Innovating Multi-user Volumetric Video Streaming through Cross-layer Design

Ding Zhang*, Bo Han*, Parth Pathak*, and Haoliang Wang**
*Department of Computer Science, George Mason University, VA, USA, **Adobe Research, USA
Email: {dzhang13, bohan, phpathak}@gmu.edu

ABSTRACT

Although existing work has demonstrated the feasibility of streaming volumetric content to a single user, there exist a number of appealing applications (*e.g.,* classroom education and collaborative design) that involve multiple concurrent users. In this position paper, we first perform a scaling experiment to demonstrate the challenges of streaming high-quality volumetric videos to multiple users and reveal the viewport-similarity opportunity that we can leverage to effectively optimize both network and computation resource utilization using multicast over mmWave. We then develop a holistic research agenda for improving the performance and quality of user experience for multi-user volumetric video streaming on commodity devices. Our proposed research includes joint viewport prediction and blockage mitigation for multiple users, intelligent multicast scheduling based on viewport similarity, customized mmWave beam design for efficient multicast, and mmWave-aware multi-user video rate adaptation. Finally, we discuss the open challenges of building a practical system with the proposed research roadmap.

1. INTRODUCTION

Volumetric video is an emerging 3D video format that enables users to explore multimedia content with six degrees of freedom (6DoF) motion, 3DoF of translational movement (X, Y, and Z) and another 3DoF of rotational movement (yaw, pitch, and roll). Different from the traditional 2D videos which are based on a fixed view, or the 360-degree videos which support only the 3DoF rotational movement, volumetric videos provide a more interactive and immersive experience for viewers. When used in augmented reality (AR) or virtual reality (VR), volumetric videos can enable many promising applications in entertainment, education, and healthcare. For example, people can track and view every movement of their favorite players during a sports event through moving around in 6DoF. Telepresence with volumetric videos can help students to attend classes remotely with an immersive learning experience. Therefore, it has been regarded as the key application in 5G and beyond [4, 21].

Although many of the applications of volumetric videos are specific to multi-user scenarios, limited attention has been paid to multi-user volumetric video streaming. The main reason is that volumetric video streaming is bandwidth-intensive

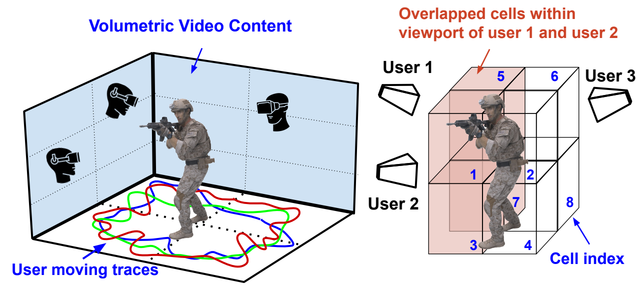


Figure 1: Multi-user volumetric video streaming with overlapped cells

and computation-intensive. Thus, streaming to a single user is already a challenging problem. Among the state-of-the-art volumetric video streaming systems, ViVo [12] exploits viewport, distance, and occlusion optimization strategies to save on average 40% bandwidth. As a result, it demands 100 to 200 Mbps data rate for a single user for point-cloud video frames with around 200K points. Transmitting high quality (high point density) frames and more complex scenes may make the bandwidth requirement even higher. GROOT [16] also requires 100 to 500 Mbps bandwidth for a single user after applying novel compression and GPU-assisted decoding schemes to reduce the computation overhead. However, in the multi-user scenario, the total bandwidth requirement increases linearly with the number of users, requiring Gbps bandwidth from the network.

The objective of this paper is to design a multi-user volumetric video system over mmWave WLANs. We show that while mmWave WLANs can provide sufficient bandwidth for supporting three to four users watching volumetric videos simultaneously, to provide high quality content and scale the system for more users, we need to leverage multicast and cross-layer design. This paper explores multiple research directions in these aspects. We list the related challenges and contributions below:

- There is limited understanding of the real performance of the multi-user volumetric video streaming system in the current WLAN. Therefore, we firstly measure the multi-user performance of state-of-the-art system ViVo [12] with 802.11ac and 802.11ad mmWave WLANs. We find that 1) the current 802.11ac WLAN cannot support multi-user volumetric video streaming even with the low quality video; 2) mmWave WLAN can support limited number of users but still need to be further optimized to support more users; 3) users share high viewport similarity when watching volumetric videos, which can be utilized in the multicast transmission to reduce the bandwidth requirement.

• In multi-user scenarios, the traditional viewport prediction methods cannot be directly used as the users might occlude the viewport of each other and affect the moving directions. We aim to build a model for joint viewport prediction for multi-users to incorporate their interactions. In addition, this multi-user viewport prediction could be exploited to estimate the blockage events in mmWave WLAN, which can enable a fast, cross-layer, proactive blockage mitigation.

