

# Optimal Streaming of 360 VR Videos with Perfect, Imperfect and Unknown FoV Viewing Probabilities

Lingzhi Zhao, Ying Cui and Chengjun Guo  
Shanghai Jiao Tong University, China

Zhi Liu  
Shizuoka University, Japan

**Abstract**—In this paper, we investigate wireless streaming of multi-quality tiled 360 virtual reality (VR) videos from a multi-antenna server to multiple single-antenna users in a multi-carrier system. To capture the impact of field-of-view (FoV) prediction, we consider three cases of FoV viewing probability distributions, i.e., perfect, imperfect and unknown FoV viewing probability distributions, and use the average total utility, worst average total utility and worst total utility as the respective performance metrics. We adopt rate splitting with successive decoding for efficient transmission of multiple sets of tiles of different 360 VR videos to their requesting users. In each case, we optimize the encoding rates of the tiles, minimum encoding rates of the FoVs, rates of the common and private messages and transmission beamforming vectors to maximize the total utility. The problems in the three cases are all challenging nonconvex optimization problems. We successfully transform the problem in each case into a difference of convex (DC) programming problem with a differentiable objective function, and obtain a suboptimal solution using concave-convex procedure (CCCP). Finally, numerical results demonstrate the proposed solutions achieve notable gains over existing schemes in all three cases. To the best of our knowledge, this is the first work revealing the impact of FoV prediction and its accuracy on the performance of streaming of multi-quality tiled 360 VR videos.

## I. INTRODUCTION

A 360 VR video is generated by capturing a scene of interest in every direction at the same time using omnidirectional cameras. A user wearing a VR headset or head mounted display (HMD) can freely watch the scene of interest in any viewing direction at any time, hence enjoying immersive viewing experience. VR has vast applications in entertainment, education, medicine, etc. A 360 VR video is of a much larger size than a traditional video. At any moment, a user watching a 360 VR video is interested in only one viewpoint, i.e., the center of one part of the 360 VR video, referred to as FoV. To improve transmission efficiency for 360 VR videos, tiling technique is widely adopted. Transmitting only the set of tiles covering the predicted FoV or FoVs with higher viewing probabilities can reduce the required communication resources, while maintaining the quality of experience (QoE) to certain extent. In addition, pre-encoding each tile into multiple representations with different quality levels allows quality adaptation according to user heterogeneity (e.g., in cellular usage costs, display resolutions of devices, channel

conditions, etc.). In this paper, we focus on wireless streaming of multi-quality tiled 360 VR videos to multiple users.

Based on viewpoint or FoV prediction, [1]–[5] study wireless streaming of one multi-quality tiled 360 VR video [1]–[3] or multiple multi-quality tiled 360 VR videos [4], [5]. Specifically, it is assumed in [1]–[5] that the FoVs that may be watched are known. Furthermore, in [3]–[5], the viewing probability distributions over those FoVs are also available. Given such FoV prediction results, in [1]–[5], the authors optimize the quality level selection and communication resource allocation to maximize the total utility [1], [4], [5], minimize the total distortion [3], or minimize the total transmission power [2].

Note that most existing works [1]–[5] rely on the assumption of perfect FoV prediction. It is unknown how FoV prediction errors influence the performance of wireless streaming of multi-quality tiled 360 VR videos. Besides, note that [1]–[5] all consider single-antenna servers, which cannot exploit spatial degrees of freedom, and hence cannot provide satisfactory performance for wireless streaming of 360 VR videos. In practice, deploying multiple antennas at a server can significantly improve the performance of wireless systems via designing efficient beamformers, and linear precoders such as zero forcing [6] are widely used. However, any linear precoding scheme can be far from optimal, as the cost to suppress interference can be high when the channels for some of the users are spatially aligned. To circumvent such limitation, the idea of rate splitting which partially decodes interference and partially treats interference as noise [7] is introduced to improve the overall performance for serving multiple users.

In this paper, we investigate wireless streaming of multi-quality tiled 360 VR videos from a multi-antenna server to multiple single-antenna users in a multi-carrier system. We consider three cases of FoV viewing probability distributions, i.e., perfect, imperfect and unknown FoV viewing probability distributions, and use the average total utility, worst average total utility and worst total utility as the respective performance metrics. We adopt rate splitting with successive decoding for efficient transmission of multiple sets of tiles from different 360 VR videos to their requesting users. In each case of FoV viewing probability distributions, we optimize the encoding rates of the tiles, minimum encoding rates of the FoVs, rates of the common and private messages and beamforming vectors to maximize the total utility. The problems in the three cases are

The work of Z. Liu was supported in part by JSPS KAKENHI under Grants 19H04092, 20H04174, and in part by ROIS NII Open Collaborative Research 2020 (20FA02).

all nonconvex optimization problems. Moreover, the optimization problem in the case of imperfect FoV viewing probability distributions is a max-min problem, and the optimization problem in the case of unknown FoV viewing probability distributions has a non-differentiable objective function. We successfully transform the problem in each case into a DC programming problem with a differentiable objective function, and obtain a suboptimal solution using CCCP [10]. Finally, numerical results show substantial gains of the proposed solutions over existing schemes in all three cases, and reveal the impact of the FoV prediction and its accuracy on streaming of multi-quality tiled 360 VR videos.

## II. SYSTEM MODEL

### A. Multi-Quality Tiled 360 VR Videos

We consider tiling to enable flexible transmission of necessary FoVs of a 360 VR video. Specifically, a 360 VR video is divided into  $X \times Y$  rectangular segments of the same size, referred to as tiles, where  $X$  and  $Y$  represent the numbers of segments in each column and each row, respectively. Define  $\mathcal{X} \triangleq \{1, \dots, X\}$  and  $\mathcal{Y} \triangleq \{1, \dots, Y\}$ . The  $(x, y)$ -th tile refers to the tile in the  $x$ -th row and the  $y$ -th column, for all  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ . For a 360 VR video, consider  $\bar{I}$  viewpoints (i.e.,  $\bar{I}$  FoVs). Denote  $\bar{\mathcal{I}} \triangleq \{1, \dots, \bar{I}\}$ . For all  $i \in \bar{\mathcal{I}}$ , let  $\mathcal{F}_i$  denote the set of tiles fully or partially included in the  $i$ -th FoV. Considering user heterogeneity, we pre-encode each tile into  $L$  representations corresponding to  $L$  quality levels using High Efficiency Video Coding (HEVC), as in Dynamic Adaptive Streaming over HTTP (DASH). Let  $\mathcal{L} \triangleq \{1, \dots, L\}$  denote the set of quality levels. For all  $l \in \mathcal{L}$ , the  $l$ -th representation of each tile corresponds to the  $l$ -th lowest quality. For ease of exposition, assume that the encoding rates of the tiles with the same quality level are identical. The encoding rate of the  $l$ -th representation of a tile is denoted by  $D_l$  (in bits/s), where  $D_1 < \dots < D_L$ .

