Multi-path Multi-tier 360-degree Video Streaming in 5G Networks*

Liyang Sun
New York University
Brooklyn, NY
ls3817@nyu.edu

Fanyi Duanmu
New York University
Brooklyn, NY
fanyi.duanmu@nyu.edu

Yong Liu
New York University
Brooklyn, NY
yongliu@nyu.edu

Yao Wang
New York University
Brooklyn, NY
yw523@nyu.edu

Yinghua Ye
Huawei Technologies
Santa Clara, CA
Yinghua.Ye@huawei.com

Hang Shi
Huawei Technologies
Santa Clara, CA
Hang.Shi@huawei.com

David Dai
Huawei Technologies
Santa Clara, CA
david.h.dai@huawei.com

ABSTRACT

360° video streaming is a key component of the emerging Virtual Reality (VR) and Augmented Reality (AR) applications. In 360° video streaming, a user may freely navigate through the captured 360° video scene by changing her desired Field-of-View. High-throughput and low-delay data transfers enabled by 5G wireless networks can potentially facilitate untethered 360° video streaming experience. Meanwhile, the high volatility of 5G wireless links present unprecedented challenges for smooth 360° video streaming. In this paper, novel multi-path multi-tier 360° video streaming solutions are developed to simultaneously address the dynamics in both network bandwidth and user viewing direction. We systematically investigate various design trade-offs on streaming quality and robustness. Through simulations driven by real 5G network bandwidth traces and user viewing direction traces, we demonstrate that the proposed 360° video streaming solutions can achieve a high-level of Quality-of-Experience (QoE) in the challenging 5G wireless network environment.

CCS CONCEPTS

• **Networks** → **Mobile networks**; *Application layer protocols*; • **Computing methodologies** → *Virtual reality*;

KEYWORDS

5G, 360° Video, Multipath, Video Streaming

*Produces the permission block, and copyright information

MMSys'18, June 12–15, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5192-8/18/06…$15.00
https://doi.org/10.1145/3204949.3204978

ACM Reference Format:
Liyang Sun, Fanyi Duanmu, Yong Liu, Yao Wang, Yinghua Ye, Hang Shi, and David Dai. 2018. Multi-path Multi-tier 360-degree Video Streaming in 5G Networks. In *MMSys'18: 9th ACM Multimedia Systems Conference, June 12–15, 2018, Amsterdam, Netherlands,* Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/3204949.3204978

1 INTRODUCTION

Virtual Reality (VR) and Augmented Reality (AR) technologies have become very popular recently. Many VR/AR applications are being rapidly commercialized in different sectors, including movie and gaming, education and training, healthcare and real estate, advertising and social media, etc. Many VR/AR applications involve streaming of 360° video. Therefore, the delivery of ultra high quality 360° video is critically important for the wide adoption of VR/AR. Compared with the traditional video streaming, 360° video streaming confronts unique new challenges. Firstly, to deliver an immersive VR experience, 360° video has much higher bandwidth requirement. A premium quality 360° video with 120 frames-per-second and 24K resolution can easily consume a bandwidth of multiple Gigabits-per-second (Gbps) [10]. For smooth rendering, video has to be streamed consistently at high rate. Meanwhile, 360° video streaming is constantly driven by user Field-of-View (FoV) changes: at any given time a user only watches a video scene within a FoV centered at certain direction and with limited horizontal and vertical spans; a user can change her FoV at any time, and she wants to the see the video scene in the new FoV immediately after her head movement. Recent subjective user study has suggested that if the video rendering latency after a FoV change, the so-called *Motion-to-Photon (MTP) latency*, is above twenty milliseconds, users will experience motion sickness [10]. This imposes stringent latency requirement for 360° video delivery.

Millimeter wave (mmWave) communications, especially operating at frequencies between 10 and 300 GHz, provide great potential for the next-generation 5G wireless networks to meet the surging

data demand of 360° video streaming. There are orders of magnitude more available spectrum in the mmWave band, where multiple gigabit-per-second data rates can be achieved [3, 16]. While 5G is capable of transmitting at speeds over multiple gigabits-per-second at the air interface, it has yet to be proven to consistently support high-throughput low-delay 360° video streaming. mmWave links are unique in that they have high bandwidth but can suffer sudden drops in bandwidth. The Friis equation [18] states that the free space loss grows with the square of the frequency. mmWave signals are also extremely susceptible to shadowing. Materials such as brick can attenuate signals by as much as 40 to 80 dB [24], and the human body itself can result in a 20 to 35 dB loss [14]. A consequence of shadowing in the mmWave regime is that the appearance of obstacles can lead to blocking with much more dramatic swings in the channel gain than in the sub-6GHz band. Thus, from a systems perspective, mmWave link bandwidths will be highly volatile. It is therefore critical that 360° video streaming protocols over 5G networks are rapidly adaptable to deliver a high level of user QoE.

In this paper, we propose novel multi-path multi-tier 360° *video streaming designs to fully explore the potential of 5G wireless networks to maximize the rendered video quality, while maintaining the streaming continuity and robustness against the inherent dynamics in both user FoV and 5G wireless links.* In the proposed framework, 360° video is encoded at multiple tiers: base-tier (BT) full-360° video chunks are coded with a basic rate and pre-fetched in a long display buffer to compensate for bandwidth variations, whereas the enhancement-tier (ET) video chunks are smaller view window coded with multiple rates and pre-fetched in a shorter buffer to ensure view prediction accuracy; correction-tier (CT) chunks code video in small view tiles that can be used to patch a pre-fetched view window at the predicted view direction to cover the actual user view window to be rendered. Video chunks from different tiers will be delivered as different sub-streams, which compete for the available network bandwidth. We study multi-tier rate allocation schemes and chunk scheduling algorithms to achieve the optimal trade-off between video quality and streaming robustness. In heterogeneous wireless networks, there are multiple network paths between the streaming server and the client, with each path having diverse throughput, latency and stability characteristics, as well as different cost and pricing structures, we study multi-path multi-tier video streaming to trade off the delivered user QoE and the incurred data communication cost. Our main contributions are three-fold.

(1) We analytically and experimentally study the optimal multi-tier rate allocation for 360° video streaming in 5G networks. Our study brings forth important understanding about the interplay between the key components of 360° video streaming, including FoV prediction accuracy, coding rate allocation, chunk pre-fetching strategy, and chunk delivery ratio.
(2) Exploiting high-throughput low-latency 5G data transfer, we propose novel tile-based FoV correction and chunk retransmission schemes to address FoV prediction errors and late chunk deliveries. For heterogeneous WiFi and 5G networks, we demonstrate that 5G channels can be utilized to significantly improve streaming quality over WiFi in a cost-effective manner.
(3) Extensive experiments are conducted to thoroughly investigate various design trade-offs. All experiments are driven by real 5G 802.11ad bandwidth traces with different levels of volatilities and real user FoV traces with diverse head movement patterns. The obtained insights can be used to guide the design of future 360° video streaming systems in 5G networks.

The rest of the paper is organized as follows. Background and related work on 360° are reviewed in Section 2. The optimal rate allocation for two-tier streaming in 5G networks is studied in Section 3. Tile-based FoV correction scheme is introduced and evaluated in Section 4. Multi-path multi-tier streaming for heterogenous WiFi and 5G networks is presented in Section 5. The paper is concluded in Section 6.

2 BACKGROUND AND RELATED WORK

In recent years, several solutions have been proposed to address 360° video streaming and can be categorized into the following three categories:

Category 1: 360° **Video Source Representation.** Videos captured from different view angles first need to be projected to a 2D plane before further processing. Facebook proposed the cube-map [11] and pyramid [12] projection methods and encoding schemes in 2016, to specifically address on-demand 360° video streaming, with 25% and 80% compression improvements reported, respectively. The Joint Video Exploration Team (JVET) also proposed a few projection solutions, including Icosahedral projection (ISP) [25], Segmented Sphere Projection (SSP) [23], Truncated Square Pyramid Projection (TSP) [4], Octahedron Projection (OHP) [13], Hybrid Cubemap Projection [5], etc.

Category 2: Source Bit Allocation. Different view regions have different perceptual quality implications, consequently deserve different numbers of coding bits. In [2], a region-adaptive smoothing scheme is proposed to reduce the bitrate spent within the polar regions of equi-rectangular 360° videos through Gaussian filtering. A 20% bitrate reduction is reported with unnoticeable perceptual quality degradation.

