6DOF Virtual Reality Dataset and Performance Evaluation of Millimeter Wave vs. Free-Space-Optical Indoor Communications Systems for Lifelike Mobile VR Streaming

Jacob Chakareski[1], Mahmudur Khan[1,2], Tanguy Ropitault[3], and Steve Blandino[3]
[1] New Jersey Institute of Technology, Newark, NJ, USA, [2] York College of Pennsylvania, York, PA, USA
[3] Wireless Networks Division, National Institute of Standards and Technology, Gaithersburg, MD, USA

***Abstract*—Dual-connectivity streaming can be a key enabler of next generation 6 Degrees Of Freedom (DOF) Virtual Reality (VR) scene immersion. Indeed, using conventional sub-6 GHz WiFi allows to reliably stream a lower-quality baseline representation of the VR content while emerging 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. In this paper, we evaluate two candidates emerging technologies, Free Space Optics (FSO) and millimeter-Wave (mmWave), which both offer unprecedented available spectrum and data rates. We formulate an optimization problem to maximize the delivered immersion fidelity of the envisioned dual-connectivity 6DOF VR streaming, which depends on the WiFi and mmWave/FSO link rates, and the computing capabilities of the server and the user's VR headset. The problem is mixed integer programming and we formulate an optimization framework that captures the optimal solution at lower complexity. To evaluate the performance of the proposed systems, we collect actual 6DOF measurements. Our results demonstrate that both FSO and mmWave technologies can enable streaming of 8K-120 frames-per-second (fps) 6DOF content at high fidelity.**

I. Introduction

Virtual reality holds tremendous potential to advance our society and is expected to impact quality of life, energy conservation, and the economy. Together with 360° video, VR can suspend our disbelief of being at a remote location, akin to *virtual human teleportation* [1, 2]. 360° video streaming to VR headsets is gaining popularity in diverse areas such as gaming and entertainment, education and training, 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 [3]. 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 need to be explored. FSO and mmWave are two emerging technologies that can enable much higher data transmission rates compared to traditional wireless systems. Henceforth, we refer to both technologies as *xGen*.

Toward this objective, we investigate an integrated dual-connectivity streaming system for future 6DOF mobile multi-user VR immersion. The proposed system is illustrated in Figure 1 and synergistically integrates parallel transmission over WiFi and xGen wireless links, scalable 360° video tiling, and edge computing, to notably advance the state-of-the-art.

In particular, our novel dual-connectivity WiFi-xGen architecture aims at using *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 xGen is used for its large transmission capacity, to send a high-quality user viewport-specific enhancement representation. The two representations are then synergistically integrated at the user to considerably augment her quality of immersion and experience. Our system is fully described in Section II, and we review related work and our main contributions next.

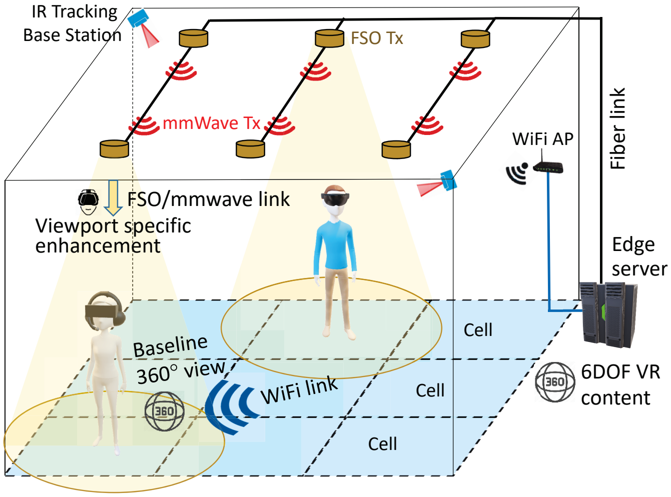

Fig. 1: 6DOF mobile VR arena WiFi-xGen scalable streaming system. WiFi delivers baseline 360° panorama of a user. Directional xGen link delivers a viewport-specific enhancement.

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 [4]. Unlike the radio

The work of Chakareski and Khan has been supported in part by NSF Awards CCF-1528030, ECCS-1711592, CNS-1836909, and CNS-1821875, and in part by NIH Award R01EY030470.

frequency spectrum, plentiful unlicensed spectrum is available for light communications, which has put FSO on the roadmap towards sixth generation (6G) networks [5]. While being a novel technology, a few studies of design concepts and experimental testbeds have already appeared [6, 7].