As illustrated in Fig. 1, we consider wireless streaming of  $K$  multi-quality tiled 360 VR videos from a server to  $K$  users, respectively. Let  $\mathcal{K} \triangleq \{1, \dots, K\}$  denote the set of user indices (video indices). We study the duration of the playback time of one group of picture (GOP), which is usually 0.5-1 seconds. A user can freely switch views, when watching a 360 VR video. Suppose the set of FoVs that may be watched by user  $k$  and the corresponding FoV viewing probability distribution have been predicted to certain extent. Let  $\mathcal{I}_k$  represent the sets of indices of the  $I_k$  FoVs of video  $k$  that may be watched by user  $k$ . Note that for all  $k \in \mathcal{K}$ ,  $\mathcal{F}_i, i \in \mathcal{I}_k$  may overlap and user  $k$  will watch one of the  $I_k$  FoVs. Suppose  $\mathcal{I}_k, k \in \mathcal{K}$  are known to the server. For all  $k \in \mathcal{K}$  and  $i \in \mathcal{I}_k$ , let  $p_{i,k}$  denote the probability that the  $i$ -th FoV of video  $k$  is viewed by user  $k$ . Here,  $p_{i,k} \geq 0, i \in \mathcal{I}_k, k \in \mathcal{K}$  and  $\sum_{i \in \mathcal{I}_k} p_{i,k} = 1, k \in \mathcal{K}$ . Denote  $\mathbf{p}_k \triangleq (p_{i,k})_{i \in \mathcal{I}_k}, k \in \mathcal{K}$ . In the following, we consider three cases of FoV viewing probability distributions.

**Perfect FoV viewing probability distributions:** In this case, FoV viewing probability distributions have been estimated by some learning methods, and the estimation errors are

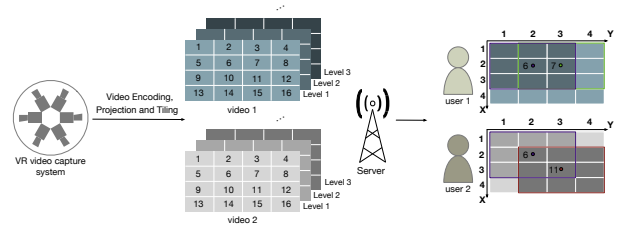


Fig. 1. System model.  $X = 4, Y = 4, \bar{I} = 16, L = 3, K = 2, \mathcal{I}_1 = \{6, 7\}, \mathcal{F}_6 = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}, \mathcal{F}_7 = \{(1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4)\}, r_{6,1} = D_2, r_{7,1} = D_1, R_{x,y,1} = D_2, (x, y) \in \mathcal{F}_6, R_{x,y,1} = D_1, (x, y) \in \mathcal{F}_7 \setminus \mathcal{F}_6, \mathcal{I}_2 = \{6, 11\}, \mathcal{F}_6 = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}, \mathcal{F}_{11} = \{(2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4), (4, 2), (4, 3), (4, 4)\}, r_{6,2} = D_2, r_{11,2} = D_3, R_{x,y,2} = D_3, (x, y) \in \mathcal{F}_{11}, R_{x,y,2} = D_2, (x, y) \in \mathcal{F}_6 \setminus \mathcal{F}_{11}$ .

negligible. That is, the exact values of  $\mathbf{p}_k, k \in \mathcal{K}$  are assumed to be known to the server.

**Imperfect FoV viewing probability distributions:** In this case, FoV viewing probability distributions have been estimated by some learning methods, with certain estimation errors. For all  $k \in \mathcal{K}$  and  $i \in \mathcal{I}_k$ , let  $\hat{p}_{i,k}$  denote the estimated probability that the  $i$ -th FoV of video  $k$  is viewed by user  $k$ , and let  $\Delta_{i,k} \triangleq p_{i,k} - \hat{p}_{i,k}$  denote the corresponding estimation error, where  $\sum_{i \in \mathcal{I}_k} \Delta_{i,k} = 0$  and  $|\Delta_{i,k}| \leq \varepsilon_{i,k}$  for some known  $\varepsilon_{i,k} \in (0, 1)$ . Assume that  $\hat{p}_{i,k}, \Delta_{i,k}, i \in \mathcal{I}_k, k \in \mathcal{K}$  are known to the server, and  $\mathbf{p}_k, k \in \mathcal{K}$  are not known to the server. Thus, the server knows that the exact FoV viewing probability distributions satisfy  $\mathbf{p}_k \in \mathcal{P}_k$ , where

$$\mathcal{P}_k \triangleq \left\{ \mathbf{p}_k \mid \underline{p}_{i,k} \leq p_{i,k} \leq \bar{p}_{i,k}, i \in \mathcal{I}_k, \sum_{i \in \mathcal{I}_k} p_{i,k} = 1 \right\}, k \in \mathcal{K},$$

with  $\underline{p}_{i,k} \triangleq \max\{\hat{p}_{i,k} - \varepsilon_{i,k}, 0\}$  and  $\bar{p}_{i,k} \triangleq \min\{\hat{p}_{i,k} + \varepsilon_{i,k}, 1\}, i \in \mathcal{I}_k, k \in \mathcal{K}$ .

**Unknown FoV viewing probability distributions:** In this case, there is no prior information about the exact FoV viewing probability distributions  $\mathbf{p}_k, k \in \mathcal{K}$ .