• Building an efficient multicast system to stream volumetric video is challenging because we need a joint user grouping and scheduling scheme to guarantee the motion-to-photon latency for all users. In addition, the directional communication in mmWave WLAN can lead to unbalanced received signal strength (RSS) which makes the multicast transmission in mmWave WLAN even harder. To solve these problems, we propose a multicast grouping and scheduling scheme to group users with highest viewport similarity to improve the multicast performance. Besides, we employ a new custom beam design scheme to overcome the unbalanced RSS problem. Our preliminary results show that the new beam design can effectively improve the multicast transmission performance in mmWave WLANs.

• To adapt to the data rate fluctuations in mmWave WLAN, we design a cross-layer bandwidth prediction scheme by combining the data rate indicators from physical layer (blockage or mobility) and application layer (buffer size or throughput). The bandwidth prediction can guide the multicast scheduler to perform one of many possible actions (prefetching, video quality adaptation or beam switching) to offer high QoE for all users.

We are currently working on realizing the above ideas and integrating them into a holistic multi-user volumetric video streaming system for mmWave WLANs.

2. RELATED WORK

Multi-user AR, VR, and 360° Video Streaming. Recent work has started to address the technical issues of supporting multiple users for AR [19, 26, 37], VR [17, 18, 20] and 360° video streaming [9, 10, 24]. For example, SPAR [26] aims to reduce the spatial inconsistency of visual content and high initialization latency in multi-user AR scenarios by adapting the communicated information to the positions of virtual objects. Coterie [20] enables multi-user VR on commodity mobile devices by exploiting the similarity of background content in consecutive frames of the same user to reduce the bandwidth consumption. Using a simulation study, Bao *et al.* [9] demonstrated the effectiveness of using multicast to transmit the shared content in 360° videos to multiple users. To the best of our knowledge, we are the first to investigate the research challenges of multi-user volumetric video streaming by jointly considering the unique features of volumetric content and mmWave networks.

Volumetric Video Streaming. Existing work on volumetric video streaming has been centered around the single-user scenario [11, 12, 16, 25, 23, 35]. For example, ViVo [12] utilizes viewport, distance, and occlusion optimizations to reduce the bandwidth requirement of mobile volumetric video streaming. GROOT [16] presents a GPU-assisted point cloud compression scheme to optimize the decoding latency of volumetric content on mobile devices. VoluSR [35] is a recent proposal that leverages 3D super-resolution to upsample point clouds at the client side for improving user experience. Different from the above work, we focus on *multi-user volumetric video streaming* and propose a research agenda that benefits from cross-layer designs by jointly considering the unique features of mmWave networks and volumetric content.

mmWave WLAN. Blockage and mobility are two key technical challenges in mmWave networks. A variety of approaches have been proposed to tackle these problems at the mmWave physical/MAC layer [14, 27, 28, 30, 31, 36]. Moreover, mmWave links have been leveraged to replace the cables that connect high-end tethered VR devices and powerful graphics-rendering machines, by designing a mirror-like transceiver to address the blockage problem [8]. In terms of multicast over mmWave, Naribole *et al.* [22] proved its feasibility for highly directional 60-GHz WLANs. In contrast, we aim to address the practical challenges when utilizing mmWave multicast for bandwidth-hungry multi-user volumetric video streaming through a *cross-layer design* that takes advantage of the plethora of existing work.

3. MOTIVATION

In this section, we first identify the challenges of supporting multiple users in volumetric video streaming, by experimentally evaluating a prototype that extends ViVo [12], the state-of-the-art volumetric video streaming system, for the multi-user scenario. We then analyze the 6DoF viewport trajectories from an IRB-approved user study that reveal significant overlaps of volumetric content consumed by different users, which motivates our proposed research agenda.

Experimental setup. Our testbed consists of a content server that hosts multiple volumetric videos. The server is a machine running Ubuntu 18.04 and it also acts as an access point (AP). It has both 802.11ac and 802.11ad wireless interfaces. We use multiple laptops as the clients, which are equipped with Intel i7 CPU (4 cores at 2.8GHZ) and 8GB RAM. This configuration is comparable to state-of-the-art smartphones such as Samsung Galaxy S21 and is slightly better than the Magic Leap One mixed reality headset, which uses the NVIDIA Jetson TX2 hardware. We could not use these mobile devices directly for our experiments because they do not have an 802.11ad interface. Both the server and the clients have an 802.11ad card with Qualcomm QCA9500 chipset and a 32-elements phased array, which is supported by the wil6210 [7] driver.