In the radio frequency spectrum, mmWave wireless communication is considered the enabling technology of next-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 [8]. Similarly, an energy efficient framework for UAV-assisted millimeter wave 5G heterogeneous cellular networks has been studied in [9].

Emerging VR applications require streaming of high fidelity real remote scene 360° video content, possibly with large 6DOF user mobility. Relative to traditional video streaming [10–15], VR-based 360° video streaming introduces further challenges by requiring an ultra high data rate, hyper intensive computing, and ultra low latency [16]. Though some advances have been made in 360° video streaming using traditional network systems, by intelligent resource allocation and content representation [17–19], 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. Essentially, conventional network systems are unable to address the above challenges, especially in the challenging context of 6DOF user mobility. This is the objective we pursue here.

The main contributions of our work 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 demonstrating that both dual-connectivity options, WiFi-mmWave/FSO, enable streaming of high fidelity 8K-120 fps 6DOF content.

II. Dual-Connectivity Systems

A. Dual-Connectivity Framework

Our novel dual-connectivity streaming framework is illustrated in Figure 1 for a VR arena scenario. In our system, N_u VR users $U = \{1, 2, ..., N_u\}$ navigate a 6DOF 360° video content in an indoor VR arena. We divide the spatial area of the arena into N_x cells of equal size. 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. Each user VR headset is dual-connectivity enabled and equipped with a WiFi and an xGen transceiver. Uplink communication between the headset and the server is carried out via WiFi, to share control information. The server controls both the WiFi uplink and downlink transmission.

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. 2). 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.

B. 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. We describe the server's operation below in detail.

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

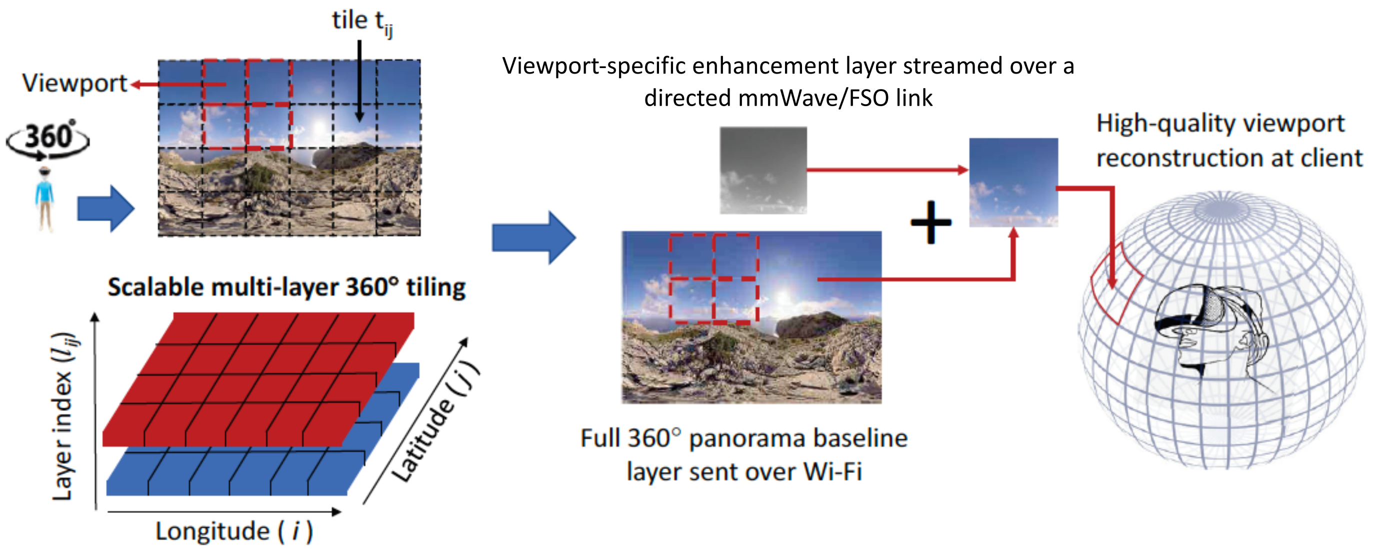
tile t_{ij}
Viewport
360°
Scalable multi-layer 360° tiling
Layer index (l_{ij})
Latitude (j)
Longitude (i)
Viewport-specific enhancement layer streamed over a directed mmWave/FSO link
+
High-quality viewport reconstruction at client
Full 360° panorama baseline layer sent over Wi-Fi

Fig. 2: 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.