For all  $k \in \mathcal{K}$ , the tiles in  $\bar{\mathcal{F}}_k \triangleq \cup_{i \in \mathcal{I}_k} \mathcal{F}_i$  may be transmitted to user  $k$ . Let  $R_{x,y,k}$  (in bits/s) denote the encoding rate of the  $(x, y)$ -th tile of video  $k$ , where

$$R_{x,y,k} \in \{0, D_1, \dots, D_L\}, (x, y) \in \bar{\mathcal{F}}_k, k \in \mathcal{K}. \quad (1)$$

Here,  $R_{x,y,k} = 0$  indicates that the  $(x, y)$ -th tile of video  $k$  will not be transmitted to user  $k$ , and  $R_{x,y,k} = D_l$  indicates that the  $l$ -th representation of the  $(x, y)$ -th tile will be transmitted to user  $k$ . To avoid degrading QoE, we consider a relative smoothness requirement for quality variation in an FoV [2]:

$$r_{i,k} \leq R_{x,y,k} \leq r_{i,k} + \delta, (x, y) \in \mathcal{F}_i, i \in \mathcal{I}_k, k \in \mathcal{K}, \quad (2)$$

where

$$r_{i,k} \in \{0, D_1, \dots, D_L\}, i \in \mathcal{I}_k, k \in \mathcal{K}. \quad (3)$$

Here,  $r_{i,k}$  (in bits/s) represents the minimum of the encoding rates of the tiles in the  $i$ -th FoV of video  $k$ , also referred to as the minimum encoding rate of the  $i$ -th FoV of video  $k$ , and indicates the quality level of the  $i$ -th FoV for user  $k$ ; and  $\delta > 0$  is a small number representing the tolerance for quality variation in an FoV (note that the quality variation for tiles in one FoV is not visible if  $\delta$  is small enough). An illustration

example can be found in Fig. 1.

Let  $U(r)$  denote the utility for an FoV with the minimum encoding rate  $r$ . Here, the utility function  $U(\cdot)$  can be any nonnegative, strictly increasing and strictly concave function, and  $U(0) = 0$ . Let  $Q^{(t)}(\mathbf{r}), t = \text{pp, ip and up}$  denote the performance metrics in the three cases of FoV viewing probability distributions. In the case of perfect FoV viewing probability distributions (i.e., case-pp), we use the average total utility as the performance metric. In the case of imperfect FoV viewing probability distributions (i.e., case-ip), we use the worst average total utility as the performance metric. In the case of unknown FoV viewing probability distributions (i.e., case-up), we use the worst total utility as the performance metric. Therefore, the performance metrics in the three cases of FoV viewing probability distributions can be written as:

$$Q^{(t)}(\mathbf{r}) = \begin{cases} \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} p_{i,k} U(r_{i,k}), & t = \text{pp}, \\ \sum_{k \in \mathcal{K}} \min_{\mathbf{p}_k \in \mathcal{P}_k} \sum_{i \in \mathcal{I}_k} p_{i,k} U(r_{i,k}), & t = \text{ip}, \\ \sum_{k \in \mathcal{K}} \min_{i \in \mathcal{I}_k} U(r_{i,k}), & t = \text{up}. \end{cases} \quad (4)$$

### B. Physical Layer Model and Transmission Scheme

The server has  $M$  antennas, and each user has one antenna. We consider a multi-carrier system. Let  $N$  and  $\mathcal{N} \triangleq \{1, \dots, N\}$  denote the number of subcarriers and the set of subcarrier indices, respectively. The bandwidth of each subcarrier is  $B$  (in Hz). We assume block fading, i.e., the channel on each subcarrier remains constant over the considered time duration. Let  $\mathbf{h}_{k,n}^H \in \mathbb{C}^{1 \times M}$  denote the downlink channel vector on subcarrier  $n$  between user  $k$  and the server. We assume that the channel state information is perfectly known at the server and the users.

For all  $k \in \mathcal{K}$ , encoded (source coding) bits of the tiles in  $\bar{\mathcal{F}}_k$  that will be transmitted to user  $k$  are ‘‘aggregated’’ into one message. We consider a rate splitting scheme [7]. Specifically, for all  $k \in \mathcal{K}$ , the aggregated message for user  $k$  is further split into a common part of rate  $d_{c,k}$  and a private part of rate  $d_{p,k}$ . Thus, we have:

$$\sum_{(x,y) \in \bar{\mathcal{F}}_k} R_{x,y,k} = d_{c,k} + d_{p,k}, \quad k \in \mathcal{K}. \quad (5)$$

Further, the common parts of the messages of the  $K$  users are combined into a common message of rate  $\sum_{k \in \mathcal{K}} d_{c,k}$ . The private part of user  $k$ 's message is also referred to as user  $k$ 's private message. The common message and the  $K$  users' private messages are then encoded (channel coding) into codewords that span over  $N$  subcarriers, respectively. Let  $\mathbf{w}_{c,n} \in \mathbb{C}^{M \times 1}$  and  $\mathbf{w}_{k,n} \in \mathbb{C}^{M \times 1}$  denote the common beamforming vector and the private beamforming vector for user  $k$  on subcarrier  $n$ , respectively. We have the total transmission power constraint:

$$\sum_{n \in \mathcal{N}} (\|\mathbf{w}_{c,n}\|_2^2 + \sum_{k \in \mathcal{K}} \|\mathbf{w}_{k,n}\|_2^2) \leq P, \quad (6)$$

where  $P$  is the total transmission power budget. Let  $s_{c,n}$  and  $s_{k,n}$  denote a symbol of the common message and a symbol of user  $k$ 's private message, which are transmitted on the  $n$ -th subcarrier, respectively. For notation simplicity, define  $\bar{\mathcal{K}} \triangleq \mathcal{K} \cup \{c\}$ . Let  $\mathbf{s}_n \triangleq (s_{k,n})_{k \in \bar{\mathcal{K}}}$  and assume that  $\mathbb{E}[\mathbf{s}_n \mathbf{s}_n^H] = \mathbf{I}$ ,  $n \in \mathcal{N}$ . We consider linear precoding on each subcarrier. The

received signal at user  $k$  on subcarrier  $n$  is given by:

$$y_{k,n} = \mathbf{h}_{k,n}^H \mathbf{w}_{c,n} s_{c,n} + \mathbf{h}_{k,n}^H \mathbf{w}_{k,n} s_{k,n} + \sum_{j \in \mathcal{K}, j \neq k} \mathbf{h}_{n,j}^H \mathbf{w}_{j,n} s_{j,n} + z_{k,n}, \quad k \in \mathcal{K}, \quad n \in \mathcal{N}, \quad (7)$$

where  $z_{k,n} \sim \mathcal{CN}(0, \sigma^2)$  represents the received Additive White Gaussian Noise (AWGN) at user  $k$  on subcarrier  $n$ .