We use the soldier volumetric video released in the 8i dy-

a) 550K points b) 430K points c) 330K points

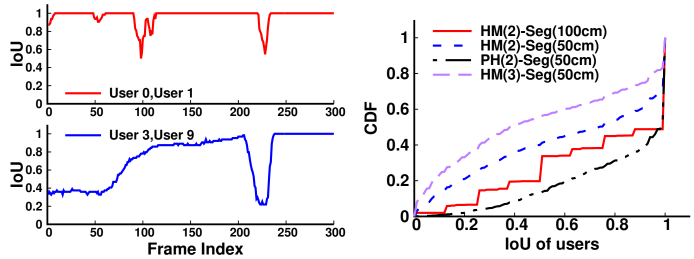
Figure 2: Soldier sequence with different qualities.

	Num. of Users	Per user data rate (Mbps)	FPS: Vanilla 330K points	FPS: Vanilla 430K points	FPS: Vanilla 550K points	FPS: Multi-user ViVo 330K points	FPS: Multi-user ViVo 430K points	FPS: Multi-user ViVo 550K points
ac	1	374	30	30	30	30	30	30
	2	180	21.5	17.4	14.1	30	28.5	21.9
	3	112	13.6	10.9	8.4	19.2	17.7	13.6
ad	1	1270	30	30	30	30	30	30
	2	575	30	30	30	30	30	30
	3	382	30	30	30	30	30	30
	4	298	30	29.3	21.8	30	30	30
	5	231	27.4	21.6	18.0	30	30	29.3
	6	175	19.8	16.5	13.2	30	27.5	21.2
	7	144	16.8	13.5	11.2	27.0	22.9	17.2

Table 1: Performance of multi-user volumetric video streaming with *vanilla* and ViVo [12] systems

namic voxelized point cloud dataset [3] as our video source. This video was captured at 30 FPS. Due to the huge size of point cloud data, we use the Draco library [2] from Google to compress this video and reduce its bandwidth consumption for streaming. We create three versions of this video with different visual qualities, by varying the numbers of points per frame. The version with the highest quality has, on average, 550K points per frame, which is the highest point density that can be decompressed by Draco at 30 FPS on the client laptops. Note that this point density doubles the highest density of the volumetric videos used in ViVo [12]. As shown in Fig. 2, these three point density levels represent videos with high, medium (430K points/frame), and low (330K points/frame) qualities. After the compression, the bitrate of these different versions ranges from 235 to 364 Mbps. We implement two video players on the client laptops. One is the vanilla system that fetches the entire point cloud for each video frame, and the other implements the viewport, occlusion, and distance optimizations in ViVo for reducing the data usage. Next, we benchmark both the vanilla system and ViVo with the above three visibility-aware optimizations when they are used for supporting multiple users, by measuring the maximum achievable frame rate.

Unicast over 802.11ac and 802.11ad for multiple users. To examine the multi-user volumetric video streaming performance, we deliver the same video content from the server to different number of concurrent users over either an 802.11ac or 802.11ad network with unicast. On our testbed, when



(a) (b)

Figure 3: Viewport similarity in multi-user volumetric video streaming (a) over time, and (b) for different device types, different partition sizes, and different group sizes.

serving a single user, the throughput is around 374 Mbps for 802.11ac and 1270 Mbps for 802.11ad. Table 1 shows the maximum frame rate (capped at 30 FPS) that can be achieved for different number of users when streaming the volumetric video with different point densities.

We have the following key observations from Table 1. First, when using 802.11ac, the vanilla system can support only one user at 30 FPS, regardless of the video quality. When multiple users watch the volumetric video simultaneously over the same 802.11ac network, the maximum achievable frame rate drops below 30 FPS, which deteriorates the user experience. ViVo can help only for the low-quality video, by adding one more user. Second, as the mmWave 802.11ad network at 60 GHz can achieve much higher throughput than 802.11ac (*e.g.,* higher than 1 Gbps without blockage), it can indeed increase the number of concurrent users to three for the medium and high quality videos and four for the low quality video. ViVo can further increase the number of users by one or two, depending on the video quality. Third, the maximum number of users that can be supported by ViVo at 30 FPS is only four for the high-quality video. However, use cases such as group-watching of sports events or AR-enhanced classroom teaching may involve more users. This limited number of users that can simultaneously watch high-quality volumetric content over an 802.11ac/802.11ad network with the state-of-the-art system motivates us to explore the opportunities for improving the performance and user experience of multi-user volumetric video streaming.

Inter-user viewport similarity. Viewport similarity has been leveraged for tile-based multi-user 360° video streaming, for which the tiles in the overlapped viewport of multiple users will be delivered using multicast [9]. However, given the 6DoF movement that is enabled by volumetric videos, it is unclear whether there still exists significant viewport similarity when multiple users consume the same volumetric video content in the 3D space. Thus, we collect the viewport trajectories of several volumetric videos from 32 participants of an IRB-approved user study. Those participants are diverse in terms of their gender (16 females), age (from 20 to 60), and education levels (bachelor, masters, and Ph.D.). We divide them into two groups, one using a smartphone and the other using a Magic Leap One headset to watch the videos using the player that we build. We denote the two groups as PH and HM, respectively. We collect the 6DoF viewport trajectories for all users at 30 Hz during the viewing sessions.