Category 3: Tile-based Streaming. A few tile-based solutions were proposed in recent years, in which increased video quality and streaming priority are assigned to tiles within regions of interest (ROI), whereas over non-ROI regions, reduced quality or priority are assigned (e.g., in the extreme case, non-ROI tiles are not even streamed at all). In [20], a few tile-based encoding and streaming solutions are proposed, including scalable coding scheme and simulcast coding scheme. Video tiles that cover the whole 360° scene are coded in multiple rates. Depending on the FoV, tiles within or close to the predicted FoV are fetched with higher bitrate while tiles far away from the predicted FoV are fetched with lower bitrate. In [17], a view prediction based framework is proposed by only fetching the video portions desirable to the end user to reduce the bandwidth consumption. A dynamic video chunk adaptation scheme is implemented to adjust tile coverage based on the view prediction accuracy. An estimated 80% maximum rate reduction is reported without considering the coding efficiency loss due to video tiling and bandwidth variations. Similar tile-based solutions can be found in [22] [9] [15], etc.

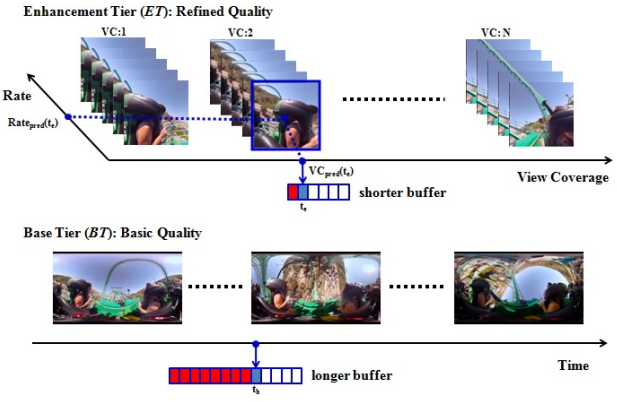


Figure 1: Two-tier 360° Video Streaming System

Algorithm 1 Two-tier 360° Video Streaming

Initialization at $t = 0$;
while (One chunk downloading is finished **or** $t = 0$) **and** display is not terminated **do**
  if $B_b <= B_b^{(T)}$ **then**
    Download next C_b;
    $t \leftarrow t + \Delta_b$;
  else
    if $B_e <= B_e^{(U)}$ **then**
      Predict bandwidth BW_t
      Refine BW_t with P-I controller
      Predict FoV P_e for next C_e;
      Request for next C_e;
      $t \leftarrow t + \Delta_e$
    else
      $t \leftarrow t + \delta$;
    end if
  end if
end while
return

3 TWO-TIER STREAMING IN 5G NETWORKS

3.1 Two-tier Streaming Overview.

In [7] [6], a two-tier 360° video streaming framework was proposed. As illustrated in Figure 1, a 360° video is partitioned into non-overlapping time segments, each segment is encoded into a BT chunk and multiple ET chunks. A BT chunk encodes the entire 360° view span (360° × 180°) at a low bitrate to provide the basic quality. BT chunks for future time segments are pre-fetched into a long streaming buffer to cope with network bandwidth variations and guarantee that any desired FoV can be rendered with minimum stalls at the client. Each ET chunk encodes video within a view window with a certain view coverage (VC) (e.g., 120° × 90°) centered at a certain direction. To provide quality differentiation, multiple ET chunks can be generated for the same view window, but coded at different bitrates. For complete coverage and smooth transition, the view windows of ET chunks in the same time segment are *overlapping* and cover the whole 360° view span. An ET chunk can be used for rendering on the client side only if it covers the user's actual view window. Since it is difficult to predict a user's view direction far into the future, only ET chunks in near future will be pre-fetched by the client. All the pre-coded BT/ET chunks are stored in the streaming server. During the streaming, the client dynamically choose precoded chunks from BT and ET tiers to download, according to the predicted view direction for the segment, the predicted download bandwidth in the next request interval, and the buffer status of each tier. The prioritied chunk scheduling is summarized in Aglorithm 1: if the BT buffer length is less than the target length $B_b^{(T)}$, the next BT chunk will be downloaded; otherwise, an ET chunk will be downloaded at the predicted direction P_e, the rate of selection the ET chunk R_e is regulated by a P-I controller, taking as inputs the estimated realtime bandwidth BW_t and the current ET buffer length B_e. More details can be found in [7]. Experiments in [7] [6] demonstrated that two-tier streaming can achieve good balance between delivered video quality and robustness against bandwidth variations and view prediction errors [6].

Table 1: Table of Notations

t	Current time
BW_t	Available bandwidth at time t
B_b (B_e)	BT (ET) buffer length
$B_b^{(T)}$ ($B_e^{(T)}$)	BT (ET) target buffer length
$B_e^{(U)}$	ET buffer length upper-bound
B_e^*	Optimal ET buffer length
R_b (R_e)	BT (ET) chunk bitrate
A_b (A_e)	BT (ET) chunk viewport coverage area
C_b (C_e)	BT (ET) video chunk
Δ_b (Δ_e or Δ_c)	Downloading time of one BT (ET or CT) chunk
P_e	Predicted FoV for ET
α	View Prediction Accuracy (VPA)
γ	Chunk Pass Rate (CPR)
η	Network Utilization Ratio (NUR)
δ	Idle time

3.2 Trajectory-based FoV Prediction

To pre-fetch an ET chunk, one needs to predict the user's view direction in near future. One approach is to predict a user's view direction in the future from the user's own past view direction trajectory, e.g., the view direction samples for the previous 60 frames, as illustrated in Figure 2. A simple approach is to use the view direction for the most recent frame as the prediction for a few frames ahead, the Last Sample Replication (LSR) scheme in Figure 2. To exploit the continuity in head movement trajectory, we can use regressions to extrapolate. One approach is linear regression based prediction, the LP scheme in Figure 2. To deal with the challenge due to occasional sudden head turning, we have developed a *truncated linear prediction* method where we only use past samples that are monotonically increasing or decreasing for extrapolation, the TLP scheme in Figure 2.

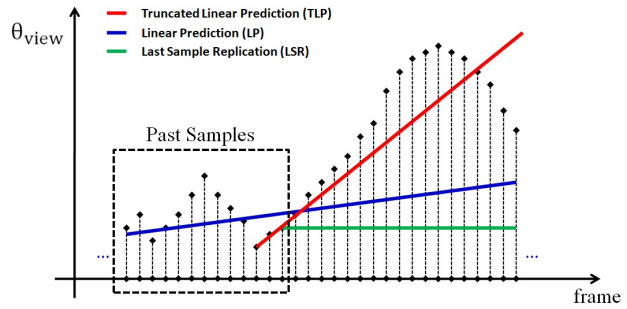


Figure 2: View Prediction Methods: TLP, LP, LSR

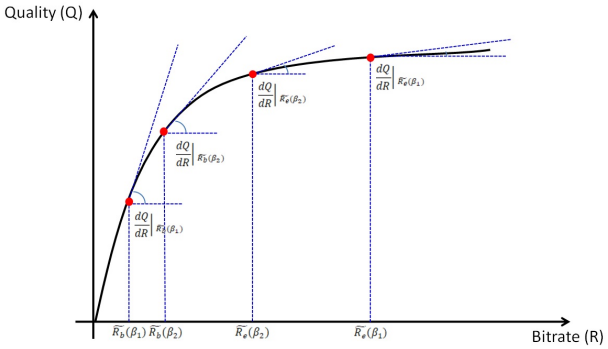


Figure 3: Sample Rate-Distortion Operation Point Analysis

3.3 Rate Allocation between BT and ET

One critical issue in the two-tier system is how to allocate the rates between the two tiers given the total sustainable video rate R_t (bits/second). One way to solve this problem is by maximizing the expected quality for a video segment. We provide a high-level formulation based on some simplistic assumptions. Because the BT rate is much lower than R_t and the BT buffer length is long, we assume that the BT chunks are mostly delivered in time for display. Consequently, for each video segment, we either receive only the BT or both the BT and ET chunks. The base-tie chunks are coded to cover the entire area of 360° video with the total rate of R_b (in bits/second) and therefore the video rendering rate is $\tilde{R}_b = R_b/A_b$ (bits/pixel), where A_b is the viewing area of the 360° video. Let R_e and A_e denote the average ET rate and the coverage area of each ET chunk, respectively. Let us assume that the ET video is coded with layered coding based on the BT decoded video, so that the pixel bit rate for an ET coded pixel is $\tilde{R}_e = R_b/A_b + R_e/A_e$. Since that the predicted view direction for a delivered video segment may not be the same as the actual user viewing direction, therefore, not all received chunks for the ET are useful. In general, only a portion of each decoded frame in the delivered ET chunk may overlap with the user's FoV for that frame. Here we introduce α to denote the average View Prediction Accuracy (VPA), namely the average overlapping ratio between the predicted view coverage and user's actual FoV, and γ to denote the average ET Chunk Pass Rate (CPR), namely the likelihood that a requested ET chunk can be delivered successfully before its display deadline. The perceived quality only depends on what is being rendered on the screen based on the user's FoV. The probability that a rendered pixel is covered by the delivered ET chunk is $\alpha\gamma$. Assuming the BT and the ET coders can be characterized by their respective quality-rate (Q-R) functions $Q_b(\tilde{R})$ and $Q_e(\tilde{R})$, where $\tilde{R}$ is bits per coded pixel, the expected rendered video quality with the constraint $R_b + R_e = R_t$ can be expressed as