We consider successive decoding at each user. Specifically, the decoding procedure for user  $k \in \mathcal{K}$  is as follows. First, user  $k$  decodes the common message by treating the interference from the  $K$  users' private messages on each subcarrier as noise. After successfully decoding and removing the common message, user  $k$  decodes his private message by treating the interference from the remaining  $K-1$  users' private messages on each subcarrier as noise. The Signal to Interference plus Noise Ratios (SINRs) of the common message and user  $k$ 's private message on subcarrier  $n$  are given by  $\frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{c,n}|^2}{\sum_{j \in \mathcal{K}} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2}$  and  $\frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{k,n}|^2}{\sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2}$ , respectively. Assuming Gaussian signaling, to guarantee that each user can successfully receive the common message and his private message, we have:

$$\sum_{k \in \mathcal{K}} d_{c,k} \leq \sum_{n \in \mathcal{N}} B \log_2 \left( 1 + \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{c,n}|^2}{\sum_{j \in \mathcal{K}} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2} \right), \quad k \in \mathcal{K}, \quad (8)$$

$$d_{p,k} \leq \sum_{n \in \mathcal{N}} B \log_2 \left( 1 + \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{k,n}|^2}{\sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2} \right), \quad k \in \mathcal{K}. \quad (9)$$

### III. PROBLEM FORMULATION AND OPTIMALITY PROPERTIES

We would like to optimize the encoding rates of the tiles  $\mathbf{R} \triangleq (R_{x,y,k})_{(x,y) \in \bar{\mathcal{F}}_k, k \in \mathcal{K}}$ , minimum encoding rates of the FoVs  $\mathbf{r} \triangleq (r_{i,k})_{i \in \mathcal{I}_k, k \in \mathcal{K}}$ , rates of the common and private messages  $\mathbf{d} \triangleq (d_{c,k}, d_{p,k})_{k \in \mathcal{K}}$  and beamforming vectors  $\mathbf{w} \triangleq (\mathbf{w}_{k,n})_{k \in \bar{\mathcal{K}}, n \in \mathcal{N}}$  to maximize the performance metrics in (4) subject to the constraints in (1), (2), (3), (5), (6), (8), (9). Note that  $\mathbf{R}$  and  $\mathbf{r}$  are discrete variables. For tractability, we consider a relaxed version of the discrete optimization, as in [8]. That is, we replace the discrete constraints in (1) and (3) with the following continuous constraints:

$$0 \leq R_{x,y,k} \leq D_L, \quad (x,y) \in \bar{\mathcal{F}}_k, \quad k \in \mathcal{K}, \quad (10)$$

$$0 \leq r_{i,k} \leq D_L, \quad i \in \mathcal{I}_k, \quad k \in \mathcal{K}. \quad (11)$$

Therefore, we formulate the following optimization problem.

*Problem 1 (Total Utility Maximization):* For  $t = \text{pp, ip, up}$ ,

$$U^{(t)*} \triangleq \max_{\mathbf{R}, \mathbf{r}, \mathbf{d}, \mathbf{w}} Q^{(t)}(\mathbf{r}) \quad \text{s.t.} \quad (2), (5), (6), (8), (9), (10), (11).$$

Let  $(\mathbf{R}^{(t)*}, \mathbf{r}^{(t)*}, \mathbf{d}^{(t)*}, \mathbf{w}^{(t)*})$  denote an optimal solution of Problem 1.

Note that  $U^{(t)*}$  depends on  $D_L$  and is not related to  $D_1, \dots, D_{L-1}$ . The performance loss induced by the continu-

ous relaxation is acceptable when  $D_2 - D_1, \dots, D_L - D_{L-1}$  are small. Thus, we focus on solving Problem 1. Problem 1 is a nonconvex problem, due to the nonconvexity of the constraints in (8) and (9). There are no effective methods for solving a general nonconvex problem optimally. Besides, Problem 1 with  $t = \text{ip}$  and Problem 1 with  $t = \text{up}$  have non-differentiable objective functions. Therefore, Problem 1 is very challenging.<sup>1</sup>

Although it is difficult to obtain a globally optimal solution of the nonconvex problem in Problem 1, we can characterize its properties.<sup>2</sup> For all  $i \in \mathcal{I}_k, k \in \mathcal{K}$ , define  $\mathcal{T}_{i,k} \triangleq \mathcal{F}_i \setminus \overline{\mathcal{F}}_k$ . Note that  $\mathcal{T}_{i,k} \cap \mathcal{T}_{j,k} = \emptyset$ , for all  $i, j \in \mathcal{I}_k, i \neq j, k \in \mathcal{K}$ .

**Theorem 1 (Optimality Properties of Problem 1):** (i) For  $t = \text{pp}, \text{ip}, \text{up}$ ,  $R_{x,y,k}^{(t)*} = \max_{i \in \mathcal{I}_k: (x,y) \in \mathcal{F}_i} r_{i,k}^{(t)*}, (x,y) \in \overline{\mathcal{F}}_k, k \in \mathcal{K}$ . (ii) For  $t = \text{pp}$  and for all  $i, j \in \mathcal{I}_k, i \neq j, k \in \mathcal{K}$ , if  $p_{i,k} \leq p_{j,k}$ , and  $|\mathcal{T}_{i,k}| > |\mathcal{T}_{j,k}| > 1$ , then  $r_{i,k}^{(\text{pp})*} \leq r_{j,k}^{(\text{pp})*}$ . For  $t = \text{ip}$  and for all  $i, j \in \mathcal{I}_k, i \neq j, k \in \mathcal{K}$ , if  $\bar{p}_{i,k} \leq \bar{p}_{j,k}$ , and  $|\mathcal{T}_{i,k}| > |\mathcal{T}_{j,k}| > 1$ , then  $r_{i,k}^{(\text{ip})*} \leq r_{j,k}^{(\text{ip})*}$ . For  $t = \text{up}$ ,  $r_{i,k}^{(\text{up})*}, i \in \mathcal{I}_k$  are identical,  $k \in \mathcal{K}$ . (iii)  $U^{(\text{pp})*} \geq U^{(\text{ip})*} \geq U^{(\text{up})*}$ .