To calculate the 3D viewport similarity for volumetric video streaming between different users, we spatially partition the original point cloud into smaller "sub" regions called *cells*, which has been used in viewport-adaptive volumetric video streaming systems such as ViVo. Each cell is independently prefetchable and decodable. We partition the entire point cloud into cells of three sizes: 25×25×25 cm^3, 50×50×50 cm^3, and 100×100×100 cm^3. After the partition, we use frustum culling [6] to determine the cells overlapping with the 3D viewport based on the position and orientation of the user and calculate a *visibility map* that records the visible cells for each user. We then define the *viewport similarity* of a group of users as the intersection over union (IoU) of their visibility maps. For example, as shown in Fig. 1, if we segment the video content into 8 cells, cells 1, 3 and 5-8 are visible for User 1, and cells 1-4, 5, and 7 are visible for User 2. Thus, cells 1, 3, 5, and 7 will be needed for both users and their viewport similarity (IoU) for this frame is 0.5.

We analyze the viewport similarity among all 32 participants and make the following observations. First, there is a significant viewport overlap between users, which means there are indeed opportunities for using multicast to reduce the required bandwidth of multi-user volumetric video streaming. Fig. 3a shows how the IoU changes when two randomly selected users watch the same video (partitioned into 50×50×50 cm^3 cells). As we can see from this figure, User 0 and User 1 watch exactly the same content most of the time. For User 3 and User 9, although the viewport similarity is low initially, it increases to 1 towards the end of the video.

Second, viewport similarity is affected by many factors including the device type, the segmentation granularity, and the group size. Fig. 3b shows the CDF of the viewport similarity (IoU) among all users for different settings. With a headset (HM(2)-Seg(50cm)), there is less viewport similarity among users because they can move relatively more freely compared to watching the video on smartphones, which is shown by the PH(2)-Seg(50cm) curve. In terms of segmentation granularity (HM(2)-Seg(50cm) vs HM(2)-Seg(100cm)), the IoU between users decreases when the number of cells increases with finer-grained segmentation. Furthermore, the number of users that we consider also affects the viewport similarity. As shown in Fig. 3b, when calculating the viewport similarity for more users (HM(3)) in a group, the IoU decreases as more users bring larger variations in their positions and orientations.

Given the above promising preliminary result on 802.11ad networks and motivated by the significant viewport similarity among users, we aim to answer the following question in this paper: how to reduce the bandwidth requirement of multi-user volumetric video streaming by exploring mmWave multicast through cross-layer design? As we can see from Table 1, the bandwidth reduction can either lead to more concurrent users or improve the quality of experience (QoE) for a given number of users.

4. RESEARCH AGENDA

4.1 Multi-user viewport prediction and blockage mitigation

Joint viewport prediction for multi-user. Exploiting viewport similarity for multicast requires us to predict the viewport for all users in advance. Existing works such as [32, 12] have shown that individual user's 6DoF can be predicted using linear regression or multilayer perceptron with a high accuracy in realtime. In multi-user scenarios, to predict the viewport of all users together, it is intuitive to combine each of them at the AP's coordinate to create a holistic view of viewport mobility of all users. However, depending on volumetric video's usage in AR or VR, it is possible that one user's mobility and viewport affects other users. For example, in case of volumetric AR, one user can occlude the view of another user looking at the volumetric video content. In such cases, it is necessary to adapt the existing models of viewport prediction to incorporate occlusions. We will explore our user study and characterization results for developing joint multi-user viewport prediction models. This model will provide us a holistic view of viewport movement of all users by considering the user interaction also. It will further facilitate the mmWave channel estimation under blockages/mobility and the multicast transmission with the viewport similarity.

Proactive blockage prediction and mitigation. mmWave link blockages can significantly reduce the link data rate and can even cause a complete outage of the link. Reinitiating beam searching to find new beams to go across the blockages will cause a delay upto 5 to 20ms. This added latency can cause video stalls and reduction in video quality which adverserly affect user's QoE. Prior work such as [30, 33] have proposed to use in-built sensors to predict user's mobility and adapt mmWave beams accordingly. However, in our multi-user scenario, one user can block the mmWave link of another user and it is not sufficient to adapt the mmWave beams to each user separately.