$$\begin{aligned} Q(R_b;\alpha,\gamma,R_t) &= \alpha\gamma Q_e(\tilde{R}_e) + (1-\alpha\gamma)Q_b(\tilde{R}_b) \\ &= \alpha\gamma Q_e\left(\frac{R_b}{A_b} + \frac{R_t - R_b}{A_e}\right) + (1-\alpha\gamma)Q_b\left(\frac{R_b}{A_b}\right) \end{aligned} \tag{1}$$

Note that α and γ are both dependent on the ET prefetching buffer length. To prefetch an ET chunk at a future time into the buffer, one has to predict the user's FoV center at that time. In general, the longer the ET prefetching buffer (measured in video time), the lower the ET VPA α. Meanwhile, longer prefetching buffer can better absorb network bandwidth variations, leading to higher chunk delivery ratio γ. In practice, γ also depends on R_b and R_e values. Here we assume that we can estimate the average network bandwidth fairly accurately and we set R_t below the estimated average bandwidth $\overline{BW}$ with a safety margin controlled by a Network Utilization Ratio (NUR) η, i.e., $R_t = R_b + R_e = \eta\overline{BW}$, so that the average γ is not affected by the individual R_b and R_e values. Therefore, for a given prefetching buffer length, α and γ can be considered as constants. Once η is configured, the optimal R_b can be solved by setting $\frac{\partial Q}{\partial \tilde{R}_b} = 0$, which yields

$$\left.\frac{\partial Q_e}{\partial R}\right|_{\tilde{R}_e^*} = \beta \left.\frac{\partial Q_b}{\partial R}\right|_{\tilde{R}_b^*}, \text{with } \beta = \left(\frac{1-\alpha\gamma}{\alpha\gamma}\right)\frac{A_e}{A_b - A_e}. \tag{2}$$

Equation (2) says that R_b should be chosen so that the Q-R slope at $\tilde{R}_e$ should be β times the slope at $\tilde{R}_b$. Figure 3 demonstrates the optimal $\tilde{R}_b^*$ and $\tilde{R}_e^*$ relations for two different β values for a hypothetical but typical Q-R curve: β_1=1/45 resulting from assuming $\alpha\gamma = 0.9$ and $A_b/A_e = 6$, and $\beta_2 = 3/35$ from assuming $\alpha\gamma = 0.7$. We see that if α and γ are both very close to 1, then β is very small, and the optimal allocation is to let $\tilde{R}_b$ be very low. This corresponds to the case that view and bandwidth prediction are both very accurate, so that a rendered pixel can almost always be covered by a delivered ET chunk. In this circumstance, it is better not to waste bits to send entire 360° scope in the base tier. When view and/or bandwidth prediction is less accurate ($\alpha\gamma$ is lower), it is better to spend more bits on the base tier, to ensure that pixels that are only rendered from BT chunks have sufficient quality. In practice, we should set $\tilde{R}_b = \max(\tilde{R}_{b,min}, \tilde{R}_b^*)$, to make sure that any FoV region that are not covered by ET chunks due to either view prediction or delivery errors can be rendered with a basic quality, with rate $\tilde{R}_{b,min}$. Note that for a chosen encoding method, one can derive the operational Q-R functions, and consequently determine the optimal R_b based on Eq. (2).

The above analysis shows that the optimal bit allocation among the two tiers depend on $\alpha\gamma$, with higher $\alpha\gamma$ leading to higher ET quality and lower BT quality. Besides the $\alpha\gamma$ factor, the optimal operation point also depends on the Q-R model for 360° video encoding. In this work, we follow the Joint Video Exploration Team (JVET)

common test conditions (CTC) and evaluation procedures for 360^o video [1], the equi-rectangular encoding statistics for four Call-for-Evidence (CfE) 8K sequences [21] are used to derive Q-R model. The rates R are High Efficiency Video Coding (HM-16.15) encoding bitrates (in kbps) with quantization parameter (QP) values of 22, 27, 32 and 37, respectively. Weighted-to-spherically-uniform peak-signal-to-noise ratio (WS-PSNR) is used as the objective quality metric. For simplicity, the Q-R relationship is modeled by a logarithmic function with two free parameters, i.e., $Q(R) = a + b \cdot \log R$, where a and b are content-dependent parameters, and are chosen to minimize the fitting mean-square-error (MSE) in our formulation. The sample Rate-Distortion (R-D) fitted curves are provided in Figure 4. Experimental results demonstrate that the parameters (a and b) derived from different sequences are pretty close, indicating a similar curve trend across different 360° videos. For simplicity, we average the a and b values across videos to derive a universal model with $a = 31.74$ and $b = 3.3$. Please note that in our rate allocation formulation, the optimal rate allocation only depends on the sum and ratio between $\tilde{R}_b^*$ and $\tilde{R}_e^*$ and is therefore video-content independent, as long as the video Q-R relationship follows a logarithmic pattern (which is usually true, as validated in Figure 4). The closed-form solutions of $\tilde{R}_e^*$ and $\tilde{R}_b^*$ are provided in Eq. (3) and Eq. (4),

$$\tilde{R}_e^* = \frac{R_t}{1+\beta} = \frac{\eta\overline{BW}}{1+\beta}, \tag{3}$$

$$\tilde{R}_b^* = \frac{\beta R_t}{1+\beta} = \frac{\beta\eta\overline{BW}}{1+\beta}, \tag{4}$$

where β is defined in Eq. (2) and is determined jointly by the view prediction accuracy (i.e., α), the chunk pass rate (i.e., γ) and the BT and ET coverage configurations (i.e., A_b and A_e). $\overline{BW}$ is the predicted average network bandwidth and η is the target network bandwidth utilization ratio.

3.4 Two-tier Streaming in 5G Networks

Based on the optimal rate allocation analysis in the previous section, we now experimentally study the two-tier coding and streaming trade-offs through detailed simulations driven by 5G network bandwidth traces.

3.4.1 5G 802.11ad Bandwidth Trace Collection. We have recently acquired a 5G wireless research platform, which provides multi-gigabit throughput over commercial 802.11ad (WiGig) with directional transceivers. Each node is equipped with a 32-element phased array antenna that supports four 2 GHz wide channels in the 60 GHz unlicensed band. Application servers can be plugged in via the Ethernet interface. As shown in Figure 5, we use two laptops to work as two end points to generate or receive data. Traffic is carried by the WiGig channel between the two antenna units. One processing unit is attached to each laptop that is transparent from the application's aspect and is able to make the antenna work as an Network Interface Card (NIC). The processing units are controlled by the central controller.

On this platform, 5G wireless bandwidth traces are collected by recording real-time throughput of *iperf*. In order to fully explore how the two-tier streaming system performs under different 5G

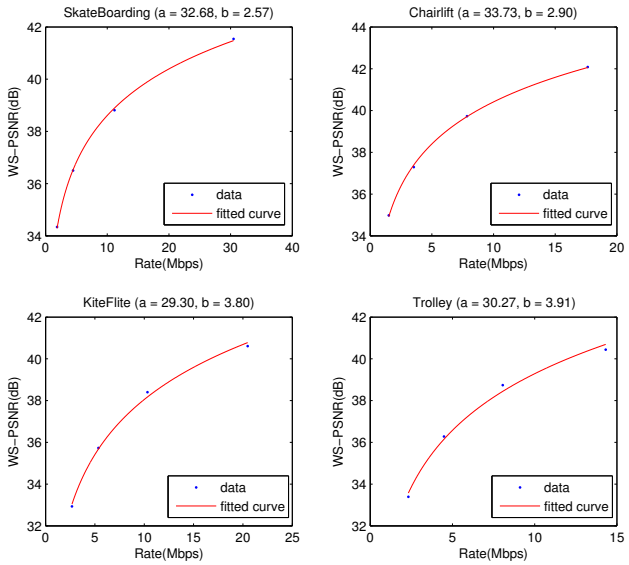


Figure 4: JVET CfE 360^o Video Rate-Distortion Curves. The bitrate is used to encode the entire 360° video in equi-rectangular format.