Statement (i) of Theorem 1 indicates that in each case, for all  $(x,y) \in \overline{\mathcal{F}}_k, k \in \mathcal{K}$ , the first inequality in (2) for at least one FoV that covers the  $(x,y)$ -th tile is active at an optimal solution. Statement (ii) of Theorem 1 indicates that in the case of perfect and imperfect FoV viewing probability distributions, an FoV with a higher viewing probability has a higher minimum encoding rate; and in the case of unknown FoV viewing probability distributions, the minimum encoding rates are identical, as all FoVs in  $\mathcal{I}_k$  are treated the same. Statement (iii) shows the relationship among the optimal values of Problem 1 for the three cases.

#### IV. SOLUTION

In this section, we obtain a suboptimal solution of Problem 1 in each case. First, we transform Problem 1 into an equivalent DC programming problem with a differentiable objective function in each case.

**Perfect FoV viewing probability distributions:** By introducing auxiliary variables and extra constraints, we can equivalently transform Problem 1 with  $t = \text{pp}$  into:

*Problem 2 (Equivalent Problem of Problem 1 with  $t = \text{pp}$ ):*

$$\max_{\mathbf{R}, \mathbf{r}, \mathbf{d}, \mathbf{e}, \mathbf{u}, \mathbf{w}} Q^{(\text{pp})}(\mathbf{r})$$

s.t. (2), (5), (6), (10), (11),

$$\sum_{k \in \mathcal{K}} d_{c,k} \leq \sum_{n \in \mathcal{N}} e_{c,n}, \quad (12)$$

$$d_{p,k} \leq \sum_{n \in \mathcal{N}} e_{k,n}, \quad k \in \mathcal{K}, \quad (13)$$

$$\sum_{j \in \mathcal{K}} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2 - \frac{\sum_{k \in \overline{\mathcal{K}}} |\mathbf{h}_{k,n}^H \mathbf{w}_{k,n}|^2 + \sigma^2}{u_{c,n}} \leq 0, \quad k \in \mathcal{K}, \quad n \in \mathcal{N}, \quad (14)$$

<sup>1</sup>The proposed method in [7] for obtaining a suboptimal rate splitting design cannot be used for solving Problem 1.

<sup>2</sup>The details are omitted due to page limitation. Please refer to [9] for the proof.

$$\sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2 - \frac{\sum_{j \in \mathcal{K}} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2}{u_{k,n}} \leq 0, \quad k \in \mathcal{K}, \quad n \in \mathcal{N}, \quad (15)$$

$$2^{\frac{e_{k,n}}{B}} \leq u_{k,n}, \quad k \in \overline{\mathcal{K}}, \quad n \in \mathcal{N}, \quad (16)$$

where  $\mathbf{e} \triangleq (e_{k,n})_{k \in \overline{\mathcal{K}}, n \in \mathcal{N}}$ ,  $\mathbf{u} \triangleq (u_{k,n})_{k \in \overline{\mathcal{K}}, n \in \mathcal{N}}$ . Let  $(\mathbf{R}^{(\text{pp})\dagger}, \mathbf{r}^{(\text{pp})\dagger}, \mathbf{d}^{(\text{pp})\dagger}, \mathbf{e}^{(\text{pp})\dagger}, \mathbf{u}^{(\text{pp})\dagger}, \mathbf{w}^{(\text{pp})\dagger})$  denote an optimal solution of Problem 2.

**Theorem 2 (Equivalence between Problem 1 with  $t = \text{pp}$  and Problem 2):**  $(\mathbf{R}^{(\text{pp})\dagger}, \mathbf{r}^{(\text{pp})\dagger}, \mathbf{d}^{(\text{pp})\dagger}, \mathbf{w}^{(\text{pp})\dagger})$  is an optimal solution of Problem 1 with  $t = \text{pp}$ .

**Imperfect FoV viewing probability distributions:**  $Q^{(\text{ip})}(\mathbf{r})$  is non-differentiable, and Problem 1 with  $t = \text{ip}$  is a max-min problem w.r.t.  $(\mathbf{R}, \mathbf{r}, \mathbf{d}, \mathbf{w}, \mathbf{p})$ . Note that max-min problems are in general very challenging. First, we replace the inner problem with its dual problem which has a differentiable objective function. Next, we introduce the auxiliary variables and extra constraints. Thus, we can equivalently convert Problem 1 with  $t = \text{ip}$  to:

*Problem 3 (Equivalent Problem of Problem 1 with  $t = \text{ip}$ ):*

$$\max_{\mathbf{R}, \mathbf{r}, \mathbf{d}, \mathbf{e}, \mathbf{u}, \mathbf{w}, \lambda \geq 0, \tau \geq 0, \gamma} \sum_{k \in \mathcal{K}} \left( \sum_{i \in \mathcal{I}_k} (\tau_{i,k} p_{i,k} - \lambda_{i,k} \bar{p}_{i,k}) - \gamma_k \right)$$

s.t. (2), (5), (6), (10), (11), (12), (13), (14), (15), (16),

$$U(r_{i,k}) + \lambda_{i,k} - \tau_{i,k} + \gamma_k \geq 0, \quad i \in \mathcal{I}_k, \quad k \in \mathcal{K},$$

where  $\lambda \triangleq (\lambda_{i,k})_{i \in \mathcal{I}_k, k \in \mathcal{K}}$ ,  $\tau \triangleq (\tau_{i,k})_{i \in \mathcal{I}_k, k \in \mathcal{K}}$ ,  $\gamma \triangleq (\gamma_k)_{k \in \mathcal{K}}$ . Let  $(\mathbf{R}^{(\text{ip})\dagger}, \mathbf{r}^{(\text{ip})\dagger}, \mathbf{d}^{(\text{ip})\dagger}, \mathbf{e}^{(\text{ip})\dagger}, \mathbf{u}^{(\text{ip})\dagger}, \mathbf{w}^{(\text{ip})\dagger}, \lambda^{(\text{ip})\dagger}, \tau^{(\text{ip})\dagger}, \gamma^{(\text{ip})\dagger})$  denote an optimal solution of Problem 3.