We will leverage our multi-user viewport prediction to build a cross-layer proactive beam adaptation and blockage mitigation scheme. Here, the viewport predictions can be leveraged for a limited beamsearching where only a small number of beams can be searched for mmWave beam adaptation. The holistic view of the multi-user viewport prediction available at the AP will be used to infer possible blockages between users and proactively mitigate them to avoid any adverse effects on QoE. Afterwards, the AP can estimate the data rate change under the potential blockages for each link and can take one of many possible actions at different layers. It can prefetch the content and schedule the future cells in current time slot so that when the user faces the blockage, it already has prefetched cells for rendering. In addition, the AP can also use the blockage prediction to adapt its beam to the user without beam searching so that the blockage can be mitigated or its effects on QoE can be reduced.

4.2 Multicast grouping and scheduling with viewport similarity

Based on the multi-user viewport prediction, we can get the potential viewport similarity among users for multicast transmission. To guarantee the QoE of users, we need to intelligently schedule the multicast transmission with unicast transmission by selecting a multicast group with high viewport similarity to reduce the required bandwidth.

In our multicast schedule, we first calculate the viewport similarity (IoU) among all possible groups based on the multi-user viewport prediction and order them from the high IoU to low IoU. We then estimate the corresponding transmission time of each group to find the group which can be transmitted within the motion-to-photon time constraint. To estimate the transmission time of a specific user or a multicast group, we need an accurate rate estimation scheme. In this work, we propose a cross-layer rate estimation scheme combing the physical layer channel estimation and application layer buffer size information, which will be discussed later.

Here, let us assume the corresponding data rate for user i is r_i. Since the frame size S is known at the server before streaming each frame, we can estimate the transmission time of each frame for each user i as $t_i = S/r_i$. If we use unicast to transmit the frame to all users one by one, the deadline of transmitting the frame is the summation of the transmission time for all users $T = \sum_i^N t_i$ where N is the total number of users watching the volumetric video. Assuming all users need to be supported at 30 FPS, $T \leq 33$ ms to guarantee that there is no stall for all users.

In terms of the multicast, we can estimate the transmission time as

$$T^m(\mathbf{k}) = S^m_{\mathbf{k}}/r^m + \sum_i^N (S - S^m_{\mathbf{k}})/r_i \tag{1}$$

where $S^m_{\mathbf{k}}$ is the size of the overlapped cells of user group $\mathbf{k}$ (a set of users), which can be calculated based on the users' 6DoF information, and r^m is the multicast data rate which is the output of our rate estimation scheme. In Equ. 1, the first part is the time to transmit the overlapped cells with multicast while the second part is the transmission time to stream the remaining requested cells with default unicast among all N users who are requesting the same frame.

To schedule the multicast transmissions, we want to find the correct balance between mixing multicast and unicast by finding the group $\mathbf{k}$ that maximizes the total throughput with the constraint of $T^m(\mathbf{k}) <= 1/F$ where the F is the required frame rate. Specifically, the throughput improvement made possible by multicast can be used to increase the transmitted frame size S (i.e., improve the video quality) or to serve more users watching the volumetric video.

4.3 Multicast with customized beam design

A key challenge with multicast is that the multicast data rate r^m for a group is limited by the lowest achievable modulation coding scheme (MCS) among all members of the group. Using this MCS guarantees a reliable multicast that can be received by all members. In case of mmWave WLANs, it is challenging to cover multiple users of a group using directional beams. If sufficiently high RSS cannot be guaranteed for all members of the group, multicast can provide even worse performance than unicast.

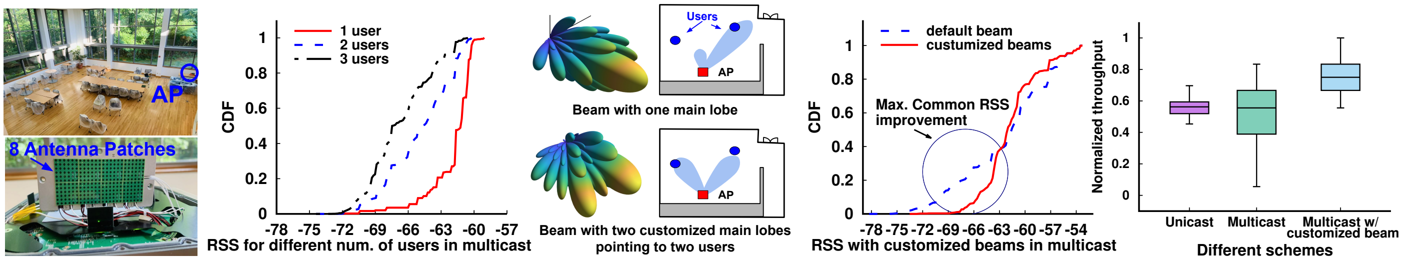
To understand the performance of multicast in mmWave WLANs, we first investigate the effectiveness of the default beam codebook used in the commercial 802.11ad devices. We use the multi-user traces collected in our user study and measure the RSS for all users in our mmWave WLAN testbed. Our mmWave testbed includes an 802.11ad router from Airfide [1] with 8 phased antenna array patches (shown in Fig. 4a) and multiple Acer laptops with 802.11ad interfaces. We modify the open-source 802.11ad driver on the laptop to extract the RSS, MCS and beamforming information to user space. Fig. 4b shows the CDF of maximum RSS that could be supported by the default codebook for multicast groups of different sizes. We find that using the default codebook, RSS of -68 dBm, which can provide approximately 384 Mbps data rate necessary for 550K points quality, can be provided for 96.5% positions for one user (no multicast), but only 79% and 60% positions for two and three users, respectively, with multicast. This is because the default codebook beams are not specifically designed to support multicast and cannot guarantee high RSS to all members in multicast groups.