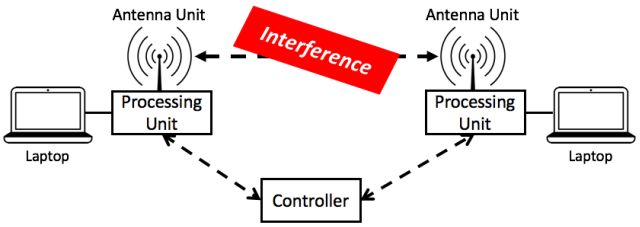
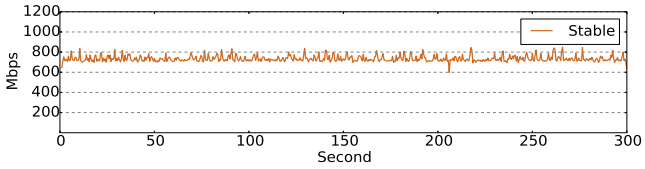


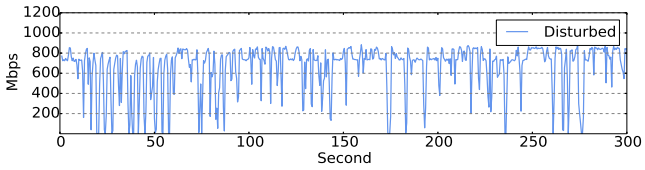
Figure 5: 5G 802.11ad Testbed for Trace Collection

scenarios, we collected three traces as illustrated in Figure 6, representing different network conditions: stable, disturbed and unstable, respectively. Bandwidth fluctuations are triggered by introducing blocking objects, e.g., metal materials or human bodies, between the pair of antenna units. The resulting bandwidth fluctuations are determined by the intensities and frequencies of the introduced blockages. For the scenario shown in Figure 6(a), transmission is finished without any blockage introduced. Throughput is stable around 730 Mbps all the time. In addition, the statistics in the first row of Table 2 also demonstrate the stability of both throughput and delay. As for the disturbed and unstable traces, blockages decrease the throughput and increase the end-to-end latency. The average throughputs decrease by 75 Mbps and 150 Mbps respectively from the stable case. According to the standard deviation (SD) in the Table 2, there are large fluctuations in both throughput and latency. In some extreme cases, the throughput drops to zero and the latency ramps up to four seconds. The unstable trace shares a similar overall throughput pattern as the disturbed one, but with more frequent and stronger interruptions. For instance, at around time

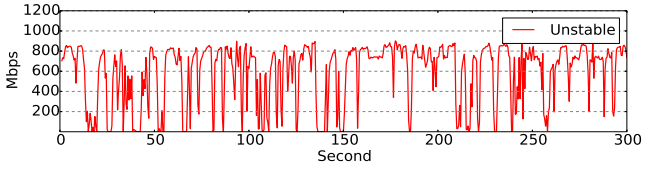
40s and 140s in Figure 6(c), the channel is unavailable for around five seconds. For all the experiments in the following, 5G related results are obtained from simulations driven by these three traces.



(a) Stable



(b) Disturbed



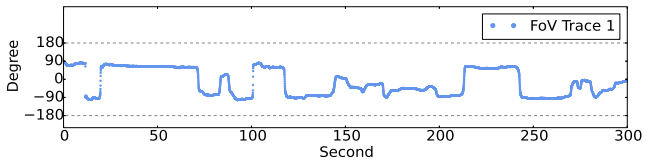
(c) Unstable

Figure 6: 5G Wireless Trace Throughput Pattern

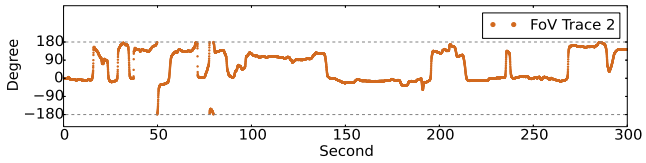
3.4.2 Optimal Rate Allocation. In order to allocate rates between BT and ET optimally, based on the Q-R model and analysis in Section 3.3, the maximum product of α and γ should be adopted. First, we measured FoV prediction accuracy α for two different FoV traces at different prediction intervals, (equivalently ET prefetching buffer lengths measured in time). As illustrated in Figure 8(a), α value decreases as the prediction interval/ET prefetching buffer length B_e increases. FoV trace 2 is more fluctuant and difficult to predict than FoV trace 1. This can be visually verified by the FoV patterns shown in Figure 7.

Chunk pass rate γ is another crucial factor, because ET chunks cannot be guaranteed to be delivered before their display deadlines under unstable network conditions. If no ET video chunk is available, even though the FoV prediction is 100% accurate, the final quality obtained by users is still low. Initially, we set $R_b + R_e$ equals to 85% of average bandwidth $\overline{BW}$ of each 5G trace, i.e., $\eta = 85\%$, and roughly allocate $R_b = 0.1 * R_e$, $R_{e1} = 0.75R_e$, $R_{e2} = R_e$ and $R_{e3} = 1.25R_e$. By fixing the BT buffer length to 10 seconds, we get the γ curve showing how CPR varies with the ET prefetching buffer length through detailed chunk-level two-tier streaming simulation of Algorithm 1 in *Python*.

Based on the obtained α and γ curves, we can plot the $\alpha\gamma$ curve and choose the optimal ET prefetching buffer length to maximize the $\alpha\gamma$ product. The optimal rate allocation for each bandwidth-FoV trace combination can also be derived using the $\alpha\gamma$ value. We set



(a) FoV Trace 1



(b) FoV Trace 2

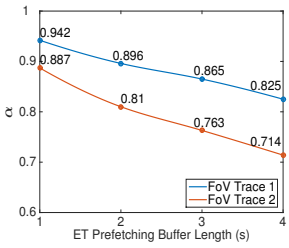
Figure 7: FoV Trace Direction (horizontal)

up three rate levels for ET: $0.75R_e$, R_e and $1.25R_e$. As discussed in Section 3.3, if α and γ are both close to 1, small value of β results in a very low R_b and high R_e. For example, based on the initial rate allocation, the optimal ET prefetching buffer length is 1s for the combination of Stable 5G trace and FoV trace 1. The corresponding $\alpha\gamma$ value of 0.94 makes the optimal ratio of R_b/R_e to be only 0.0456. The new optimal rate allocation generated from this $\alpha\gamma$ pair is much different from the initial allocation. Due to the extremely high value of R_e, when we simulate the two-tier streaming system again with the updated rate allocation, the CPR γ becomes much smaller than the simulation with the initial rate allocation with short ET prefetching buffer. For instance, γ value changes from 0.998 to 0.738 under the new rate allocation. To obtain a stable rate allocation, we numerically solved a fixed-point problem by regenerating the γ curve and iterating the whole process until the γ curve and rate allocation converge. Figure 8(b) represents the final converged values of γ, and it becomes obvious from Figure 8(c) and 8(d) that the optimal ET prefetching buffer lengths of most bandwidth-FoV combinations are larger than the ones with the initial rate allocation. Based on the converged $\alpha\gamma$ value, the optimal ET prefetching buffer lengths and rate allocation configurations are shown as Table 3. The analysis in Section 3.3 is also confirmed here: with stable network condition and steady FoV direction movement, R_b only occupies a small portion, and R_e uses the most of the available rate; the rate allocation is more balanced when $\alpha\gamma$ reduces due to unstable network throughput and/or fluctuant FoV change.