Note that  $\lambda_{i,k}$ ,  $\tau_{i,k}$  and  $\gamma_{i,k}$  are dual variables of the inner problem, corresponding to  $p_{i,k} \leq \bar{p}_{i,k}$ ,  $p_{i,k} \geq \underline{p}_{i,k}$  and  $\sum_{i \in \mathcal{I}_k} p_{i,k} = 1$ , respectively.

**Theorem 3 (Equivalence between Problem 1 with  $t = \text{ip}$  and Problem 3):**  $(\mathbf{R}^{(\text{ip})\dagger}, \mathbf{r}^{(\text{ip})\dagger}, \mathbf{d}^{(\text{ip})\dagger}, \mathbf{w}^{(\text{ip})\dagger})$  is an optimal solution of Problem 1 with  $t = \text{ip}$ .

**Unknown FoV viewing probability distributions:**  $Q_{\text{up}}(\mathbf{r})$  is non-differentiable. First, we cast Problem 1 with  $t = \text{up}$  in hypograph form which has a differentiable objective function. Next, we introduce the auxiliary variables and extra constraints. Thus, we can equivalently convert Problem 1 with  $t = \text{up}$  to:

*Problem 4 (Equivalent Problem of Problem 1 with  $t = \text{up}$ ):*

$$\max_{\mathbf{R}, \mathbf{r}, \mathbf{d}, \mathbf{e}, \mathbf{u}, \mathbf{y} \geq 0, \mathbf{w}} \sum_{k \in \mathcal{K}} y_k$$

s.t. (2), (5), (6), (10), (11), (12), (13), (14), (15), (16),

$$y_k \leq U(r_{i,k}), \quad i \in \mathcal{I}_k, \quad k \in \mathcal{K},$$

where  $\mathbf{y} \triangleq (y_k)_{k \in \mathcal{K}}$ . Let  $(\mathbf{R}^{(\text{up})\dagger}, \mathbf{r}^{(\text{up})\dagger}, \mathbf{d}^{(\text{up})\dagger}, \mathbf{e}^{(\text{up})\dagger}, \mathbf{u}^{(\text{up})\dagger}, \mathbf{y}^{(\text{up})\dagger}, \mathbf{w}^{(\text{up})\dagger})$  denote an optimal solution of Problem 4.

**Theorem 4 (Equivalence between Problem 1 with  $t = \text{up}$  and Problem 4):**  $(\mathbf{R}^{(\text{up})\dagger}, \mathbf{r}^{(\text{up})\dagger}, \mathbf{d}^{(\text{up})\dagger}, \mathbf{w}^{(\text{up})\dagger})$  is an optimal solution of Problem 1 with  $t = \text{up}$ .

Note that in Problem 2, Problem 3 and Problem 4,  $\mathbf{u}, \mathbf{e}$  are auxiliary variables, and (12), (13), (14), (15), (16) are extra

constraints. Furthermore, notice that the constraints in (12) and (13) are convex with respect to (w.r.t)  $(\mathbf{d}, \mathbf{e})$ , the constraints in (16) are convex w.r.t  $(\mathbf{e}, \mathbf{u})$ , and each constraint function in (14) and (15) can be regarded as a difference of two convex functions w.r.t.  $(\mathbf{u}, \mathbf{w})$ . Therefore, Problem 2, Problem 3 and Problem 4 are DC programming problems with differentiable objective functions, and their stationary points can be obtained by CCCP [10]. The main idea is to solve a sequence of successively refined approximate convex problems, each of which is obtained by linearizing the second convex part in (14), (15) and preserving the remaining convexity. For each problem, we can run CCCP multiple times with different feasible initial points to obtain multiple stationary points, and choose the stationary point with the best objective value as a suboptimal solution. Please refer to [9] for details.

## V. NUMERICAL RESULTS

In the simulation, we consider the streaming of five 360 VR video sequences, i.e., *Diving*, *Rollercoaster*, *Timelapse*, *Venice*, *Paris*, provided by [11]. They are indexed by 1, 2, 3, 4 and 5, respectively. As illustrated in Fig. 2, we divide each 360 VR video into  $8 \times 8$  tiles, i.e.,  $X = 8, Y = 8$ , and choose  $\bar{I} = 64$  FoVs, each of size  $3 \times 3$  (in the number of tiles). We use *Kvazaar* as the 360 VR video encoder, and set  $D_l, l \in \mathcal{L}$  according to Table I. For each video sequence, based on the viewpoint data of 59 users provided by [11], we obtain a viewpoint sequence for each user, with one viewpoint for every  $1/3$  second. We view users 2, 8, 24, 32 and 40 as the users who request videos 1, 2, 3, 4 and 5, i.e., users 1, 2, 3, 4 and 5, respectively. For video  $k$ , the viewpoint sequences from the other users besides user  $k$  are used for FoV prediction of user  $k$ . For user  $k$ , consider the prediction of the 3-rd element in his viewpoint sequence given the 2-nd element. Let  $i_k$  denote the index of the FoV of user  $k$  corresponding to the 2-nd element in his viewpoint sequence. Set  $\mathcal{I}_k = \{i_k - 8, i_k - 1, i_k, i_k + 1, i_k + 8\}$ , which contains  $i_k$  and the indices of the neighbouring FoVs of FoV  $i_k$ . Let  $n_{i_k, i}$  denote the number of users with the 2-nd element and 3-rd element in his viewpoint sequence for video  $k$  being  $i_k$  and  $i$ , respectively, where  $i \in \mathcal{I}_k$ . Then, we calculate the FoV viewing probabilities according to  $p_{i, k} = \frac{n_{i_k, i}}{\sum_{i \in \mathcal{I}_k} n_{i_k, i}}$ ,  $i \in \mathcal{I}_k, k \in \mathcal{K}$ . The values of  $p_{i, k}, i \in \mathcal{I}_k, k \in \mathcal{K}$  are given in Table II. We set  $\hat{p}_{i, k} = p_{i, k}, \varepsilon_{i, k} = \varepsilon, i \in \mathcal{I}_k, k \in \mathcal{K}, B = 1\text{MHz}, N = 128, P = 30\text{ dBm}$  and  $\sigma^2 = 10^{-9}\text{ W}$ . Besides, we choose  $\mathbf{h}_{k, n}, k \in \mathcal{K}, n \in \mathcal{N}$  randomly and independently according to  $\mathcal{CN}(0, \mathbf{1}_{M \times M})$ , and set the pass loss  $\beta_k = 1, k \in \mathcal{K}$ . We evaluate the average performance metrics over 100 random realizations of  $\mathbf{h}_{k, n}, k \in \mathcal{K}, n \in \mathcal{N}$ . We adopt the utility function presented in [12], i.e.,  $U(r) = 0.6 \log(1000 \frac{r}{D_L})$ . For ease of presentation, in the following, the performance metrics in all three cases are referred to as total utility, unless otherwise specified. In case- $t$ , the proposed solution of Problem 1 is referred to as Prop- $t$ , and the feasible solution of the original discrete problem, which is constructed based on Prop- $t$  (as illustrated in Section III), is referred to as Prop-Disc- $t$ , where  $t = \text{pp}, \text{ip}, \text{up}$ .