To address the problem of poor and unbalanced RSS in multicast provided by the default codebook, we aim to design a customized beams that can provide high RSS to all users of a given multicast group. The main idea here is to create the customized beams by combining the antenna weight vectors of beams pointing to individual users while constraining the total transmit power. Assume the antenna weight vector used for transmitting to User 1 is $\mathbf{w}_1$ and RSS is Δ_1, and $\mathbf{w}_2$ and Δ_2 are the weight vector and RSS for User 2, we define the combined antenna weight vector for the new beam as

$$\mathbf{w} = \frac{1}{\Delta_1 + \Delta_2}(\Delta_2 \mathbf{w}_1 + \Delta_1 \mathbf{w}_2) \tag{2}$$

The objective here is to design beams that can not only cover the users of a group with high viewport similarity but can also provide high RSS and higher common MCS supported by all users. We note that we can directly use the predicted 6DoF information at the server to select the individual beams and combined beams for the AP without beam searching. Besides, our custom beam design just requires the RSS value to normalize the antenna weights, instead of the complete channel state information (CSI) as in [29, 14] because the separated users have independent receive chain.

To investigate the effectiveness of the proposed beam design, we run the multicast of 2 users case with our custom beams and default beams in a commercial mmWave channel simulator Remcom [5]. As shown in Fig. 4d, by using the new combined weight, the two users can achieve much higher common RSS values(in the circle), leading to



(a) (b) (c) (d) (e)

Figure 4: (a) mmWave WLAN measurement setup (b) The default beams cannot support an efficient multicast for multi-user. (c-d) With the customized beam design in multicast, we can efficiently transmit the overlapped cells between users. (e) Preliminary performance of the multicast system with customized beams

higher common MCS. We also notice that when both users have high RSS, we should directly use the common beam which has already covered the users. Lastly, we compare the unicast, multicast with default beams, and multicast with custom multi-lobe beams in terms of normalized throughput with 2 users case(Fig.4e). We find that multicast with the default beams cannot always improve the network data rate but can in fact sometimes even reduce the data rate. This is due to the unbalanced RSS which reduces the MCS that can be supported by the multicast group. However, using multicast with our customized beams can effectively increase the data rate and the performance gains can be higher with more number of users in multicast group.

4.4 Cross-layer video rate adaptation

To offer better QoE for multi-user volumetric video streaming, we propose a cross-layer rate adaption scheme. We identify two important issues when exploring the design space. First, how to accurately estimate the link bandwidth (the data rate r_i and r^m in Equ. 1) for unicast and multicast transmissions? In the multi-user scenario, an inaccurate bandwidth estimation will affect the quality of experience for all users in the same multicast group. To better adapt to network bandwidth changes, we will propose a cross-layer solution that combines the mmWave channel information (*e.g.,* RSS) with the application layer information such as the buffer size of video player [13]. Second, how to judiciously react to the bandwidth fluctuation caused by the blockage of mmWave links? The possible reactions include: 1) prefetching more cells for a priority group of users with low predicted bandwidth; 2) regrouping the multicast and unicast transmissions to find better viewport overlaps to save more bandwidth; 3) generate a new beam to cover a different physical path to improve the channel quality. Our rate adaptation scheme will intelligently select different solutions according to the mmWave link status.

5. OPEN CHALLENGES

Multi-user viewport prediction and blockage mitigation. One important challenge of using multi-user viewport prediction to estimate the blockage and link bandwidth is that blockage does not always cause link outage [28]. It is still possible to operate the link at a degraded performance level. This requires developing a model to quantify the degradation of the link performance with respect to different levels of blockage and integrating it to the multicast scheduler to offer a high QoE for all users.