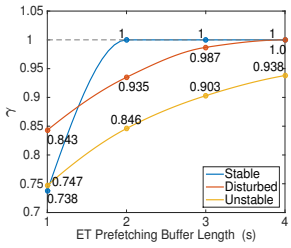
3.4.3 Detailed QoE Evaluation. The two most important factors influencing user QoE of the traditional 2D planar video are video rate and freezing. The same should hold for 360° video. To improve QoE, first of all, a streaming system should be designed to avoid video freezing even under unstable network conditions. Then, on the premise of meeting the first requirement, the system should try to provide users with higher video rate. In our two-tier system, the main purpose of adopting 10s buffer length and 360° panoramic view port for BT chunks is to satisfy the first rule. 10s prefetched video can cope with bandwidth fluctuation or temporary network

Table 2: 5G Wireless Traces Information

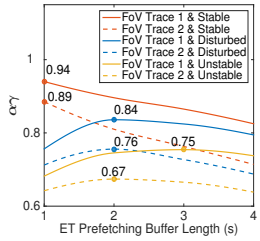
Traces	Throughput (Mbps)					Latency (ms)				
	Mean	SD	Max	Min	Med	Mean	SD	Max	Min	Med
Stable	734.34	22.51	832.5	662.0	729.5	10.45	3.48	48.60	2.31	10.4
Disturbed	659.67	206.73	866.5	0	735.5	64.46	213.5	1695.0	0.716	8.575
Unstable	585.31	277.10	874.5	0	715.25	164.77	491.33	4119.0	0.920	8.80



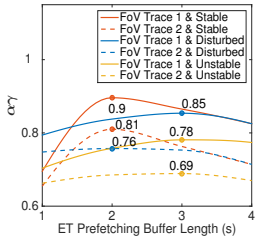
(a) View Prediction Accuracy α Curve



(b) Chunk Pass Rate γ Curve



(c) Initial $\alpha\gamma$ Curve



(d) Converged $\alpha\gamma$ Curve

Figure 8: Optimal Rate Allocation with Different Combinations of 5G Wireless Bandwidth Traces and FoV Traces

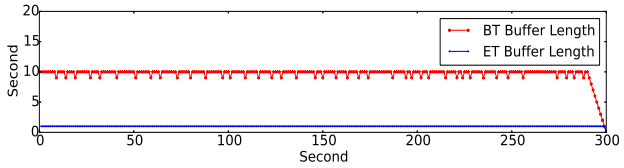
Table 3: Optimal Rate Allocation

5G Trace	FoV Trace	$B_e{}^*$(s)	$\alpha\gamma$	Rate Allocation (Mbps)			
				R_b	R_{e_1}	R_{e_2}	R_{e_3}
Stable	FoV 1	2	0.90	45.7	433.9	578.5	723.1
	FoV 2	2	0.81	89.8	400.8	534.4	668.0
Disturbed	FoV 1	3	0.85	62.7	373.5	498.0	622.5
	FoV 2	2	0.76	103.4	343.0	457.4	571.7
Unstable	FoV 1	3	0.78	83.3	310.7	414.3	517.8
	FoV 2	3	0.69	120.9	282.5	376.6	470.8

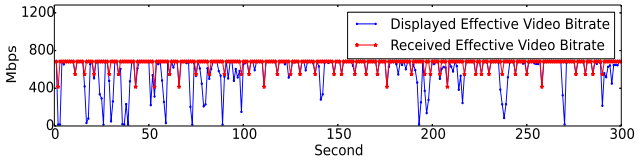
outage. A panoramic view port guarantees no matter how users change their view direction, the corresponding part of content is always available. On top of BT chunks, ET chunks with higher bitrate but smaller coverage are introduced to improve the final rendered video quality when user FoV can be predicted.

We evaluate the QoE of two-tier streaming under different bandwidth-FoV combinations. For the combination of stable network trace and fluctuant FoV movement, the optimal ET prefetching buffer length B_e^* is 2s ($B_e^{(T)} = B_e^* - 1 = 1s$) and the optimal rate allocation is shown in Table 3. From Figure 9(a), we find that BT buffer keeps stable at around 10s all the time, indicating that no video freezing occurs.

In addition, the ET buffer length also is well controlled around the target value throughout the whole process. In most of the time, ET chunks with the highest bitrate 668.01 Mbps are downloaded. The total of R_b and R_{e3} roughly equals to the average bandwidth. As long as ET with bitrate R_{e3} is requested, the downloading time of one 1s ET chunk is about one second. However, the performance curve in Figure 9(b) illustrates even though ET video is available, the final displayed video bitrate can still be affected by sudden user FoV movement. Referring to FoV trace in Figure 7(b), the frequent and irregular FoV movements at time of 30-100s, around 200s and 240s result in poor displayed video quality in Figure 9(b).



(a) Buffer Length



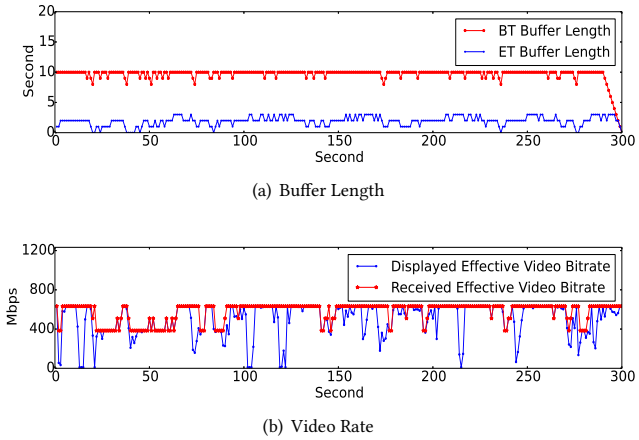
(b) Video Rate

Figure 9: Stable Bandwidth Trace & FoV Trace 2

Table 4: Impact of Network Utilization

η	B_e^*(s)	α	γ	Rate Alloc (Mbps)		QoE	PAN Rate (Mbps)
				R_b	R_e		
65%	2	0.896	0.982	34.8	394.0	161.0	71.46
75%	2	0.896	0.969	48.6	446.1	177.4	82.46
85%	3	0.865	0.987	62.7	498.0	**187.1**	93.45
100%	3	0.865	0.959	84.2	575.5	186.8	109.95
110%	3	0.865	0.928	139.3	586.3	182.2	120.94

System performance in another combination of less stable network bandwidth and steady FoV is also evaluated. Different from the previous combination, the optimal ET buffer length becomes 3s ($B_e^{(T)} = B_e^* - 1 = 2s$). This is because the steady FoV is easy to predict and a longer prefetching buffer can improve the chunk pass rate under the unstable network condition. As illustrated in Figure



(a) Buffer Length

(b) Video Rate

Figure 10: Disturbed Bandwidth Trace & FoV Trace 1

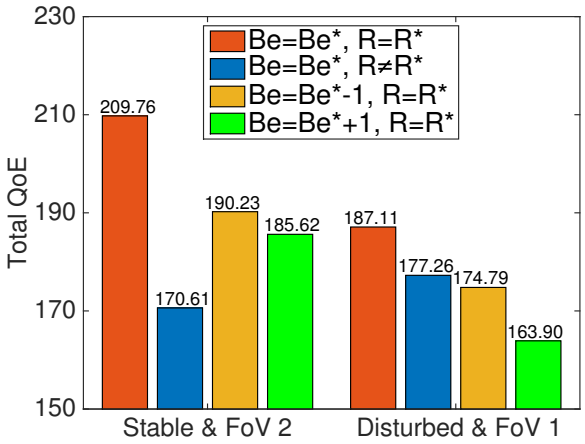


Figure 11: Performance Comparison between Optimal and Non-optimal Configurations

10(a), both BT and ET buffers become more fluctuant, the ET buffer that may even decrease to zero. As shown in Figure 6(b), from 30s to 90s, the network bandwidth changes dramatically between zero and 800 Mbps. Correspondingly, from Figure 10(b), the system chooses to download ET chunks with a lower rate, e.g. R_{e1}, to preserve the ET buffer and improve ET chunk pass rate γ.

To cope with the high volatility of 5G connections, how the delivered video quality is correlated with network utilization should be discussed. As mentioned in Section 3.3, the network utilization ratio factor η decides how much network resource is used for transmitting BT and ET chunks. Different network utilizations result into different CPR. Not only the total video rate, but also the allocation ratio between R_b and R_e are affected by η. We use the same model to generate the optimal rate allocations under different network utilizations, and the corresponding QoE. To represent the final video quality properly, we use the logarithmic (base 10) of the displayed video rate divided by a fixed number, e.g. 100 Mbps. Results for five different network utilizations under the same bandwidth-FoV combination (Disturbed & FoV Trace 1) are shown in Table 4. The overall video quality is the best when 85% bandwidth is utilized. If η is smaller than 85%, *ET* chunk rate is low and the available bandwidth cannot be utilized efficiently; on the contrary, when η is larger than 85%, ET chunk delivery ratio drops, leading to lower rendered quality. The last column in the table presents the maximum effective video bitrate in which all bandwidth is assigned to panoramic (PAN) video without ET. Even though panoramic video achieves a higher bitrate, without ET, the final bitrate could be displayed to users are still very low even compared with R_{e1} in two-tier streaming system.