TABLE I. Encoding rates (in Mbit/s) for  $L = 3, 5, 7$ .

$L$	$D_l, l \in \mathcal{L}$
3	$D_1 = 14.46, D_2 = 52.97, D_3 = 87.75$ .
5	$D_1 = 14.46, D_2 = 37.10, D_3 = 52.97, D_4 = 69.53, D_5 = 87.75$ .
7	$D_1 = 14.46, D_2 = 37.10, D_3 = 46.20, D_4 = 52.97, D_5 = 59.45, D_6 = 69.53, D_7 = 87.75$ .

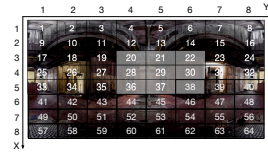


Fig. 2. Illustration of tiles, viewpoints and FoVs in a 360 VR video. The grey area represents FoV 29 which is centered at viewpoint 29.

We consider wireless streaming of the five 360 VR videos given in Table II to the five users, respectively. First, we investigate properties of the proposed solutions in the three cases. Fig. 3 (a) illustrates the total utility versus the number of quality levels  $L$ . Notice that the total utility of Prop- $t$  does not change with  $L$ , for all  $t = \text{pp}, \text{ip}, \text{up}$ . From Fig. 3 (a), we can see that in case- $t$ , the gap between the total utilities of Prop- $t$  and Prop-Disc- $t$  decreases with  $L$ . Furthermore, the gap is small when  $L$  is large, implying that the performance loss due to continuous relaxation is negligible at a large  $L$ . Fig. 3 (b) shows the worst average total utility versus the estimation error bound  $\varepsilon$ . Note that the worst average total utility of Prop-up is irrelevant to  $\varepsilon$ . We can see that in the case of imperfect FoV viewing probability distributions, the worst average total utility of Prop-ip is greater than those of Prop-pp and Prop-up, which reveals the importance of explicitly considering imperfectness of the predicted FoV viewing probability distributions in this case; and the worst-case average total utility of Prop-up is greater than that of Prop-pp when  $\varepsilon$  is large, as Prop-up is designed to maximize the worst-case total utility and does not depend on any information of FoV viewing probability distributions. Furthermore, the gain of Prop-ip over Prop-pp increases with  $\varepsilon$ , as it is more important to take into account of FoV prediction error when  $\varepsilon$  is large; and the gain of Prop-ip over Prop-up decreases with  $\varepsilon$ , as the imperfect FoV viewing probability distributions become less important when  $\varepsilon$  is large. Fig. 4 illustrates the heatmap of the encoding rates of all tiles for user 4 given by the proposed solutions in the three cases. From Fig. 4 (a) and (b), we can see that the encoding rates of the tiles in an FoV with a larger viewing probability are higher. From Fig. 4 (c), we can tell that the encoding rates of the tiles given by Prop-up are identical. Such observations are in accordance with the optimality properties in Statement (ii) of Theorem 1.

Next, we compare the total utilities of the proposed solution and three baseline schemes, namely BL-SDMA-OPT, BL-SDMA-ZF, BL-OFDMA-MRT, in each case. In BL-SDMA-OPT and BL-SDMA-ZF, SDMA is adopted. In BL-SDMA-OPT, beamforming optimization is considered on each sub-carrier, and a suboptimal solution of Problem 1 with  $d_{c, k} = 0, k \in \mathcal{K}$  and  $\mathbf{w}_{c, n} = 0, n \in \mathcal{N}$  (which is a DC programming) is obtained using CCCP. In BL-SDMA-ZF, zero forcing beamforming is adopted on each subcarrier, and an optimal solution

TABLE II. Prediction parameters.

$k$	Video sequence	User	Current FoV	Predicted FoVs	FoV viewing probability distributions
1	Diving	2	28	$\mathcal{I}_1 = \{20, 27, 28, 29, 36\}$	$(P_{20,1}, P_{27,1}, P_{28,1}, P_{29,1}, P_{36,1}) = (0.4138, 0.1724, 0.2414, 0.1667, 0.0417)$
2	Rollercoaster	8	21	$\mathcal{I}_2 = \{13, 20, 21, 22, 29\}$	$(P_{13,2}, P_{20,2}, P_{21,2}, P_{22,2}, P_{29,2}) = (0, 0.4615, 0.3077, 0.0769, 0.1538)$
3	Timelapse	24	24	$\mathcal{I}_3 = \{16, 23, 24, 17, 32\}$	$(P_{16,3}, P_{23,3}, P_{24,3}, P_{17,3}, P_{32,3}) = (0.1481, 0.037, 0.2963, 0.5185, 0)$
4	Venice	32	29	$\mathcal{I}_4 = \{21, 28, 29, 30, 37\}$	$(P_{21,4}, P_{28,4}, P_{29,4}, P_{30,4}, P_{37,4}) = (0.25, 0.375, 0.25, 0.0625, 0.0625)$
5	Paris	40	18	$\mathcal{I}_5 = \{10, 17, 18, 19, 26\}$	$(P_{10,5}, P_{17,5}, P_{18,5}, P_{19,5}, P_{26,5}) = (0.375, 0.5, 0.125, 0, 0)$