Multicast with customized beams. Even though we have shown the effectiveness of multicast with customized beams, there are a few challenges in the design on COTS devices. First, multi-lobe beams can cause interference between users of the same group due to reflections from nearby objects. This means that we need to probe the newly designed beam to verify that it will not cause too much interference before using it. Second, the imperfections in the antenna hardware of COTS devices can produce non-negligible sidelobes that can reduce the effectiveness of custom beams. Lastly, the MAC layer multicast implementation needs heavy engineering of the driver or even the firmware.

Cross-layer video rate adaptation. In contrast to traditional video rate adaptation algorithms that are executed by the players on the client side [15, 34], our proposed algorithm for multi-user volumetric video streaming should be realized at a central point (*e.g.,* on an edge server) and has several unique challenges. A key issue is that the cross-layer approach to choose the proper video encoding bitrates for multiple users leads to a much larger design space when formulating the optimization problem, compared to the conventional application-layer schemes that consider only a single user. For example, in a multi-user scenario, it is common that different users may have heterogeneous hardware settings with diverse computation capabilities. Two users in the same multicast group could have mobile devices with different decoding performance, and one may experience a significantly longer decoding latency than the other. As a result, our rate adaptation algorithm needs to consider such heterogeneity in selecting the video quality for a group of users.

6. CONCLUSION

In this position paper, we first show the challenges of using current 802.11 networks to stream volumetric videos to multiple concurrent users. Motivated by the 3D viewport similarity that we identify from a trajectory trace with 32 users, we then propose a holistic research agenda by leveraging cross-layer design to improve the performance and user experience of multi-user volumetric video streaming over customized mmWave multicast. We conclude the paper with several remaining open challenges for building a practical system, which we hope could stimulate more discussion and research on this topic.

7. REFERENCES

[1] Airfide afn2200. https://airfidenet.com/.

[2] Google draco library. https://github.com/google/draco.

[3] Jpeg pleno database: 8i voxelized full bodies (8ivfb v2) - a dynamic voxelized point cloud dataset. http://plenodb.jpeg.org/pc/8ilabs.

[4] Next generation display media in 6g. https://research.samsung.com/next-generation-display-media.

[5] Remcom Wireless InSite 3D Wireless Prediction Software. https://www.remcom.com/wireless-insite-em-propagation-software.

[6] View frustum culling. http://www.lighthouse3d.com/tutorials/view-frustumculling/.

[7] wil6210 module. https://git.kernel.org/pub/scm/linux/kernel/git/kvalo/ath.git/tree/drivers/net/wireless/ath/wil6210.

[8] O. Abari, D. Bharadia, A. Duffield, and D. Katabi. Enabling high-quality untethered virtual reality. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*, pages 531–544, Boston, MA, Mar. 2017. USENIX Association.

[9] Y. Bao, T. Zhang, A. Pande, H. Wu, and X. Liu. Motion-prediction-based multicast for 360-degree video transmissions. In *2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 1–9, 2017.

[10] J. Chen, X. Qin, G. Zhu, B. Ji, and B. Li. Motion-Prediction-based Wireless Scheduling for Multi-User Panoramic Video Streaming. In *Proceedings of IEEE INFOCOM*, 2021.

[11] S. Gül, D. Podborski, J. Son, G. S. Bhullar, T. Buchholz, T. Schierl, and C. Hellge. Cloud Rendering-based Volumetric Video Streaming System for Mixed Reality Services. In *Proceedings of ACM MMSys*, 2020.

[12] B. Han, Y. Liu, and F. Qian. ViVo: Visibility-Aware Mobile Volumetric Video Streaming. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, MobiCom '20, New York, NY, USA, 2020. Association for Computing Machinery.

[13] T.-Y. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson. A buffer-based approach to rate adaptation: Evidence from a large video streaming service. In *Proceedings of the 2014 ACM Conference on SIGCOMM*, SIGCOMM '14, page 187–198, New York, NY, USA, 2014. Association for Computing Machinery.

[14] I. K. Jain, R. Subbaraman, and D. Bharadia. Two beams are better than one: Enabling reliable and high throughput mmwave links, 2021.

[15] J. Jiang, V. Sekar, and H. Zhang. Improving Fairness, Efficiency, and Stability in HTTP-Based Adaptive Video Streaming with FESTIVE. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '12, page 97–108, New York, NY, USA, 2012. Association for Computing Machinery.

[16] K. Lee, J. Yi, Y. Lee, S. Choi, and Y. M. Kim. GROOT: A Real-Time Streaming System of High-Fidelity Volumetric Videos. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, MobiCom '20, New York, NY, USA, 2020. Association for Computing Machinery.

[17] Y. Li and W. Gao. MUVR: Supporting Multi-User Mobile Virtual Reality with Resource Constrained Edge Cloud. In *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, pages 1–16, 2018.

[18] X. Liu, C. Vlachou, F. Qian, C. Wang, and K.-H. Kim. Firefly: Untethered multi-user VR for commodity mobile devices. In *2020 USENIX Annual Technical Conference (USENIX ATC 20)*, pages 943–957. USENIX Association, July 2020.

