li, and Y. Ji, "Point cloud video streaming: Challenges and solutions," *IEEE Network Magazine*, vol. 35, no. 5, pp. 202–209, Sep. 2021.
[21] J. Chakareski and P. Frossard, "Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 207–218, Apr. 2006, special issue on Distributed Media Technologies and Applications.
[22] A. B. Reis, J. Chakareski, A. Kassler, and S. Sargento, "Distortion optimized multi-service scheduling for next generation wireless mesh networks," in *Proc. IEEE INFOCOM Int'l Workshop on Carrier-grade Wireless Mesh Networks*, San Diego, CA, USA, Mar. 2010.
[23] J. Chakareski, J. Apostolopoulos, W.-T. Tan, S. Wee, and B. Girod, "Distortion chains for predicting the video distortion for general packet loss patterns," in *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing*, vol. 5. Montreal, Canada: IEEE, May 2004, pp. 1001–1004.
[24] J. Chakareski, J. Apostolopoulos, S. Wee, W.-T. Tan, and B. Girod, "Rate-distortion hint tracks for adaptive video streaming," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 10, pp. 1257–1269, Oct. 2005, special issue on Analysis and Understanding for Video Adaptation.
[25] J. Chakareski and P. Chou, "RaDiO Edge: Rate-distortion optimized proxy-driven streaming from the network edge," *IEEE/ACM Trans. Networking*, vol. 14, no. 6, pp. 1302–1312, Dec. 2006.
[26] ——, "Application layer error correction coding for rate-distortion optimized streaming to wireless clients," *IEEE Trans. Communications*, vol. 52, no. 10, pp. 1675–1687, Oct. 2004.
[27] A. T. Hussein and J. M. Elmirghani, "10 Gbps mobile visible light communication system employing angle diversity, imaging receivers, and relay nodes," *OSA Journal of Optical Communications and Networking*, vol. 7, no. 8, pp. 718–735, 2015.
[28] "NVIDIA GeForce GTX 980." https://www.techpowerup.com/gpu-specs/geforce-gtx-980.c2621.
[29] "GPU Systems." https://www.supermicro.com/en/products/GPU/.
[30] "NVIDIA Video Encode and Decode GPU Support Matrix." https://developer.nvidia.com/video-encode-and-decode-gpu-support-matrix-new.
[31] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable extensions of the high efficiency video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, 2015.

[32] J. Chakareski, "Viewport-adaptive scalable multi-user virtual reality mobile-edge streaming," *IEEE Trans. Image Processing*, Dec. 2020.
[33] "QualComm Adreno 650 GPU." https://www.qualcomm.com/snapdragon/processors/comparison.
[34] G. Xu, "Global optimization of signomial geometric programming problems," *European journal of operational research*, vol. 233, no. 3, pp. 500–510, 2014.
[35] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[36] H. Willebrand and B. S. Ghuman, *Free space optics: enabling optical connectivity in today's networks*. SAMS publishing, 2002.
[37] Z. Zeng, M. D. Soltani, M. Safari, and H. Haas, "Angle diversity receiver in LiFi cellular networks," in *Proc. IEEE International Conference on Communications*, Shanghai, China, May 2019, pp. 1–6.
[38] D. Brennan, "Linear diversity combining techniques," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 331–356, 2003.
[39] A. Maltsev, A. Pudeyev, I. Karls, I. Bolotin, G. Morozov, R. Weiler, M. Peter, and W. Keusgen, "Quasi-deterministic approach to mmwave channel modeling in a non-stationary environment," in *2014 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2014, pp. 966–971.
[40] A. P. Punnen and K. Nair, "Improved complexity bound for the maximum cardinality bottleneck bipartite matching problem," *Discrete Applied Mathematics*, vol. 55, no. 1, pp. 91–93, 1994.
[41] S. Blandino, J. Senic, C. Gentile, D. Caudill, J. Chuang, and A. Kayani, "Markov multi-beamtracking on 60 GHz mobile channel measurements," *IEEE Open Journal of Vehicular Technology*, vol. 3, pp. 26–39, 2022.
[42] "Virtual Museum." https://assetstore.unity.com/packages/3d/environments/museum-117927.
[43] "SteamVR SDK." https://github.com/ValveSoftware/openvr.
[44] "Opentrack tracking." https://github.com/opentrack/opentrack.
[45] M. Khan and J. Chakareski, "NJIT 6DOF VR Navigation Dataset," https://www.jakov.org.
[46] NIST and Università di Padova, "Q-D realization software," https://github.com/wigig-tools/qd-realization.
[47] H. Assasa, J. Widmer, T. Ropitault, and N. Golmie, "Enhancing the ns-3 ieee 802.11ad model fidelity: Beam codebooks, multi-antenna beamforming training, and quasi-deterministic mmwave channel," in *Proceedings of the 2019 Workshop on Ns-3*, ser. WNS3 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 33–40. [Online]. Available: https://doi.org/10.1145/3321349.3321354
[48] H. Assasa, J. Widmer, T. Ropitault, A. Bodi, and N. Golmie, "High fidelity simulation of IEEE 802.11ad in ns-3 using a quasi-deterministic channel model," in *Proceedings of the 2019 Workshop on Next-Generation Wireless with ns-3*, 2019, pp. 22–25.
[49] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck, "Improving virtual reality streaming using HTTP/2," in *Proc. Multimedia Systems Conference*. ACM, Jun. 2017, pp. 225–228.