Finally, we compare the streaming performance of the optimal configuration derived from our model in Section 3.3 with other non-optimal configurations. Firstly, we compare the performance of a streaming system with optimal ET buffer and optimal rate allocation with a streaming system with the optimal ET buffer but non-optimal rate allocation. We use the rate allocation generated from the initialized $\alpha\gamma$ value as the non-optimal rate allocation. We also evaluate the performance of another two configurations with the optimal rate allocation, but non-optimal buffer lengths: $B_e^* + 1$ and $B_e^* - 1$. From the result presented in Figure 11, we find for both bandwidth-FoV combinations, the optimal configuration based on our model performs much better than the non-optimal ones.

4 TILE-BASED FOV PREDICTION CORRECTION

4.1 Opportunistic FoV Correction

In the current two tier design, ET video chunk is downloaded based on viewport prediction using LP as presented in Section 3.2. The prediction accuracy drops as prediction interval increases. The frame-level FoV prediction accuracies for the two sample FoV traces in Figure 7 are plotted in Figure 12. The view prediction accuracy for close-by frames is very high (i.e., up to 95%), but prediction accuracy drops drastically when the prediction interval increases. In two-tier streaming, a pre-fetched ET chunk is useless for rendering if it does not cover the actual user view window. The high-throughput low-latency data transmission capability of 5G makes it possible for *realtime FoV correction*: immediately before the rendering of a buffered ET chunk, one can re-do the view direction for the chunk with shorter leading time τ (thus higher prediction accuracy) and download additional *video tiles* to patch the buffered chunk to better cover the actual user FoV. Therefore, beyond the original two-tier design, a third "correction" tier of video tiles are introduced into our system to effectively compensate for the FoV prediction errors using a new FoV prediction (using TLP in Section 3.2) with a shorter leading time. All the three tiers of video are streamed over the same 5G channel and compete for transmission opportunities. We extend the prioritized two-streaming algorithm in Algorithm 1 to schedule the transmission of correction video tiles. In our design, correction tier yields to the BT and ET tiers, and will be transmitted only when BT and ET buffers are sufficiently long for continuous streaming. As a result, the correction tiles will be transmitted with the lowest priority.

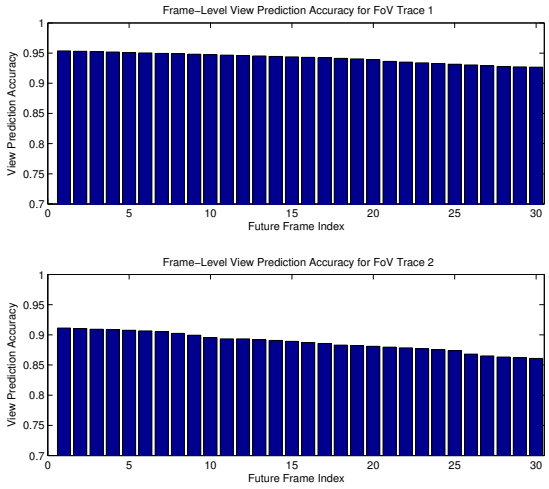


Figure 12: Frame Level Prediction

Algorithm 2 Opportunistic FoV Prediction Correction

Initialization at $t = 0$;
while One chunk downloading is finished **or** $t = 0$ **and** display is not terminated **do**
  if $B_b <= B_b^{(T)}$ **then**
    Download next C_b;
    $t \leftarrow t + \Delta_b$;
  else
    $\tau \leftarrow \lceil t \rceil - t$;
    if $\tau <= T_l$ **and** $C_e^{\lceil t \rceil}$ exists in B_e **then**
      RequestCorrection$(t, C_e^{\lceil t \rceil})$
    else
      if $B_e <= B_e^{(U)}$ **then**
        Predict bandwidth BW_t
        Refine BW_t with P-I controller
        Predict FoV P_e for next C_e;
        Request for next C_e;
        $t \leftarrow t + \Delta_e$
      else
        $t \leftarrow t + \delta$;
      end if
    end if
  end if
end while
return

function RequestCorrection(t, C_e)
  Make new FoV prediction P_e' for C_e;
  Compare P_e' and P_e and calculate missing FoV A_e;
  Request correction tiles C_c covering A_e;
  Estimate download duration Δ_c for correction tiles;
  $t \leftarrow t + \Delta_c$;
  if t exceeds $\lceil t \rceil + T_d$ **then**
    Discard current correction downloading;
  end if
return
end function

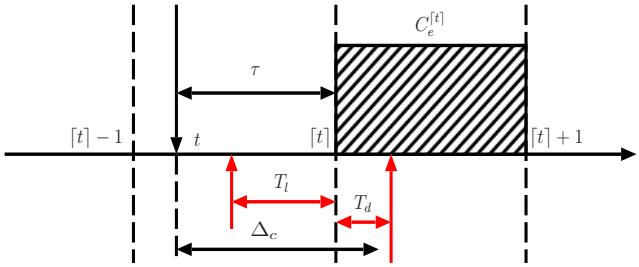


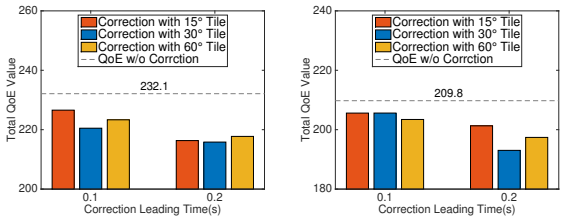
Figure 13: 5G FoV Correction Leading Time

As detailed in Algorithm 2, BT is given the highest priority. After the previous chunk download finishes, if the BT buffer length is less than the target value, a new BT chunk will be requested immediately. Otherwise, we have a chance for ET or CT downloading. More specifically, to achieve high prediction accuracy, we only do correction for the next ET chunk to be rendered. As illustrated in Figure 13, at time t, if the next ET chunk to be rendered at time $\lceil t \rceil$ is in the buffer, one can redo FoV prediction with leading time of $\tau = \lceil t \rceil - t$. To achieve high prediction accuracy, we require $\tau \leq T_l$, where T_l is a pre-set threshold. Based on the FoV discrepancy between the buffered chunk and the new prediction, one can calculate the total size of the patching tiles, and estimate the download latency of the CT tiles as Δ_c. If the download completion time is well beyond the rendering time of the chunk, the correction becomes useless. So that the correction for a chunk will be granted only if the download completion time is no later than $\lceil t \rceil + T_d$, where T_d is the delay tolerance parameter, i.e., $\Delta_c \leq \tau + T_d$. As a result, the mechanism to request for a prediction correction data is designed to be opportunistic. If either condition is not met, system will request the next ET video chunk.

4.2 Impact of Correction Leading Time and Tile Size

T_l is a parameter that controls whether a correction will be requested. A larger T_l results in more corrections to be requested and downloaded, but degrades the FoV prediction accuracy. Obviously, the setting of T_l reflects the trade-off between value of α and γ discussed in Section 3.3. In addition, as correction is a tile-based strategy, the span size of each tile also plays an important role: larger tile size could preserve a higher coding efficiency, but the patching granularity is compromised. For example, with 60° tile span size, even for only one or two degree difference between the original FoV and the new correction prediction, a correction tile with 60° has to be delivered. The impact of tile size should also be evaluated. Figure 14 illustrates the results of different configurations of T_l and tile span size. Six combinations of two FoV traces and three 5G throughput traces are tested. FoV Trace 1 with prediction accuracy value of 0.942 is easier to predict than FoV Trace 2 whose accuracy is only 0.887 when prediction interval is 1s ahead. The three 5G throughput traces discussed in Section 3.4.1 are utilized to simulate different network conditions. The original ET viewport span is still 150°, and the user VP coverage is 120° horizontal. The rate allocations for different bandwidth-FoV combination are assigned according to the optimal values summarized in Table 3.