[19] Z. Liu, G. Lan, J. Stojkovic, Y. Zhang, C. Joe-Wong, and M. Gorlatova. CollabAR: Edge-assisted Collaborative Image Recognition for Mobile Augmented Reality. In *Proceedings of ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2020.

[20] J. Meng, S. Paul, and Y. C. Hu. Coterie: Exploiting frame similarity to enable high-quality multiplayer vr on commodity mobile devices. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '20, page 923–937, New York, NY, USA, 2020. Association for Computing Machinery.

[21] A. Narayanan, E. Ramadan, J. Carpenter, Q. Liu, Y. Liu, F. Qian, and Z.-L. Zhang. A first look at commercial 5g performance on smartphones. In *Proceedings of The Web Conference 2020*, WWW '20, page 894–905, New York, NY, USA, 2020. Association for Computing Machinery.

[22] S. Naribole, E. Knightly, S. Naribole, and E. Knightly. Scalable multicast in highly-directional 60-ghz wlans. *IEEE/ACM Trans. Netw.*, 25(5):2844–2857, Oct. 2017.

[23] J. Park, P. A. Chou, and J.-N. Hwang. Rate-Utility Optimized Streaming of Volumetric Media for Augmented Reality. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1):149–162, 2019.

[24] C. Perfecto, M. S. Elbamby, J. D. Ser, and M. Bennis. Taming the Latency in Multi-User VR 360°: A QoE-Aware Deep Learning-Aided Multicast Framework. *IEEE Transactions on Communications*, 68(4):2491–2508, 2020.

[25] F. Qian, B. Han, J. Pair, and V. Gopalakrishnan. Toward practical volumetric video streaming on commodity smartphones. In *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications*, HotMobile '19, page 135–140, New York, NY, USA, 2019. Association for Computing Machinery.

[26] X. Ran, C. Slocum, Y.-Z. Tsai, K. Apicharttrisorn, M. Gorlatova, and J. Chen. Multi-user augmented reality with communication efficient and spatially consistent virtual objects. In *Proceedings of the 16th International Conference on Emerging Networking EXperiments and Technologies*, CoNEXT '20, page 386–398, New York, NY, USA, 2020. Association for Computing Machinery.

[27] S. Sur, I. Pefkianakis, X. Zhang, and K.-H. Kim. Towards scalable and ubiquitous millimeter-wave wireless networks. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, MobiCom '18, page 257–271, New York, NY, USA, 2018. Association for Computing Machinery.

[28] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra. BeamSpy: Enabling Robust 60 GHz Links under Blockage. In *Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation*, NSDI'16, page 193–206, USA, 2016. USENIX Association.

[29] S. Wang, J. Huang, X. Zhang, H. Kim, and S. Dey. X-array: Approximating omnidirectional millimeter-wave coverage using an array of phased arrays. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, MobiCom '20, New York, NY, USA, 2020. Association for Computing Machinery.

[30] T. Wei and X. Zhang. Pose Information Assisted 60 GHz Networks: Towards Seamless Coverage and Mobility Support. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, MobiCom '17, page 42–55, New York, NY, USA, 2017. Association for Computing Machinery.

[31] T. Wei, A. Zhou, and X. Zhang. Facilitating Robust 60 GHz Network Deployment by Sensing Ambient Reflectors. In *Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation*, NSDI'17, page 213–226, USA, 2017. USENIX Association.

[32] T. Xu, B. Han, and F. Qian. Analyzing Viewport Prediction under Different VR Interactions. In *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies*, CoNEXT '19, page 165–171, New York, NY, USA, 2019. Association for Computing Machinery.

[33] Z. Yang, P. H. Pathak, Y. Zeng, and P. Mohapatra. Sensor-assisted codebook-based beamforming for mobility management in 60 ghz wlans. In *MASS '15*.

[34] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP. In *Proceedings of ACM SIGCOMM*, 2015.

[35] A. Zhang, C. Wang, B. Han, and F. Qian. Efficient volumetric video streaming through super resolution. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, HotMobile '21, page 106–111, New York, NY, USA, 2021. Association for Computing Machinery.

[36] D. Zhang, M. Garude, and P. H. Pathak. MmChoir: Exploiting Joint Transmissions for Reliable 60GHz MmWave WLANs. In *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Mobihoc '18, page 251–260, New York, NY, USA, 2018. Association for Computing Machinery.

[37] W. Zhang, B. Han, P. Hui, V. Gopalakrishnan, E. Zavesky, and F. Qian. CARS: Collaborative Augmented Reality for Socialization. In *Proceedings of International Workshop on Mobile Computing Systems and Applications (HotMobile)*, 2018.