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 $R_{ij,\max}$.

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 [21] to enable the server to build this information and benefit our analysis and optimization of the resource allocation carried out by the server.

3) GOP-tile decoding at 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 tradeoffs that can be leveraged in our analysis and optimization.

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^w_u and $\sum_{u \in U} C^w_u \leq C^w$. We formulate the delay of streaming the baseline representation of the entire 360° panorama to user u as $\tau^w_u = \frac{\sum_{ij} R_{ij,w} \Delta T}{C^w_u}$.

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

C. User headset modeling

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 use an FSO transceiver on the headset, adopted from our recent work [22], as the xGen transceiver. For a VR-arena with mmWave transmitters, the xGen transceiver on the headset is a mmWave transceiver [8].

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. The maximum decoding speed of the headset is $z_u \geq z^w_u + z^x_u$, where z^w_u is the speed allocated for decoding the GOP-tiles (baseline representation) received over the WiFi link and z^x_u 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 $\tau^{z,w}_u = \frac{\sum_{ij} R_{ij,w} \Delta T}{z^w_u}$ 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.

D. *User viewport reconstruction error*

We leverage our recent modeling advances [17] 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.

III. Problem Formulation

Our objective is to minimize the aggregate reconstruction error 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:

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

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

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

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

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

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

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 $|M_u^r| \geq 0$ 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 subproblem for user u assigned to xGen transmitter x as

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

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

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 $M_{u,k}^{\mathrm{r}}$ 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 II-B. For each $M_{u,k}^{\mathrm{r}}$, we find the streaming data rates $\{R_{ij,x,k}^\star\}$ and $\{R_{ij,w,k}^\star\}$ associated with the baseline and enhancement representations, and the user's headset decoding speed allocations $\{z_{u,k}^{x\star}\}$ and $\{z_{u,k}^{w\star}\}$, and rendering speed allocations $\{r_{u,k}^{x\star}\}$ and $\{z_{u,k}^{w\star}\}$, for which the reconstruction distortion $D_{u,k}^\star$ is minimum. Finally, we select the value $k^\star$ for which for which $D_{u,k}^\star$ is the lowest and this completes the solution to (7). We describe our proposed approach in more detail in the following section.

IV. Computing Optimal Resource Allocation

When the selection of GOP-tiles to be streamed in raw format is fixed, i.e., for a given value of k and $M_{u,k}^{\mathrm{r}}$, 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 II-D), 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 [23]. In particular, according to this method, for a constraint which is a ratio of posynomials, the denominator posynomial can be approximated into a monomial. This will enable us to reformulate all constraints in (9) involving ratios as posynomials, to solve (9) as GP. We formulate an iterative method towards this objective. At each iteration t, we convert the constraints (2) and (10) into respective posynomial functions. Space constraints prevent us to include the resulting expressions here. Then, the optimization problem to be solved at iteration t is:

$$\min_{\substack{\{R_{ij,xw,k}\},\ r^{w}_{u,k},\ z^{w}_{u,k},\\ \{R_{ij,w,k}\},\ r^{x}_{u,k},\ z^{x}_{u,k}}} D^{xw}_{u,k}(t), \tag{12}$$
$$\text{s.t.} \quad (2), (6), (10), (11).$$

Here, (12) is a GP problem and we can solve it optimally. We carry out the optimization iteratively until $|D^{xw}_{u,k}(t) - D^{xw}_{u,k}(t-1)| \leq \epsilon$, for some small $\epsilon \geq 0$. When this condition is met, we obtain the optimal value of the objective function in (12) as $D^{\star}_{u,k} = \sum_{ij \in M^{r}_{u,k}} p^{u}_{ij} a_{ij} R^{b_{ij}}_{ij,max} + D^{xw}_{u,k}(t)$, the optimal streaming data rate $\{R^{\star}_{ij,xw,k}\} = \{R_{ij,xw,k}(t)\}$, the optimal headset decoding speed allocations $z^{w\star}_{u,k} = z^{w}_{u,k}(t)$ and $z^{x\star}_{u,k} = z^{x}_{u,k}(t)$, and rendering capability allocations $r^{w\star}_{u,k} = r^{w}_{u,k}(t)$ and $r^{x\star}_{u,k} = r^{x}_{u,k}(t)$, 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^{r}_{u} by finding the k and $M^{r}_{u,k}$ that result in the smallest $D^{\star}_{u,k}$. We formally write this optimization as:

$$D^{\text{opt}}_{u} = \min D^{\star}_{u,k}, \tag{13}$$
$$\begin{matrix} M^{r,\text{opt}}_{u}, & z^{\text{opt}}_{w}, & r^{\text{opt}}_{w}, \\ \{R^{\text{opt}}_{ij,xw}\}, & z^{\text{opt}}_{x}, & r^{\text{opt}}_{x} \end{matrix} = \operatorname*{arg\,min}_{\substack{M^{r}_{u,k},\ z^{w\star}_{u,k},\ r^{w\star}_{u,k},\\ \{R^{\star}_{ij,xw,k}\},\ z^{x\star}_{u,k},\ r^{x\star}_{u,k}}} D^{\star}_{u,k}.$$