Tile sizes of 15°, 30° and 60° and leading time of 0.1s and 0.2s are used for comparison. The coding efficiency of different tile sizes are configured according to our simulation over JVET 360° video sequences, and 20% , 10% and 5% extra bitrate is consumed for tile size of 15°, 30° and 60° respectively. While comparing the performance of $T_l = 0.1s$ and $T_l = 0.2s$, in most cases, the former one performs better. Correcting original FoV video $0.2s$ ahead is not tight enough and wastes bandwidth, and the delivered QoE is often worse than the QoE without correction. When $T_l = 0.1s$, comparing the performances of all six bandwidth-FoV combinations, result shows that correction is helpful for the three combinations whose optimal ET buffer length B_e^* equals to 3s, but is harmful for the overall performance when B_e^* is short. In addition, simulation results show that tile sizes do have performance impacts under all bandwidth-FoV and leading time combinations. However, from our current results, there is no tile-size which is universally optimal across all combinations.



(a) Stable & FoV Trace 1 (b) Stable & FoV Trace 2

(c) Disturbed & FoV Trace 1 (d) Disturbed & FoV Trace 2

(e) Unstable & FoV Trace 1 (f) Unstable & FoV Trace 2

Figure 14: Prediction Correction Performance

4.3 Impact of Network Utilization

In the previous section, we evaluated two-tier system performance with different network utilizations, and find 85% is the optimal ratio. But when correction is considered, performance under each specific network utilization might become different. For example, for a low network utilized scenario, both BT and ET buffer are more likely to reach their upper-bounds and result in idle channel. During this idle time, correction tiles can be downloaded to improve the quality. On the other hand, if the BT or ET chunk downloading has occupied a large portion of the network resource, the chance of downloading correction tiles is low. Correction performance under different network utilizations is reported in Table 5. We use different leading time and tile size combinations to evaluate the optimal configuration and maximum quality. It's easy to find lower utilization has more significant improvement after correction. If the utilization is close to or greater than the available bandwidth, correction degrades on the performance. Like 100° or 110°, the improve ratios are negative with correction enabled.

Table 5: Impact of Network Utilization on correction

NUR η	65%	75%	85%	100%	110%
Two-tier quality	161.0	177.4	187.1	186.8	182.2
Quality with correction	164.7	180.6	190.4	182.6	174.0
Improve Ratio	2.30%	1.80%	1.76%	−2.25%	−4.50%
B_e^* (s)	2	2	3	3	3

5 MULTI-PATH MULTI-TIER STREAMING IN HETEROGENEOUS WIFI/5G NETWORKS

In the previous two sessions, we have studied multi-tier 360° video streaming in 5G networks, now we study the performance of multi-tier streaming in heterogeneous WiFi/5G networks. In the envisioned heterogeneous network, a 360° video client can connect to the streaming server simultaneously through WiFi access point and 5G base station, as illustrated in Figure 15. There are many strategies to allocate multiple tiers to multiple paths. Due to the space limit, we present our results for one specific strategy.

5.1 WiFi for Two-tier, 5G for Correction and Retransmission

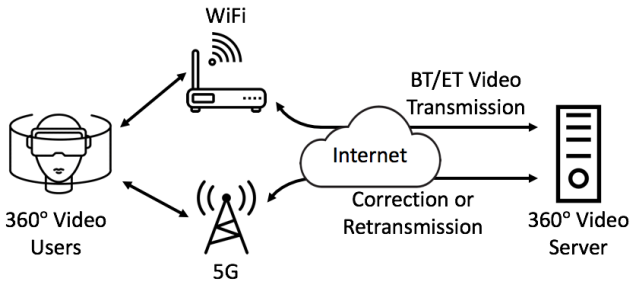
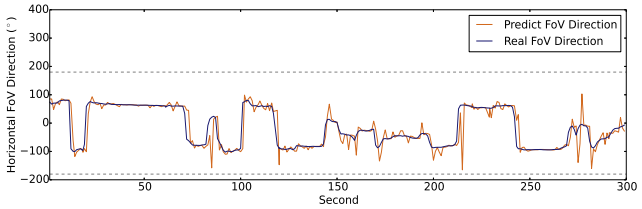


Figure 15: 360° Video Streaming over Multipath in Heterogeneous Networks

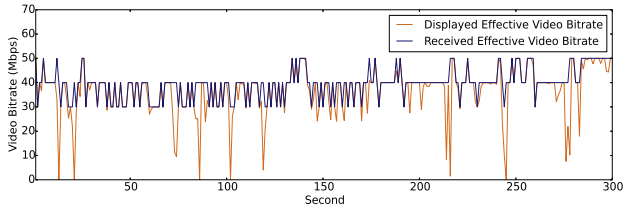
Since 5G is still in its infancy, it data pricing mechanism is still uncertain. Lots of questions are open: whether 5G network data usage will be expensive or not? whether users will be charge based on data volume or connection time? For illustration purpose, we study one case where users are offered with a free WiFi connection

with moderate bandwidth, and another high bandwidth but priced 5G connection. Fortunately, leveraging on multipath transmission like [8] [19], users flexibly combine WiFi and 5G connections to achieve their desired performance-cost trade-off. One 360° video streaming solution is to download all the original BT/ET video chunks over the free WiFi channel as the volume of 360° video is large and users might not watch a 360° video with tens of minutes using a priced 5G connection if free WiFi is available. However, the correction over 5G idea discussed in Section 4 is still viable in this scenario. Compared with normal BT/ET video data, the quantity of data used for correction is small and correction tiles can be transmitted over 5G with high throughput and low delay. Furthermore, due to the network bandwidth fluctuations, ET video is not guaranteed to be delivered all the time over WiFi. If WiFi cannot deliver an ET chunk before its display deadline, we can *retransmit* the whole ET chunk with the highest video bitrate through the high speed 5G channel to improve user QoE. The coding efficiency for the whole ET chunk is higher than CT tiles.

One WiFi trace with average throughput of 43 Mbps and the stable 5G network trace are utilized to simulate the heterogeneous network environment. Using model developed in Section 3.3, we calculate the optimal rate allocation matching the average WiFi bandwidth as:R_b: 3 Mbps, R_{e1}: 30 Mbps, R_{e2}: 40 Mbps and R_{e3}: 50 Mbps. Simulation is running over the FoV trace 1 illustrated as Figure 7(a). This FoV trace is stable for most of the time but with some sudden and sharp FoV movement at time of about 15s, 75-120s and 245s, etc. The prediction method for the original ET is LP. Correspondingly, the displayed effective bitrate visible to users decreases significantly at these timestamps.



(a) FoV Prediction



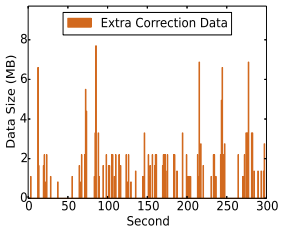
(b) Displayed Effective Video Bitrate

Figure 16: 360° Video Performance over WiFi

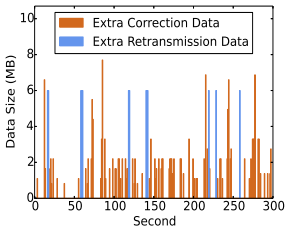
5.2 Cost Effectiveness of Streaming over 5G

In order to evaluate whether the extra data usage over 5G is desirable, we calculate and compare the final 360° video display bitrate as a function of how much 5G data is used or how much time the 5G channel is occupied. Correction and retransmission are still based on 30° tiles. Leading time T_l of correction and retransmission are $0.05s$ and $0.1s$ respectively, and termination threshold T_d is $0.1s$ for both. If only video correction is allowed and no retransmission, as shown in Figure 17(a), correction data is requested frequently with data size ranging from 1 MB to nearly 7 MB. especially when the FoV direction changes quickly. On top of correction, if retransmission is also available, we count the data delivered over 5G for both correction and retransmission in Figure 17(b). With the current optimal rate allocation for this WiFi trace, ET γ value is comparatively high. Because of this, the chance for retransmission is very low, compared with correction.

Result in Figure 18(a) indicates that, benefiting from correction, the final $\alpha\gamma$ value increases considerably over single path streaming. For example, during the time from 135s to 175s, $\alpha\gamma$ is dramatically improved. We can also conclude from the Table 6 that the overall video bitrate is improved by 1.9 Mbps. Meanwhile, one of the drawback is that correction does nothing if the next ET chunk is not available. Retransmission can be introduced to cope with this case through retransmitting a new ET chunk. At the time of 60s and 120s in Figure 18(b), retransmission is triggered and the highest video bitrate is displayed. As explained by the last row in Table 6, the average effective video bitrate increases by 2.53 Mbps with retransmission enabled. Users may be concerned with how much resources are consumed and how much 5G cost they have to pay to get the video quality improvement. We record the 5G data usage and 5G channel usage time. The last two columns in Table 6 is calculated through dividing the number of total effective gain by the corresponding 5G data usage or channel time. Result illustrates that correction is slightly more efficient than retransmission. So whether correction and retransmission over 5G will be enabled to improve 360° streaming over WiFi is a trade off between the video quality improvement and the cost incurred for using 5G channel.