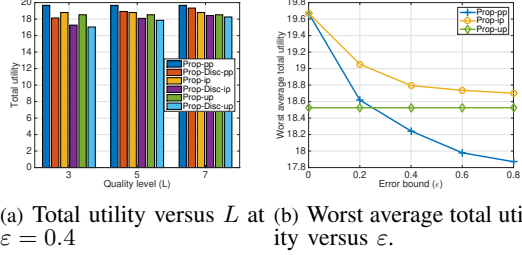


Fig. 3. Total utility comparison of the proposed solutions at  $M = 64$ .

of Problem 1 with the zero forcing beamformers (which is convex) is obtained by standard convex optimization methods. In BL-OFDMA-MRT, OFDMA is adopted, the maximum ratio transmission (MRT) is adopted for each user, and an optimal solution of Problem 1 with the MRT beamformers is obtained by continuous relaxation and the KKT conditions. Fig. 5 illustrates the total utility versus the number of transmit antennas  $M$ . From Fig. 5, we can see that in each case, the proposed solution outperforms the baseline schemes, and the total utility of each scheme increases with  $M$ . In each case, the gain of the proposed solution over BL-SDMA-OPT arises from the fact that the cost to suppress interference in BL-SDMA-OPT can be high, while the rate splitting scheme partially decodes interference and partially treats interference as noise; the gain of the proposed solution over BL-SDMA-ZF arises from the fact that the cost to suppress interference in BL-SDMA-ZF can be high and the beamforming directions are optimized; and the gain of the proposed solution over BL-OFDMA-MRT is due to effective spatial multiplexing.

Finally, from Fig. 3 (a) and Fig. 5, we can see that  $U^{(pp)\dagger} > U^{(ip)\dagger} > U^{(up)\dagger}$ , where  $U^{(t)\dagger}$  represents the total utility of Prop- $t$ , where  $t \in \{pp, ip, up\}$ . The relationship among the total utilities of the suboptimal solutions of Problem 1 in the three cases is the same as that of the optimal solutions in the three cases (which is shown in Statement (iii) of Theorem 1).

## VI. CONCLUSION

In this paper, we investigated optimal designs, based on rate splitting with successive decoding, for streaming multi-quality tiled 360 VR videos from a multi-antenna server to multiple single-antenna users in a multi-carrier system. We considered three cases of FoV viewing probability distributions and proposed three respective performance metrics for total utility. In each case, we optimized the rate adaptation, resource allocation and beamforming to maximize the total utility. We successfully transformed the challenging nonconvex

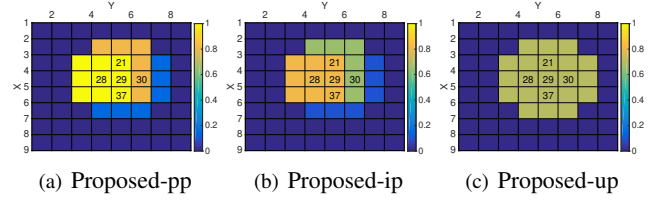


Fig. 4. Encoding rates of the tiles for user 4 given by Prop-pp, Prop-ip and Prop-up at  $M = 64$ ,  $\varepsilon = 0.4$ .

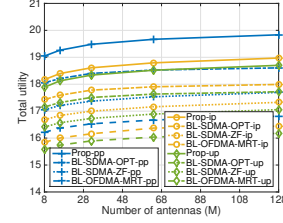


Fig. 5. Total utility comparison between the proposed solutions and baseline schemes at  $\varepsilon = 0.4$ .

problem in each case into a DC programming problem, and obtained a suboptimal solution using CCCP. Finally, numerical results demonstrated the proposed solutions achieve notable gains over existing schemes in all three cases, and revealed the impact of FoV prediction and its accuracy on streaming of multi-quality tiled 360 VR videos.

## REFERENCES

- [1] K. Long, C. Ye, Y. Cui, and Z. Liu, "Optimal multi-quality multicast for 360 virtual reality video," in *Proc. of IEEE GLOBECOM*, Dec. 2018, pp. 1–6.
- [2] K. Long, Y. Cui, C. Ye, and Z. Liu, "Optimal wireless streaming of multi-quality 360 VR video by exploiting natural, relative smoothness-enabled and transcoding-enabled multicast opportunities," major revision, *IEEE Trans. Multimedia*.
- [3] J. Zou, C. Li, C. Liu, Q. Yang, H. Xiong, and E. Steinbach, "Probabilistic tile visibility-based server-side rate adaptation for adaptive 360-degree video streaming," *IEEE JSTSP*, vol. 14, no. 1, pp. 161–176, Jan. 2020.
- [4] W. Huang, L. Ding, G. Zhai, X. Min, J.-N. Hwang, Y. Xu, and W. Zhang, "Utility-oriented resource allocation for 360-degree video transmission over heterogeneous networks," *Digit. Signal Process.*, vol. 84, pp. 1–14, Oct. 2018.
- [5] J. Yang, J. Luo, D. Meng, and J. Hwang, "QoE-driven resource allocation optimized for delay-sensitive VR video uploading over cellular network," in *Proc. of IEEE ISCC*, Jun. 2019, pp. 1–6.
- [6] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.
- [7] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: a rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov. 2016.
- [8] J. Chakareski, R. Aksu, X. Corbillon, G. Simon, and V. Swaminathan, "Viewport-driven rate-distortion optimized 360° video streaming," in *Proc. of IEEE ICC*, May. 2018, pp. 1–7.
- [9] L. Zhao, Y. Cui, C. Guo, Z. Liu and S. Yang "Optimal wireless streaming of multi-quality 360 VR videos with perfect, imperfect and unknown FoV viewing probabilities," *Technical Report*, 2020. [Online]. Available: <http://iwtc.sjtu.edu.cn/personal/yingcui/publicans.html>
- [10] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.
- [11] X. Corbillon, F. De Simone, and G. Simon, "360-degree video head movement dataset," in *Proc. of ACM MMSys*, June 2017, pp. 199–204.
- [12] W. Zhang, Y. Wen, Z. Chen, and A. Khisti, "QoE-driven cache management for HTTP adaptive bit rate streaming over wireless networks," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1431–1445, Oct. 2013.