This completes the solution to the problem in (7).

V. xGen connectivity Maintenance

A. Free-Space Optics

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

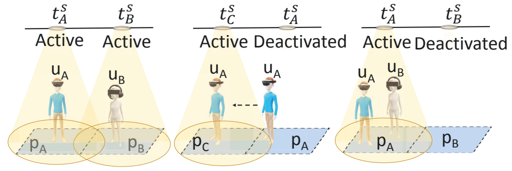

Fig. 3: Electronic steering.

1) 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 (see Fig. 3). We define this switching of transmitter assignment for a user as *electronic steering*. Also, when multiple users are within the same cell, they are all assigned to the same transmitter and allocated an equal share of its data rate.

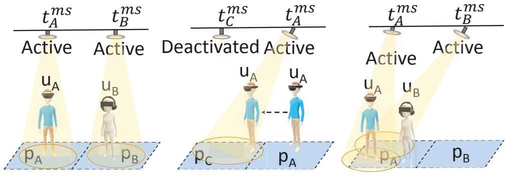

Fig. 4: Mechanical steering.

2) 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 (see Fig. 4). The server uses the tracking information to steer a transmitter towards its assigned user 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 $d_{u,x}$. Every ΔT time units, a user-to-transmitter re-assignment is performed such that the smallest $s_{u,x}$ is maximized, or equivalently the biggest $d_{u,x}$ is minimized.

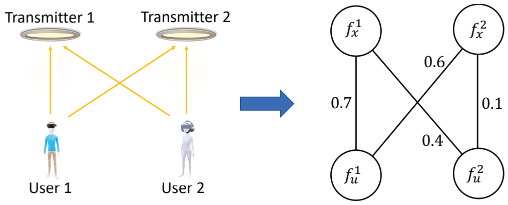

Fig. 5: Bipartite graph example for 2 transmitters and 2 users.

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 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.,

$$\min_{\pi \in \Pi} \max_{(f_u^1, f_x^2) \in \pi} \omega_{(f_u^1, f_x^2)}, \tag{14}$$

where Π comprises all the possible maximum matchings. For the graph in Fig. 5, 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 solve the problem in (14) using the BM algorithm proposed in [24].

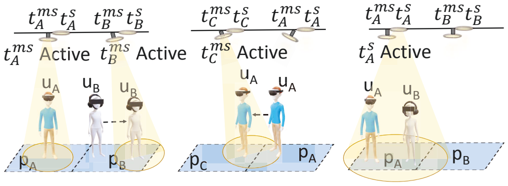

Fig. 6: Electro-mechanical steering.

3) 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 (Fig. 6).

B. Millimeter Wave

We define two different mmWave connectivity maintenance: mmWave Same Channel (MMWSC) and mmWave Different Channel (MMWDC).

1) MMWSC: In this scheme, all the mmWave transmitters are configured to operate in the same channel. The users are associated to the transmitter which yields the highest receive power for a given 6DOF position.

2) MMWDC: In this scheme, each of the mmWave transmitters are configured to operate on a different channel. At the beginning of the simulation, every user is associated to the transmitter yielding the highest receive power and stays associated to this transmitter for the entire VR session.

VI. WiFi-xGen Performance Evaluation

Here, we carry out performance evaluation of the two proposed WiFi-xGen dual-connectivity streaming systems. Our simulation experiments leverage real 6DOF navigation measurements to incorporate realistic body and head movements comprising VR navigation in the performance evaluation.