(a) Extra Correction Data



(b) Extra Correction/Retransmission Data

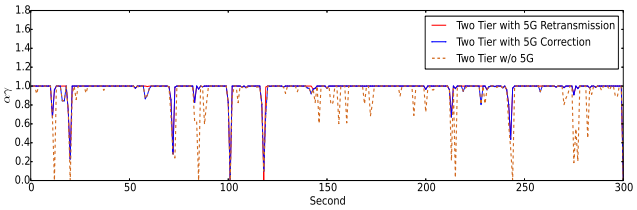
Figure 17: Correction and Retransmission over 5G Channel

6 CONCLUSIONS

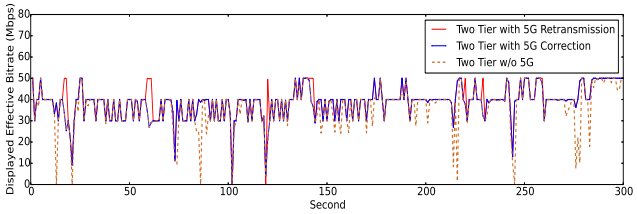
In this paper, we developed novel multi-path multi-tier 360° video streaming solutions for 5G wireless networks. To efficiently utilize the high bandwidth available in 5G and cope with high bandwidth volatility, we analytically and experimentally studied the optimal rate allocation between the base tier and enhancement tier. We

Table 6: Cost Efficiency of Correction/Retransmission over 5G Channel

	Average Effective Bitrate (Mbps)	Total Effective Gain (Mb)	Effective Bitrate Gain (Mbps)	Extra Data Usage (Mb)	Extra Time Usage (s)	Gain per Extra Data	Gain per Extra Time
Traditional Two Tier	36.67	∅	∅	∅	∅	∅	∅
Correction over 5G	38.57	570	1.90	1636	2.24	0.348	254.46
Retransmission over 5G	39.20	759	2.53	2286	3.12	0.332	243.27



(a) Effective ET Ratio



(b) Displayed Effective Video Bitrate

Figure 18: 360° Video Performance with Multipath

showed that the optimal trade-off is determined by the product of FoV prediction accuracy and chunk delivery success ratio. Leveraging on high-throughput and low-latency 5G data transfer, we developed novel tile-based FoV correction scheme that patches a buffered enhancement tier chunk right before its display deadline. In heterogeneous WiFi and 5G networks, we showed that 5G channel can also be used in a cost-efficient manner to retransmit chunks that cannot be delivered in time through WiFi. Important insights obtained from our analysis and simulation study driven by real 5G network traces and user FoV traces can be used to guide the design of future 360° video streaming systems in 5G wireless networks.

REFERENCES

[1] E. Alshina, J. Boyce, A. Abbas, and Y. Ye. 2017. *JVET common test conditions and evaluation procedures for 360° video, JCTVC Doc. G1030.*

[2] M. Budagavi, J. Furton, G. Jin, A. Saxena, J. Wilkinson, and A. Dickerson. 2015. 360 degrees video coding using region adaptive smoothing. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*. IEEE, 750–754.

[3] R. Daniels and R. Heath. 2007. 60 GHz wireless communications: emerging requirements and design recommendations. *IEEE Vehicular Technology Magazine* 2, 3 (September 2007), 41 – 50.

[4] G. Van der Auwera, Hendry M. Coban, and M. Karczewicz. 2016. *Truncated Square Pyramid Projection (TSP) For 360 Video, JVET Doc. D0071.*

[5] F. Duanmu, Y. He, X. Xiu, P. Hanhart, Y. Ye, and Y. Wang. 2018. Hybrid Cubemap Projection Format for 360-Degree Video Coding. In *Proceedings of the 2017 Data Compression Conference (DCC '18)*. 263–264.

[6] Fanyi Duanmu, Eymen Kurdoglu, S. Amir Hosseini, Yong Liu, and Yao Wang. 2017. Prioritized Buffer Control in Two-tier 360 Video Streaming. In *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network (VR/AR Network '17)*. ACM, New York, NY, USA, 13–18. https://doi.org/10.1145/3097895.3097898

[7] F. Duanmu, E. Kurdoglu, Y. Liu, and Y. Wang. 2017. View direction and bandwidth adaptive 360 degree video streaming using a two-tier system. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. 1–4. https://doi.org/10.1109/ISCAS.2017.8050575

[8] Alan Ford, Costin Raiciu, Mark Handley, and Olivier Bonaventure. 2013. *TCP extensions for multipath operation with multiple addresses.* Technical Report.

[9] Mario Graf, Christian Timmerer, and Christopher Mueller. 2017. Towards Bandwidth Efficient Adaptive Streaming of Omnidirectional Video over HTTP: Design, Implementation, and Evaluation. In *Proceedings of the 8th ACM on Multimedia Systems Conference*. ACM, 261–271.

[10] Huawei. 2016. *Whitepaper on the VR-Oriented Bearer Network Requirement(2016).* Technical Report. Huawei Technology. Available at http://www-file.huawei.com/~/media/CORPORATE/PDF/white%20paper/whitepaper-on-the-vr-oriented-bearer-network-requirement-en.pdf.

[11] Evgeny Kuzyakov. 2015. Under the hood: Building 360 video. (2015). https://code.facebook.com/posts/1638767863078802/under-the-hood-building-360-video/

[12] Evgeny Kuzyakov. 2016. Next-generation video encoding techniques for 360 video and VR. (2016). https://code.facebook.com/posts/1126354007399553/next-generation-video-encoding-techniques-for-360-video-and-vr/

[13] H.-C. Lin, C.-Y. Li, J.-L. Lin, S.-K. Chang, and C.-C. Ju. 2016. *An efficient compact layout for octahedron format, JVET Doc. D0142.*

[14] Jonathan S Lu, Daniel Steinbach, Patrick Cabrol, and Philip Pietraski. 2012. Modeling human blockers in millimeter wave radio links. *ZTE Communications* 10, 4 (2012), 23–28.

[15] Stefano Petrangeli, Filip De Turck, Viswanathan Swaminathan, and Mohammad Hosseini. 2017. Improving Virtual Reality Streaming using HTTP/2. In *Proceedings of the 8th ACM on Multimedia Systems Conference*. ACM, 225–228.

[16] Zhouyue Pi and Farooq Khan. 2011. An introduction to millimeter-wave mobile broadband systems. 49, 6 (June 2011), 101 – 107.

[17] Feng Qian, Lusheng Ji, Bo Han, and Vijay Gopalakrishnan. 2016. Optimizing 360 Video Delivery over Cellular Networks. In *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges (ATC '16)*. ACM, New York, NY, USA, 1–6.

[18] T. S. Rappaport. 2002. *Wireless Communications: Principles and Practice* (second ed.). Prentice Hall, Upper Saddle River, NJ.

[19] Liyang Sun, Guibin Tian, Guanyu Zhu, Yong Liu, Hang Shi, and David Dai. 2018. Multipath IP Routing on End Devices: Motivation, Design, and Performance. In *IFIP Networking*.

[20] Y.-K. Wang, Hendry, and M. Karczewicz. 2016. *Tile based VR video encoding and decoding schemes, JCTVC Doc. X0077.*

[21] M. Wien, V. Baroncini, J. Boyce, A. Segall, and T. Suzuki. 2017. *Joint Call for Evidence on Video Compression with Capability beyond HEVC, JCTVC Doc. F1002.*

[22] Alireza Zare, Alireza Aminlou, Miska M Hannuksela, and Moncef Gabbouj. 2016. HEVC-compliant tile-based streaming of panoramic video for virtual reality applications. In *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 601–605.

[23] C. Zhang, Y. Lu, J. Li, and Z. Wen. 2016. *segmented sphere projection (SSP) for 360-degree video content, JVET Doc. D0030.*

[24] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport. 2013. 28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city. In *2013 IEEE International Conference on Communications (ICC)*. 5163–5167.

[25] Minhua Zhou. 2016. *A study on compression efficiency of icosahedral projection, JVET Doc. D0023.*