A. 6-DOF Navigation Measurements

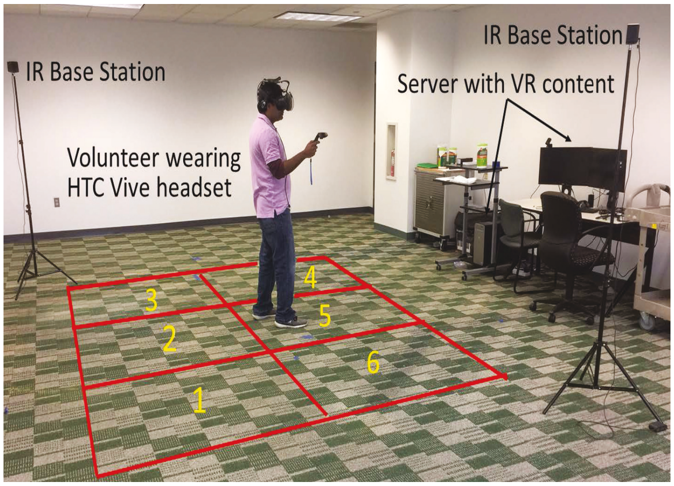

Fig. 7: 6-DOF navigation data measurement session.

The 6-DOF 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 7, where the users navigated the 6-DOF VR content *Virtual Museum* [25] 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 [26] and Opentrack [27] to record the navigation information for the users in our arena system, as they were being tracked during a session (see Section II-A). 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 [28].

B. *Simulation Setup*

We explore two simulation settings associated with each WiFi-xGen streaming system investigated in this paper.

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 $\beta = 0.75\Theta_d$ [22]. The tracking data accuracy is ±1 mm. The system-level results are obtained via a Matlab implementation.

2) WiFi-mmWave: We equip the VR arena with six mmWave transmitters, one for each cell. Each transmitter and VR Headset is equipped with a 16 phased-antenna array disposed in a rectangular 2×8 configuration to enable beamforming in both azimuth and elevation. The millimeter-wave propagation is generated using the open source NIST Quasi-Deterministic channel model implementation [29], which can accurately predict the channel characteristics for millimeter wave frequencies. The system level results are obtained via an NS-3 IEEE 802.11ad implementation.

For both scenarios, we assess the immersion fidelity (quality) of the viewport of user u via the luminance (Y) Peak Signal to Noise Ratio (Y-PSNR) of the expected viewport distortion experienced by the user over a GOP, computed as $10\log_{10}(255^2/\sum_{ij\in M_u} p^u_{ij} D_{ij})$. We model the distortion terms D_{ij} associated with the GOP tiles m_{ij} comprising the present 360° video viewpoint/panorama of the user, using the popular 8K 360° video sequence *Runner* [30], scalable encoded at different data rates and 120fps temporal frame rate. We compute the Y-PSNR for every GOP and the average GOP Y-PSNR across the entire session.

C. *Results and analysis*

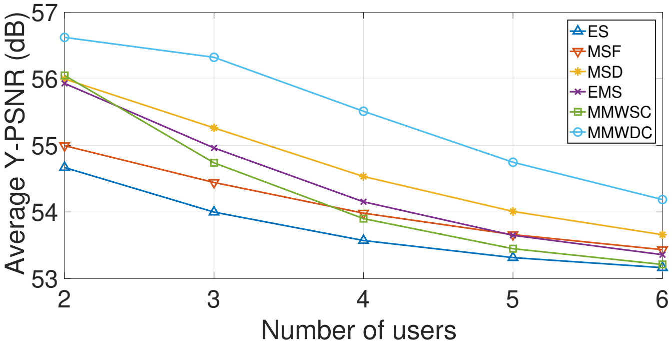

Fig. 8: Immersion fidelity for different xGen connectivity maintenance methods.

We can see in Figure 8 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.

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.

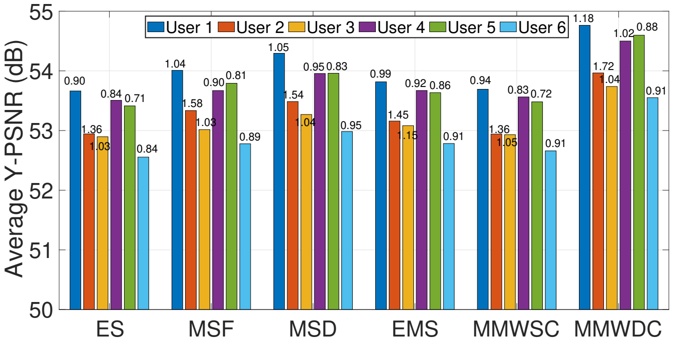

Fig. 9: Immersion fidelity and its standard deviation (on top of each bar) per user (six users in the arena).

Figure 9 shows 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, 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.

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 10 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. Hence, streaming of 8K-120fps 6-DOF content at high fidelity is still achieved for all users. Here, *MSD* and *MMWDC* deliver the highest immersion fidelity.

1) Comparison to the (conventional) state-of-the-art: To have an understanding of 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

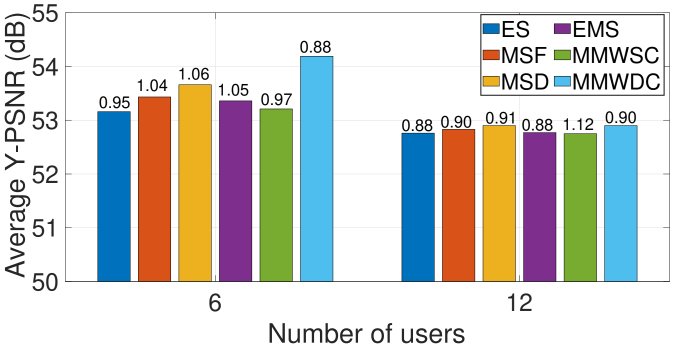

Fig. 10: Immersion fidelity for different xGen connectivity in an overloaded VR-arena.

content to users in our system via WiFi [31]. As anticipated, the 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.

VII. 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. 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 all the connectivity methods in either setting can enable streaming of 8K-120 fps 6DOF content at high fidelity, thereby advancing the conventional state-of-the-art considerably.

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] ——, "UAV-IoT for next generation virtual reality," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5977–5990, 2019.

[3] 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.

[4] S. Dimitrov and H. Haas, *Principles of LED light communications: Towards networked Li-Fi*. Cambridge University Press, 2015.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] J. Chakareski, S. Naqvi *et al.*, "An energy efficient framework for UAV-assisted millimeter wave 5G heterogeneous cellular networks," *IEEE Trans. Green Communications and Networking*, vol. 3, no. 1, Mar. 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] J. Chakareski, J. Apostolopoulos, S. Wee, W.-T. Tan, and B. Girod, "R-D hint tracks for low-complexity R-D optimized video streaming," in *Proc. IEEE Int'l Conf. Multimedia and Expo*, Taipei, Taiwan, Jun. 2004.

[12] 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.

[13] 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.

[14] J. Chakareski, R. Sasson, A. Eleftheriadis, and O. Shapiro, "System and method for low-delay, interactive communication using multiple TCP connections and scalable coding," U.S. Patent 7 933 294, Apr. 26, 2011.

[15] J. Chakareski, R. Sasson, and A. Eleftheriadis, "System and method for the control of the transmission rate in packet-based digital communications," U.S. Patent 7 701 851, Apr. 20, 2010.

[16] B. Begole, "Why the Internet pipes will burst when virtual reality takes off," *Forbes Magazine*, Feb, 2016.

[17] 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.

[18] 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.

[19] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer!" in *Proc. IEEE Int'l Symp. Multimedia*, Dec. 2016.

[20] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable extensions of the high efficiency video coding standard," *IEEE Trans. Circuits and Systems for Video Technology*, 2015.

[21] J. Chakareski, "Viewport-adaptive scalable multi-user virtual reality mobile-edge streaming," *IEEE Trans. Image Processing*, Dec. 2020.

[22] M. Khan and J. Chakareski, "Visible light communication for next generation untethered virtual reality systems," in *Proc. IEEE Int'l Conf. Communications Workshops*, Shanghai, China, May 2019, pp. 1–6.

[23] G. Xu, "Global optimization of signomial geometric programming problems," *Elsevier European Journal of Operational Research*, 2014.

[24] 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.

[25] "Virtual Museum." https://assetstore.unity.com/packages/3d/environments/museum-117927.

[26] "SteamVR SDK." https://github.com/ValveSoftware/openvr.

[27] "Opentrack tracking." https://github.com/opentrack/opentrack.

[28] M. Khan and J. Chakareski, "NJIT 6DOF VR Navigation Dataset," https://www.jakov.org.

[29] NIST and Università di Padova, "Q-D realization software," https://github.com/wigig-tools/qd-realization.

[30] X. Liu, Y. Huang, L. Song, R. Xie, and X. Yang, "The SJTU UHD 360-degree immersive video sequence dataset," in *Proc. IEEE Int'l Conf. Virtual Reality and Visualization (ICVRV)*, 2017, pp. 400–401.

[31] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation," in *Proc. ACM Multimedia Systems Conference*, Taipei, Taiwan, June 2017, pp. 261–271.